

**KORPUSAVUSTEINEN DISKURSSINTUTKIMUS *DIABETEKSEN*
DISKURSSIPROSODIASTA YLEN VERKKOUUTISISSA**

Peppi Hälvä
Kandidaatintutkielma
Suomen kieli
Kieli- ja viestintätieteiden
laitos
Jyväskylän yliopisto
Kevät 2021

JYVÄSKYLÄN YLIOPISTO

Tiedekunta Humanistis-yhteiskuntatieteellinen tiedekunta	Laitos Kieli- ja viestintätieteiden laitos
Tekijä Peppi Hälvä	
Työn nimi Korpusavusteinen diskurssintutkimus <i>diabeteksen</i> diskurssiprosodiasta Ylen verkkouutisissa.	
Oppiaine suomen kieli	Työn laji kandidaatintutkielma
Aika 04/2021	Sivumäärä 22
<p>Tiivistelmä</p> <p>Tutkimuksessa tarkastellaan diabetessairauteen liittyviä diskurssiprosodioita Ylen verkkouutisissa. Tutkimuksen tarkoituksena on selvittää, miten diabeteksestä uutisoidaan ja miten kielelliset valinnat vaikuttavat lukijalle välittyvään tietoon. Tutkimus kuuluu korpuslingvistiikan alalle, ja se on toteutettu korpusavusteisena diskurssintutkimuksena. Metodina tutkimuksessa on käytetty kollokaatioanalyysia. Tutkimuskysymykset tässä tutkimuksessa ovat: Mitä kollokaatteja sana <i>diabetes</i> saa Ylen verkkouutisten korpuksessa? Millaisia diskurssiprosodioita <i>diabetes</i>-sanaan liittyy?</p> <p>Tutkimuksen aineistona on käytetty Ylen verkkouutisista tehtyä tekstikorpusta. Se on vuosilta 2011–2018, ja se on Kielipankin ylläpitämä korpus Korp-palvelussa. Haku on toteutettu <i>diabetes</i>-hakusanalla kyseisessä palvelussa. Aineiston siistimisen jälkeen sen siirrettiin AntConc-ohjelmaan, jossa <i>diabetekselle</i> laskettiin tilastollisesti merkitsevät kollokaatit. Tuloksia tuli 368, joista tarkasteltiin sataa merkitsevintä. Analyysissa kollokaatit jaoteltiin ensin merkitysryhmiin aineiston pohjalta ja niiden jälkeen diskurssiprosodioihin.</p> <p>Diskurssiprosodioita nousi aineistosta esiin kolme: <i>diabeteksen sairastaminen, tiede ja tutkimus sekä yhteiskunta ja ympäristö</i>. Diskurssiprosodiat olivat pääsääntöisesti sävyiltään neutraaleja, mutta yksittäiset kollokaatit saattoi tulkita negatiivissävytteisiksi. Diskurssiprosodiat toivat esiin sairauden laatua, yleisyyttä, tutkittavuutta sekä yhteyttä muihin sairauksiin. Tutkimuksen tuloksia voidaan hyödyntää esimerkiksi vertailukohtana muissa korpusavusteisissa diskurssintutkimuksissa tai pohjana jatkossa tehtäville saman aiheen tutkimuksille.</p>	
Asiasanat diabetes, diskurssiprosodia, kollokaatti, korpuslingvistiikka, Yle	
Säilytyspaikka Jyväskylän yliopisto	
Muita tietoja	

SISÄLLYS

1	JOHDANTO	1
2	KORPUSLINGVISTIIKKA	3
	2.1 Korpusavusteinen diskurssintutkimus.....	4
	2.2 Diskurssiprosodia	5
3	AINEISTO JA MENETELMÄT	7
	3.1 Ylen verkkouutiskorpus.....	7
	3.2 Korp-konkordanssityökalu.....	8
	3.3 AntConc-konkordanssiohjelma	9
4	DIABETES-TERMIN DISKURSSIPROSODIOIDEN ANALYYSI.....	11
	4.1 <i>Diabeteksen sairastaminen</i>	12
	4.2 <i>Tiede ja tutkimus</i>	14
	4.3 <i>Yhteiskunta ja ympäristö</i>	16
5	PÄÄTÄNTÖ	20
	5.1 Yhteenveto ja pohdinta	20
	5.2 Mahdollisuudet jatkotutkimuksiin.....	22
	LÄHTEET	23

LIITTEET

1 JOHDANTO

Diabetes, vanhalta nimeltään sokeritauti, on pitkäaikaissairaus, joka vaatii jatkuvaa päivittäistä hoitoa. Se johtuu haiman heikkenneestä, vaillinaisesta tai loppuneesta kyvystä tuottaa elimistölle välttämätöntä insuliinia. Diabeteksen kirjo on laaja ja erilaisia diabetestyyppejä tunnetaan nykyisin lukuisia, mutta tunnetuimmat tyypit ovat tyyppin 1 ja 2 diabetes sekä raskausajan diabetes. Suomessa on noin 50 000 tyyppiä 1 sairastavaa diabeetikkoa, 400 000 tyyppiä 2 sairastavaa sekä vuosittain 10 000 odottavaa äitiä sairastuu raskausajan diabetekseen. Lisäksi tietämättään tyyppin 2 diabeetikoita arvellaan olevan Suomessa noin 50 000–100 000. (Diabetesliitto 2021.) Kuten luvuista voi päätellä, on diabetes todella yleinen sairaus Suomessa, mutta myös maailmalla. Maailman terveysjärjestön tilastojen mukaan noin 422 miljoonaa ihmistä sairastaa diabeteksen eri muotoja tällä hetkellä maailmassa. Varsinkin tyyppiä 2 sairastavia diabeetikoita on maailmanlaajuisesti todella paljon ja tilastollista nousua on myös nähtävissä. (World Health Organization 2021.) Uutisoinnissa diabetes nousee esille tasaisin väliajoin. Pöllänen ja Vartiainen ovat tutkimuksessaan todenneet, että uutisointi diabeteksestä kytkeytyy kampanjoiden ja juttusarjojen avulla lukijoiden elintapojen muuttamiseen (Pöllänen & Vartiainen 2006). Myös Bednarek on tuonut esiin, että diabeetikoista ja diabeteksestä uutisointi painottuu negatiiviseen sisältöön, vaikkakaan uutisointi ei saa tarpeeksi näkyvyyttä (Bednarek 2020). Mediatekstit luovat todellisuudesta omia muunnelmiaan, jotka vaihtelevat tekstintuottajien yhteiskunnallisen aseman, etujen ja päämäärien mukaan. Muunnelmat syntyvät niistä valinnoista, joita tehdään tekstien tuottamisen aikana. (Fairclough 1995: 136.) Uutisointi diabeteksestä on joko vajavaista tai keskittynyt yksittäisiin ilmiöihin ja todella laajoihin kokonaisuuksiin.

Tässä kandidaatintutkielmassani tutkin *diabetes*-termiin liittyvää uutisointia ja sen mukanaan tuomaa tietoa ja käsitystä diabeteksestä sairautena. Aiheeni pohjautuu vahvasti uutisointiin ja siihen, miten sairaudesta kirjoitetaan. Tutkin *diabeteksen* kanssa esiintyviä sanoja sekä niistä ilmeneviä diskursseja. Aineistonani toimii Ylen verkkouutisten korpus, sillä Liikenne- ja viestintäministeriön hallinnonalan alla toimiva Yle (Yle 2019) oli vuonna 2019 tehdyn Kunnallisalan kehittämissäätiön

kyselyn mukaan suomalaisten mielestä luetettavin uutisoija yhdessä Suomen tietotoimiston (STT) kanssa (Kunnallisanalan kehittämissäätiö 2019). Näin ollen tutkimuksen kohteena on merkittävä suomalainen uutisointikanava. Aineiston keruussa sekä analyysissä olen käyttänyt korpusavusteisen diskurssintutkimuksen menetelmiä. Tässä tutkimuksessa tutkimuskysymykseni ovat:

- 1) Mitä kollokaatteja sana *diabetes* saa Ylen verkkouutisten korpuksessa?
- 2) Millaisia diskurssiprosodioita *diabetes*-sanaan liittyy?

Tutkimukseni on merkittävä, sillä median harhaanjohtava uutisointi ja vanhentuneeseen tietoon perustuvat huhupuheet levittävät väärää tietoa diabeteksestä (Luoma 2015). On tähdellistä tutkia, kuinka diabeteksestä todella uutisoidaan, sekä millaisia käsityksiä lukijalle todella lukemastaan muodostuu ja mitkä seikat uutisoinnissa korostuvat. Vastaavaa tutkimusta ei Suomessa ole oikeastaan tehty. Aikaisemmat suomalaiset korpusavusteiset diskurssintutkimukset liittyvät vahvasti keskustelupalstojen kielenkäytön analysointiin (Jantunen 2018, Ainiala & Jantunen 2019 sekä Johansson, Jantunen, Heimo, Ahonen & Laippala 2018). Uutisointiin liittyvää korpustutkimusta on tehnyt Annunen Tampereen yliopistosta tutkien seksuaalivähemmistöistä uutisointia (Annunen 2020). Lisäksi Suomessa diabeteksestä uutisointiin keskittyvää tutkimusta on tehty ainoastaan 2. tyypin diabetekseen keskittyen (Pöllänen & Vartiainen 2006, Paju 2006). Ulkomailla on julkaistu diabeteksen uutisointiin keskittyvää tutkimusta (ks. Bednarek 2020), sekä diskurssintutkimusta AIDS:ia sairastavista naisista (Hoosen & Collins 2004) sekä diskurssianalyysia rintasyövän uutisoinnista kanadalaisissa medioissa (McGannon 2016), mutta Suomessa tällaiselle tutkimukselle on tilaa. Varsinkin korpusavusteisen diskurssintutkimuksen kentällä on tilaa uusille tutkimuksille. Tässä tutkimuksessa tarkastelen ensin tutkimuksen teoreettisia sekä käytännön taustoja luvuissa 2 sekä 3, ja etenen sitten omaan analyysiini sekä pohdintaan luvuissa 4 ja 5.

2 KORPUSLINGVISTIikka

Korpuslingvistiikka on kielentutkimuksen osa-alue, joka keskittyy kielen muodon ja toiminnan tutkimiseen sekä analysoimiseen käyttäen apuna tietokoneistettua tekstikorpusta (Partington, Duguid & Taylor 2015 : 5). Korpuslingvistisessä tutkimuksessa analysoidaan siis kielipiirteiden, esimerkiksi *diabetes*-sanana, esiintymiä korpuksissa, jotka edustavat tutkittavan kielenkäytön aluetta (Tyrkkö & Taavitsainen 2012: 309).

Korpuksat ovat laajoja tekstikokoelmia, jotka ovat koottu ja ryhmitelty nimenomaan tutkimuksen tarpeisiin (Nevalainen & Rissanen 2012: 320). Ne ovat kokoelmia strukturoiduista, selkeiden periaatteiden mukaan luoduista tekstivalikoimista. Niiden avulla voidaan selvittää empiirisesti kielenkäytön ominaisuuksia valitussa tekstiympäristössä. Korpusten perusteella saadut tulokset voi yleistää koko kieltä tai sen tiettyä genreä ja tekstilajia koskevaksi (Tyrkkö & Taavitsainen 2012: 309). Kuitenkin korpuksen perusteella saadut tulokset ovat sidoksissa sen aikaan, paikkaan ja kontekstiin (Partington, Duguid & Taylor 2015 : 6). Silloin ne ovat kuvauksia pienestä palasta kielenkäyttöä. Valmiiksi koottujen korpusten laajuuden myötä pystytään tunnistamaan kielen variaatioita toistuvissa leksikaalisissa rakenteissa, mitä ei normaalista kielenkäytöstä voisi analysoida (Sinclair 1991: 171). Korpuksat ovat tehokas tapa tehdä erilaisia automatisoituja hakuja juuri niiden tutkimustarkoituksiperäisyyden vuoksi (Luodonpää-Manni, Hamunen, Konstenius, Miestamo, Nikkanen & Sinnemäki 2020 : 460). Valmiiksi koottu aineisto tarjoaa lähes rajattomat mahdollisuudet erilaisten kielenpiirteiden tutkimiseen. Korpuslingvistiikka on kielentutkimuksen alana todistepainotteinen, jossa kielenkäyttötilanteiden ja kontekstien tuotoksia ja jälkeenjättämiä tekstejä analysoidaan (Partington, Duguid & Taylor 2015 : 5). Korpuslingvistiikassa käytettyjä tyypillisiä analyysimenetelmiä ovat muun muassa konkordanssi- sekä kollokaatioanalyysi, joiden avulla voidaan selvittää listamaisesta tekstimassasta analysoitavan sanan kanssa tavallista tiheämmin esiintyvät eli kollokoivat sanat.

Korpuslingvistiikassa analysoidaan tilastollisesti merkitseviä sanoja niiden saaman frekvenssin eli esiintymistiheyden mukaan (Partington, Duguid & Taylor 2015 : 8).

2.1 Korpusavusteinen diskurssintutkimus

Korpusavusteinen diskurssintutkimus yhdistää korpuslingvistiikan ja diskurssintutkimuksen työtapoja ja periaatteita. Korpusavusteinen diskurssintutkimus eli Corpus-assisted discourse studies (CADS) määritellään korpuslingvistiikan tavoin kielentutkimuksen muodoksi, jossa analysoidaan kielen muotoa ja käyttöä. Sitä kuitenkin analysoidaan kommunikatiivisena diskurssina, jossa tietokoneellisen korpuksen avulla tehdään diskurssien analyysia. CADS yhdistää kvalitatiivisen diskurssintutkimuksen sekä kvantitatiivisen korpustutkimuksen. (Partington, Duguid & Taylor 2015: 10.) Korpusavusteinen diskurssintutkimus sopii omaan tutkimukseeni juuri siksi, että se antaa määrällisen työkalun aineiston keräämiseen sekä laadullisen analyysin tulosten analysointiin. Perinteisen korpuslingvistiikan rinnalle diskurssianalyysi tuokin juuri laajan analyysimahdollisuuden tilastollisten tapojen ulkopuolelta (Partington, Duguid & Taylor 2015: 12). Tutkimustavat täydentävät toisiaan, vaikkakin pitkään ajateltiin, ettei kvantitatiivista ja kvalitatiivista menetelmää sovi sekoittaa. Korpuslingvistiikan avulla saadaan kuitenkin liitettyä kielenkäyttötilanteiden analyysiin tietoa, jota ei diskurssianalyysin menetelmin voitaisi päätellä. Laajojen aineistojen ja tilastojen analyysin avulla tuloksista saadaan yleistettävämpiä. (Partington, Duguid & Taylor 2015: 11.)

Diskurssilla tarkoitetaan laajuudessaan lauserajat ylittävää, kontekstissaan sidonnaista kielenkäyttöä. Sen ajatellaan olevan myös kielenkäyttöä, joka heijastaa sosiaalista toimintaa ja sen funktionaalista lähtökohtaa. (Heikkinen 2012: 94–95.) Sinclair ilmentää, että diskurssi on kielen merkitys lauseen yläpuolella (Sinclair 1991 : 172). Diskurssi määritellään laajasti sosiaalisen toiminnan ja kielenkäytön yhteenkietoutumiseksi (Luodonpää-Manni ym. 2020: 375), ja tutkimuksessa keskiö on kielenkäytön tilanteissa sekä laajemmin kulttuurisissa ja yhteiskunnallisissa tilanteissa. Korpusavusteisessa diskurssintutkimuksessa diskurssin ajatellaan olevan toistuvien kielenpiirteiden kokonaisuus, joka ilmentää yhteiskunnassa vallalla olevia, yhteisiä sosiaalisia merkityksiä. Diskurssintutkimuksessa päämääränä on muiden laadullisten tutkimusten tapaan ymmärtää ihmisten käyttäytymistä ja hahmottaa niitä tulkintoja ja merkityksiä, joita rakennetaan sosiaalisessa vuorovaikutuksessa muiden ihmisten kanssa (Pietikäinen & Mäntynen 2019 : 211). Kieli on pelkistymätön osa sosiaalista elämää (Fairclough 2003 : 2), ja sitä tutkimalla pyritään kuvaamaan selittämään ja ymmärtämään ilmiöitä ja niiden luonnetta (Pietikäinen & Mäntynen

2019 : 210). Lähestymistapana diskurssintutkimuksessa on merkitysten rakentuminen kielen ja sosiaalisen toiminnan kautta, jossa keskitytään niin pieneen yksityiskohtaiseen kielenkäyttöön sekä laajempien verkostojen analyysiin tekstipiirteiden ja sosiaalisten rakenteiden välillä (Pietikäinen & Mäntynen 2019 : 210, Fairclough 2003 : 2-3). Kielen monipuolisuus todellisissa tilanteissa sosiaalisena käytäntönä kiteytyy (Pietikäinen & Mäntynen 2019 : 37) diskurssintutkimuksesta hyvin myös korpuslingvistiikkaan, jossa kielenpiirteet usein ovat autenttista kielenkäytön ympäristöistä.

Kieltä tuotetaan fraseologisesti, muistiin varastoituneiden, tunnettujen ilmausten kautta. (Jantunen 2018b : 27.) Fraseologisuus korostaa lauseyhteyttä, sanojen yhdessä esiintymisen eli kollokoinnin tarkoitusta ja merkitystä (Stubbs 2001), ja korpuslingvistiikka sekä diskurssintutkimus korpusavusteisesti keskittyy näiden vakiintuneiden leksikaalisten yksiköiden tutkimiseen (Jantunen 2009 : 357). Kielitieteessä diskurssilla on monia käyttöyhteyksiä, mutta kaikessa se viittaa kielen ja sen ympäristön väliseen analyysiin sekä merkityksiin (Heikkinen 2012: 94-95). Diskurssianalyysi on nimenomaan kielenkäytön analyysia, joka keskittyy vahvasti asiayhteyteen ja kontekstiin (Partington, Duguilt & Taylor 2015: 2) ja siihen, mitä kielellä tehdään ja miten kieli sekä kielenkäyttö yhdistyy muihin yhteiskunnallisiin ja sosiaalisiin rakenteisiin (Luodonpää-Manni ym. 2020 : 374). Diskurssianalyysissa kyse on myös luettujen aiheiden ja lukijalle tarkoitettujen tietojen ymmärtämisestä sekä niiden yhdistämisestä ymmärtämäämme, aikaisemmin kerättyyn tietokokonaisuuteen. (Yule 1996 : 139-140.) Diskurssianalyysin tavoitteena onkin yhdistää kielen säännönmukaisuuksia sosiaalisen toiminnan ja kontekstin säännönmukaisuuksiin (Luodonpää-Manni ym. 2020 : 374).

2.2 Diskurssiprosodia

Diskurssiprosodialla tarkoitetaan tässä kielentutkimuksessa kielellisten ilmausten ominaisuutta esiintyä niille tyypillisissä diskurssikonteksteissa. Ilmiössä tutkitaan kielenkäytön tarkoitusta luoda koherenssia diskursseissa ja kuinka tavanomaisesti tietty kielellinen valinta liitetään kyseisiin diskursseihin. Esimerkiksi verbivalinta *aiheuttaa* (*cause*) liitetään usein epämiellyttäviin asioihin, samoin kuin *hankkia* (*provide*) taas esiintyy yhteyksissä, joissa puhutaan halusta tai tarpeesta. Diskurssiprosodiat ilmaisevat puhujan asennetta tai pragmaattisia funktioita, joten valittu kielimuoto osoittaa kokemuksia ja vakiintuneita käsityksiä aiheesta. Ja koska ilmaisut edustavat puhujan mielipiteitä, diskurssiprosodiat ilmaisevat myös puhujan syytä tehdä kyseinen lausunto. (Stubbs 2001: 65-66.) Diskurssiprosodiaa lähestytään tarkastelemalla kollokaatteja ja etsimällä niitä yhdistäviä kontekstuaalisia merkityksiä

(Jantunen 2018b: 29). Ryhmittely ja siten myös diskurssiprosodian analyysi perustuu ajatukseen, jonka mukaan yhtä tiettyä diskurssia tuotetaan ja toisinnetaan useiden eri kielenpiirteiden, tässä tutkimusyhteydessä siis kollokaattien, avulla (Jantunen 2018b: 36).

Diskurssiprosodian ilmiölle on ollut tarjolla korpusavusteisen diskurssintutkimuksen kehittymisen myötä useita eri nimityksiä. Stubbs nimeää diskurssiprosodian terävöittävän terminä puhujan ja kuulijan välistä suhdetta, mutta myös niiden vaikutusta luoda diskurssien koherenssia eli yhteensopivuutta (Stubbs 2001: 66). Partington taas kuvailee samaa ilmiötä yksinkertaisesti kollokaattien väleinä ja suhteina toisiinsa (Partington, Duguil & Taylor 2015: 58). Sanavalinnat ja niiden lataamat mielikuvat vaikuttavat lukijan käsitykseen aiheesta, ja lopulta koko prosodia onkin juuri lukijan käsityksen varassa. Kollokaatteja ja diskursseja pohtiessa on kuitenkin otettava huomioon myös kollokaatteja laajempi merkitysyhteys. Pelkästään yksi sana yhdessä toisen kanssa ei välttämättä riitä luomaan negatiivista tai positiivista vaikutusta. On tarkasteltava koko lauseen merkitystä ja sävyä tehdäkseen päätelmiä kirjoittajan tarkoittamasta tyylistä. (Stubbs 2001: 66.)

3 AINEISTO JA MENETELMÄT

Tässä luvussa käsittelen tämän tutkimuksen aineistoa, sen keräämistä ja sen käsittelemiseen käyttämiäni menetelmiä. Tarkastelen lyhyesti Ylen verkkouutisten korpusta, josta keräsin aineistoni sekä kahta työkalua, joiden avulla keräsin ja rajasin korpusaineistoani.

3.1 Ylen verkkouutiskorpus

Ylen verkkouutiskorpus on Kielipankin ylläpitämä julkinen korpus, joka on kooltaan 18 708 231 lausetta ja yli 207 miljoonaa sanetta. Korpus on luotu vuonna 2019 ja se pitää sisällään kaikki osoitteessa yle.fi julkaistut artikkelit vuosilta 2011–2018. (Kielipankki 2021.) Korpus on melko uusi ja kooltaan myös huomattavasti pienempi kuin esimerkiksi suosittu Suomi24-keskustelupalstan korpus, jota esimerkiksi Jantunen (2018) on analysoinut. Tämä tulee huomioida aineiston keräämisessä ja merkitsevien sanojen tarkastelussa, sillä aineisto on eri kontekstista ja virallista tarkoitusta varten kirjoitettua. Käytettävän korpuksen ollessa julkinen ovat myös siitä saadut aineistot toistettavissa (Stubbs 2001: 123). Tämä vahvistaa tulosten luotettavuutta ja käytettävyyttä.

Korpus koostuu Ylen uutisartikkeleista, joten myös tekstien alkuperäinen tarkoitus tulee huomioida. Sisältö korpuksessa on kielelliseltä muodoltaan asiallista ja siistiä, eikä siinä näy kirjoittajan henkilökohtainen mielipide. Mukana aineistossa on kaikki Ylen määrättyinä vuosina julkaistut artikkelit, joten mukana on myös esimerkiksi haastatteluja ja sitaatteja. Tämä tulee myös huomioida tuloksia tarkastellessa, mutta myös tutkimuseettisiä kysymyksiä pohtiessa. Korpus on koottu tutkimusta varten ja se on vapaassa käytössä, joten aineiston saatavuuden kannalta eettisiä ongelmia ei tule. Toki aineistossa ilmenee henkilöiden nimiä, joita tulee käsitellä arvokkaasti, mutta tässä kielitieteellisessä tutkimuksessa ei nimiin kiinnitetä

huomiota. Lisäksi nämä nimet ovat osa alkuperäisiä verkkouutisia, jotka ovat kaikkien luettavissa myös Yle.fi -sivustolla. Ylen verkkouutiskorpuksessa artikkeleiden virkkeet on sekoitettu, eikä aineisto tue kappalekontekstia (Kielipankki 2019). Tutkimukseni kannalta Ylen korpus on sopiva, koska kieli on neutraalia ja sen tarkoitus on olla informatiivista.

3.2 Korp-konkordanssityökalu

Korp-konkordanssityökalu on Kielipankin ylläpitämä selainpohjainen sivusto, jossa on mahdollista tehdä konkordanssihakuja tekstiaineistoista morfosyntaktisen jäsennyksen perusteella (Kielipankki 2021). Konkordanssilla tarkoitetaan Korp-käyttöliittymän luomaa vertikaalisesti luettavaa listausta sanojen esiintymisestä lähikontekstissaan (Jantunen 2018b: 24), yksinkertaisuudessaan siis tekstin sanojen listausta (Sinclair 1991: 170). Käytännössä Korp-sivusto antaa lyhyenä tekstikontekstina listauksen kaikista yhteyksistä, jossa haettu sana ilmenee. Haku voidaan toimittaa joko yhdellä sanalla tai useamman sanan yhdistelmällä, sekä myös esimerkiksi erikoismerkillä. Aineisto muodostuu siis täysin käyttäjän hakusanan sekä asetusten mukaan.

Tässä tutkimuksessani aineisto on haettu hakusanalla *diabetes* Ylen korpuksessa kokonaisuudessaan aikaväliltä 2011–2018, sillä sisällöltään korpus on vielä melko suppea verrattuna muihin korpuksiin. En ole rajannut vuosia enempää, jotta tuloksista saataisiin mahdollisimman kattavat ja analysoitavaa olisi tarpeeksi. Haku on muotoiltu sanan perusmuodolla haettavaksi virkkeen sisältä. Vielä laajemman tuloksen saisi hakemalla tuloksia virkkeen sijasta kappaleen sisältä (Jantunen 2018: 7), mutta Ylen korpuksessa tähän ei ole mahdollisuutta. Haussa on lisäksi määritelty hakusanan ja kollokaatin väliseksi suhteeksi neljä sanaa oikealle ja neljä sanaa vasemmalle (4L–4R), sillä suomen kielessä tähän akseliin mahtuu tyypillinen keskipituinen lause (Jantunen 2018b: 29). Konkordansseja haulla tuli 3754. Saadut konkordanssit siirrettiin kokonaisuudessaan Excel-taulukko-ohjelmaan, jossa aineistoa karsittiin. Aineistosta poistettiin kaksoiskappaleet sekä lemmatisoiduista että lemmatisoimattomista konkordansseista. Lemmatisoidulla tarkoitetaan tässä tutkimuksessa perusmuotoon muutettua sanaa. Poistojen jälkeen konkordansseja jäi 3066 ja 41 040 sanetta. Excelistä lemmatisoitu aineisto siirrettiin Word-tiedostoon, jossa siitä poistettiin vielä ylimääräiset erikoismerkit sekä turhat välilyönnit. Ennen AntConc-ohjelmaan viemistä tarkastin vielä aineistosta manuaalisesti mahdolliset muotoiluvirheet, joissa lemmatisoitu muoto ei vastannut merkitykseltään taivutettua sanaa. Esimerkiksi *huolestuttava* partisiippi muuttui muotoon *huolestuttaa* ja verbin *sairastaa* partisiippi *sairastava* muotoon *sairastapa*. Nämä korjattiin manuaalisesti.

3.3 AntConc-konkordanssiohjelma

AntConc-konkordanssiohjelma on ilmaiseksi tietokoneelle ladattava ohjelma, jolla pystytään laskemaan yleisimpiä tilastollisia todennäköisyyksiä käsiteltävistä tekstiaineistoista (Partington, Duguilt & Taylor 2015: 344). Tässä tutkimuksessa on valittu käytettäväksi kollokaatioanalyysi. Siinä lasketaan MI-testillä kahden tai useamman sanan esiintyminen yhdessä tilastollisesti merkittävän monta kertaa. MI-testillä (Mutual Information) tarkoitetaan tilastollista testausta, jossa hakusanan ja sen ympäröivien sanojen myötäesiintymisen todennäköisyyttä mitataan. Testi perustuu kahden sanan relatiivisen aseman huomioimiseen, yksinkertaisesti siihen, kuinka monta kertaa sana x esiintyy lähellä sanaa y . (Baker 2010: 24.) Kollokaatioanalyysi perustuu sanojen taipumukseen esiintyä yhdessä keskenään samaan merkitysryhmään kuuluvien sanojen kanssa. Merkitysryhmillä tässä tutkimuksessa tarkoitetaan kollokaattiryhmiä, joissa kollokaateilla on samankaltainen merkitys. AntConc-ohjelmassa on valmiiksi sisäänohjelmoitu työkalu, joka määrittää kollokaatit halutuille sanoille. Tämän avulla tuloksina saadut kollokaatit voivat olla odottamattomia, joka tekee analyysistä joustavaa ja monipuolista. (Luodonpää-Manni ym. 2020: 497–498.) MI-testin avulla kollokaatioanalyysissä saadaan selville vahvemmin yhdessä merkitsevästi esiintyvät sanat, sillä testi ottaa huomioon myös yhdessä esiintymisen määrän lisäksi erillään esiintymisen. Näistä esiintymistä testi antaa MI-luvun, jonka ollessa suurempi kuin kolme, tulos on tilastollisesti merkittävä ja myötäesiintymä tapahtuu usein. (Baker 2010: 24–25). Nämä MI-testin tulokset ovat kokonaisuudessaan tarkasteltavissa osana liitettä 1.

Tähän tutkimukseen valittiin tehtäväksi juuri MI-testi, koska sen antaman arvon avulla voidaan tarkastella sanoja riippumatta niiden frekvenssistä eli esiintymismäärästä. Jos käytettäisiin kollokaatioanalyysin sijaan avainsana-analyysia, ottaisi se huomioon myös kollokaattien esiintymismäärän. (Luodonpää-Manni ym. 2020: 498). MI-testin hyvänä puolena nähdään siis tulosten arvaamattomuus, sillä yhteisesiintymistä ei voi nähdä suoraan aineistosta. Lisäksi MI-testi on tähän tutkimukseen hyvä, koska se yleensä nostaa merkitseviksi sanoiksi sisältösanoja eikä niinkään kieliopillisia funktiosanoja (Jantunen 2018b: 29). AntConc-ohjelmassa rajattiin lisäksi kollokaattien vähimmäisfrekvenssiksi 10, jotta kokonaisfrekvenssiluku saatiin tarpeeksi isoksi ja näin tulokset merkitseviksi.

AntConc antaa myös kollokaattien yhteydessä esiintymisyhteyden konkordanssin muodossa, josta kollokaatin merkityksen ja toimintatarkoituksen voi tarkistaa. Saatujen yhteyksien avulla luotiin kaikille kollokaateille esimerkkilauseet yhteyksineen, jotta kollokaattien hallitseminen ja luokittelu helpottuisi. Merkitysryhmiin jaon taustalla on tässä ajatus seuraavasta vaiheesta, jossa kollokaatteja tarkastellaan niiden diskurssiprosodioiden mukaan. Merkitysten ja

merkitysryhmien kautta jako diskurssiprosodioihin sujuu helpommin ja sanojen todellinen merkitys on jo tiedossa. Oman haasteensa luokitteluun toi kuitenkin kokonaisuudessaan aihepiirin tiivis yhtäläisyys. Monet sanat olisivat voineet kuulua moneen eri ryhmään, sillä kaikki sanat kietoutuvat sairastamisen ympärille. AntConcista saadut sata merkitsevintä kollokaattia ajettiin Exceliin, jossa niistä muodostettiin merkitysryhmiä. Sadan merkitsevimmän sanan joukkoon valikoituivat lähinnä sisältösanat substantiivit, adjektiivit, verbit ja adverbit, joilla viitataan tekstissä ulkomaailman asioihin. Näillä sanoilla on tyypillisesti prototyyppiset merkityksensä, johon niitä käytetään. (VISK § 438.) Siksi analyysi myös kohdistuu pääasiassa niihin sanoihin. Myös ryhmien ulkopuolelle jääviä epäselviä sanoja ilmeni. Ne poistettiin luokittelusta kokonaan lähinnä niiden funktionaalisuuden takia, sillä niillä on selkeämmin vain kieliopillinen ja tekstin kulkuun liittyvä tehtävä (VISK § 438). Näitä sanoja oli 6%, *vaikkapa, paitsi, erityisesti, eli, joku ja ehkä*. Tyyliltään sisältösanat olivat neutraaleja, mikä ei ole yllättävää uutisartikkelien yhteydessä.

4 DIABETES-TERMIN DISKURSSIPROSODIOIDEN ANALYYSI

Tässä luvussa käsitellään analyysini etenemistä, *diabetes*-sanan kollokaatteja ja niistä muodostuvista merkitysryhmistä ilmeneviä diskurssiprosodioita. AntConc-ohjelma antoi 368 kollokaattia *diabetekselle*, joista tässä tutkimuksessa ja luvussa tarkastelen sataa merkitsevintä. Alaluvuissa käsittelen merkitysryhmistä muodostuvia diskurssiprosodioita *diabetes*-termille tarkemmin erikseen. Tuon esille analyysin aikana esiinnousseita kohtia sekä pohdin yleisesti analyysin etenemistä.

Merkitysryhmien kollokaateista muodostui kolme diskurssiprosodiaa. Kaikista selkein prosodia oli *tiede ja tutkimus*, sillä tutkimukseen liittyvä sisältösanasto erottui muista aiheista. Määrällisesti suurin prosodia oli kuitenkin *yhteiskunta ja ympäristö*. Kolmanneksi prosodiaksi muodostui *diabeteksen sairastaminen*. Alla olevassa taulukossa näkyy kollokaattien prosentuaalinen jakautuminen diskurssiprosodioihin. Huomioitavaa on, että kollokaatti *tausta* on kahdessa prosodiassa sen monimerkityksisyyden takia. Seuraavaksi käsittelen prosodioita yksitellen erikseen.

Taulukko 1. *Diabetes*-termin kollokaattien diskurssiprosodioiden prosentuaalinen määrä

Diskurssiprosodia	prosenttiosuus
<i>diabeteksen sairastaminen</i>	32 %
<i>tiede ja tutkimus</i>	27 %
<i>yhteiskunta ja ympäristö</i>	36 %
poistetut	6 %

4.1 Diabeteksen sairastaminen

Diskurssiprosodioista *diabeteksen sairastaminen* koostuu eri *diabetes*-substantiiviva täydentävistä substantiivimääritteistä. Lisäksi ryhmässä on verbejä sekä muutama henkilöä tarkoittava termi. Ryhmän rajaaminen osoittautui osin todella vaikeaksi, sillä tähän prosodiaan olisi voinut kuulua melkein mikä vain *diabeteksen* kanssa kollokoiva sana. Kuitenkin lopulliset rajanvedot on tehty ajatellen konkreettisesti itse sairautta tai sairastamisen aikana tapahtuvaa toimintaa. Käsittelen ryhmää ensimmäisenä, koska siinä on konkreettisesti kyse diabeteksen sairastamisesta ja siihen vahvasti liittyvistä asioista. Diabeteksen sairastamiseen ja sairauteen itseensä liittyvät kollokaatit näkyvät merkitysryhminä alla olevassa taulukossa.

Taulukko 2. Diskurssiprosodia *diabeteksen sairastaminen*.

<p><i>diabetestyytit</i> tyyppi, lapsuusikä, nuoruustyyppi, nuoruusikä, kakkostyyppi, esiaste, aikuistyyppi, ykköstyyppi, aikuisikä, ykkönen, sokeritauti</p> <p><i>sairastumisen syy</i> seuraus, synty, merkitys, perimä, geeni, hoito, puhkeaminen, tausta</p> <p><i>sairauden kulku</i> kehittyä, vaatia, parantaa, sairastaa, hoitaa, potea, sairastua, tuntea</p> <p><i>muut sairaudet</i> liitännäissairaus, pitkäaikaissairaus, lisäsairaus</p> <p><i>henkilöt</i> sairastava, sairastavia</p>
--

Esimerkeissä 1 ja 2 on nähtävissä diabeteksen laatua kuvaavia määritesubstantiiveja. Määritteenä kollokaatit ovat nimikemääritteitä, jotka toimivat substantiivin edellä (VISK § 563). Kollokaateilla kuvataan, mitä diabetestyyppiä tarkoitetaan.

- 1) Tulokset kuitenkin viittaavat siihen, että **nuoruustyyppin** *diabetekseen* on mahdollista kehittää lääkkeitä, jotka voivat pysäyttää sairauden etenemisen.
- 2) Hiljattain julkaistun tutkimuksen mukaan Suomessa voitaisiin liikuntaan ja terveelliseen ruokavalioon keskittymällä vähentää **kakkostyyppin** *diabeteksen* hoitokuluja jopa 700 miljoonalla vuodessa.

Diabeteksen sairastumiseen johtavia syitä esitellään substantiivein esimerkeissä 3, 4 ja 5. Näistä selviää suorasti, mikä vaikuttaa sairastumiseen ja miten. Esimerkki 6 ilmaisee kuvainnollisesti sairastumisen taustoja mahdolliseen tutkimustietoon nojaten. Asia on sairauden taustalla kuvainnollisesti huomaamattomissa, mutta silti merkittävässä asemassa (Kielitoimiston sanakirja s.v. tausta). Esimerkki 7 ilmaisee sairastamisen aikana tapahtuvaa toimintaa inessiivisijamuodolla viitaten sisäpaikallissijalla hoidon tapahtumiin. (VISK § 81).

- 3) Osa ruokavalion terveyshyödyistä saattaa jäädä saamatta henkilöillä, joiden **perimä** altistaa esimerkiksi *diabetekselle*.
- 4) Lihavuuden ja kakkostyyppin *diabetekseen* vaikuttavien **geenien** tutkimusta voidaan hyödyntää, kun tautien hoitoa kehitetään.
- 5) Verensokerin nopeita, jatkuva heilahteluja kannattaa välttää myös fyysisen terveyden kannalta: insuliinitasapainon järkkyminen voi johtaa *diabeteksen puhkeamiseen*, jos siihen on alttius.
- 6) Jos lapsi sairastuu alle puolivuotiaana *diabetekseen*, on sairauden **taustalla** yleensä mutaatio jossakin insuliinia tuottavien beetasolujen toiminnan kannalta tärkeässä geenissä.
- 7) Oulun Diabetesyhdistyksessä ollaan huolestuneita hallituksen lääkekorvaussäästöjen vaikutuksista tyyppin 2 *diabeteksen hoidossa*

Sairastamista kokonaisuutena ja kronologisena tapahtumana kuvaavat esimerkit 8 ja 9. Intransitiiviverbeillä (VISK § 459) kuvataan sairastamisen ajallista kestoa sekä sairauden aikana tapahtuvaa toimintaa.

- 8) Tutkimuksessa laihduttaminen **paransi** *diabeteksen* joka toiselta potilaalta, ja he myös pääsivät eroon diabetes- ja verenpainelääkityksistä.
- 9) Suomalaisista arviolta 500 000 **sairastaa** *diabetesta*.

Diabetesta kuvataan myös esimerkeissä 10 ja 11 laajemmilla sairautta tarkoittavilla termeillä. Nimikemääritteet (VISK § 563) kuvaavat diabeteksen luonnetta ja statusta sairautena, ja niistä lukijalle muodostuu käsitys sairaudesta. Liitesana *lisä-* esimerkin 11 yhdyssanassa kuvaa sairauden tyyppiä.

- 10) Pikkukeskosilla on lisääntynyt riski aikuisiän **pitkäaikaissairauksiin** kuten osteoporoosiin, *diabetekseen* ja sydän- ja verisuonisairauksiin.
- 11) **Lisäsairauksien** riski syntyy, kun *diabetes* on kestänyt pitkään ja hoitotasapaino on huono.

Diabetesta sairastavia henkilöitä kuvataan kollokaateilla *sairastava* ja *sairastavia* (esimerkki 12). VA-partisiipilla on luotu *sairastaa*-verbikannasta syntaktisesti

adjektiivin kaltaisesti toimiva sana *sairastava* (VISK § 122, 297). VA-partisiippi ilmaisee päättymätöntä toimintaa (VISK § 522), kuten myös esimerkissä. Tässä virkeyhteydessä adjektiivi toimii substantiivin tapaisesti absoluuttisena adjektiivina kuvaten henkilön ominaisuuksia (VISK § 606).

- 12) Yksi lasillinen puna- tai valkoviiniä aterian yhteydessä voi tuoreen tutkimuksen mukaan parantaa aikuistyyppin *diabetesta sairastavien* kolesterolin ja glukoositasoja.

Selkeästi tässä diskurssiprosodiassa nousivat esiin varsinkin diabetesta kuvaavat termit, sairastamisen kulkua kuvaavat verbit sekä yleiset huomiot diabeteksen sairastamisesta ja sairaudesta.

4.2 *Tiede ja tutkimus*

Tieteeseen ja tutkimukseen liittyvä diskurssiprosodia on toinen esiinnoussut diskurssiprosodia. Kollokaateista 27 prosenttia kuului tähän prosodiaan. Ryhmä koostuu verbeistä, substantiiveista sekä adjektiiveista. Verbit kuvaavat tutkimuksen tapahtumista ja tulosten ominaisuuksia. Substantiivit kuvaavat tutkimuksen kohdetta, tulosta tai tieteellistä julkaisua. Adjektiiveja käytetään kuvaamaan tulosten tai sairauden laatua.

Taulukko 3. Diskurssiprosodia *tiede ja tutkimus*.

tutkimuksen kohde ja tieteellinen julkaisu

hiiri, silmä, care, lehti

tutkimuksen tekeminen

yleistyä, paljastaa, ennustaa, altistaa, liittyä, lisääntyä, julkaista, löytää, heikentää, johtua, ilmetä

sairauden ja oireiden kuvaus

varsinainen, tyypillinen, todennäköinen, mahdollinen, herkkä, vaikea

taudin ehkäisy

ehkäisy, rokote, ennaltaehkäisy, riski; seuloa

paikka ja määrä

tausta

Tutkimuksen kohdetta ja tieteellistä julkaisua kuvataan esimerkeissä 13 ja 14. Substantiivit ilmaisevat tutkimuksen tekemisen kohdetta tai siihen tarvittavaa asiaa sekä tutkimuksen julkaisua konkreettisessa muodossaan. Esimerkissä 13 adessiivisijamuodolla ilmaistaan adverbien kaltaisesti tapaa (VISK § 387), vaikka sana ei varsinaisesti yksinään ole tapaa ilmaiseva. Esimerkissä 14 *care*-kollokaatti tarkoittaa yhdysvaltalaisista lehtijulkaisusarjaa, joka on täysin keskittynyt diabeteksen käsittelemiseen. Se on tieteellinen julkaisu, jonka tarkoituksena on tarjota tieteellistä ja ajankohtaista tietoa terveydenhoitoalan ammattilaisille diabeteksen hoidosta ja uusista tapahtumista tutkimuskentällä (Diabetes Care 2021).

- 13) Koe tehtiin **hiirillä**, joille oli aiheutettu *diabeteksen* kaltainen tila lääketieteellisesti tai geenimuuntelulla.
- 14) Tutkimus on hyväksytty julkaistavaksi arvostetussa lääketieteen alan *Diabetes Care* -julkaisusarjassa.

Esimerkit 15, 16 ja 17 kertovat tutkimuksessa selvinneistä syy-seuraussuhteista. Nämä virkkeet toimivat hyvin informatiivisina, johon vaikuttaa hyvin informatiivisten verbien käyttö. Merkitysryhmän verbit ilmaisevat tutkimuksen tuloksiin ja tutkimuksen tekemiseen liittyvää toimintaa. Esimerkissä 15 verbi *ennustaa* on tuloslauseessa arvioiva verbi (VISK § 904), joka ilmaisee epävarmaa lopputulosta. Esimerkkien 16 ja 17 verbit *johtua* ja *ilmetä* ovat intransitiiviverbejä, joilla ilmaistaan muutosta tai ilmiötä (VISK § 459).

- 15) Tulokset osoittivat, että niin painoindeksi kuin vyötärön ympärys **ennustivat** tyypin 2 *diabeteksen* vaarasta itsenäisesti.
- 16) Nuoruustyypin *diabetes* **johtuu** insuliinia tuottavien haimasolujen tuhoutumisesta.
- 17) Äidin kalan syöminen raskauden ja imetyksen aikana voi pienentää perinnölliseen riskiryhmään kuuluvien vauvojen vaaraa sairastua ykköstyypin *diabetekseen* myöhemmin lapsuudessa, **ilmenee** uudesta suomalaistutkimuksesta.

Diabetesta sairautena sekä sen oireita kuvataan modaalisilla adjektiiveilla esimerkeissä 18 ja 19. Varmuuden astetta kuvataan adjektiiveilla *tyypillinen* ja *mahdollinen*. (VISK § 1588.) Ne informoivat lukijalle sairauden mahdollisuuksia sekä yleisyyttä.

- 18) Tämä selvisi tutkimuksessa, jossa oli mukana 40-vuotiaita lappilaisia, joilla oli selkeä riski sairastua **tyypillisiin** kansantauteihimme, kakkostyypin *diabetekseen* ja sepelvaltimotautiin.
- 19) Tutkimuksissa yhdistettiin aikaisempien tutkimusten aineistoja tarkoituksena muodostaa kokonaiskuva kalan **mahdollisista** vaikutuksista aikuistyypin *diabetekseen*.

Sairauden ehkäisy ja siihen liittyvät seikat näkyvät esimerkeissä 20, 21 ja 22. Monissa yhteyksissä puhutaan juuri sairauden ehkäisemisestä tutkimuksen keinoin sekä tutkimuksessa esiin tulleista seikoista sairauden lisääntymisestä. Näissä yhteyksissä lukija saa kuvan tehtävästä seuloonasta ja sairauden eteen tehtävästä ehkäisytyöstä. Esimerkissä 20 *rokote* eli lääketieteellinen valmiste immuniteetin saamiseksi sairautta vastaan (kielitoimiston sanakirja s.v. rokote) ilmaisee tehtävää lääketieteellistä tutkimusta sairauden parantamiseksi. Esimerkissä 21 *riskillä* tarkoitetaan epäedullisen tapahtuman mahdollisuutta (kielitoimiston sanakirja s.v. riski) eli sairastumista ja sen negatiivista leimaa. Esimerkin 21 kaltaisia yhteyksiä ilmeni tuloksissa paljon. Esimerkki 22 kertoo tutkimuksen tarkoituksista sekä sen tekemisestä O-verbillä (VISK § 366) *seuloa*, joka ilmaisee lääketieteessä tautitapausten lajittelua tutkimuksissa (kielitoimiston sanakirja s.v. seuloa).

- 20) Suomalaiset tutkijat ovat kehittäneet **rokotteen** *diabetesta* eli sokeritautia vastaan.
- 21) Istuminen kasvattaa **riskiä** *diabetekseen*, sydän- ja verisuonisairauksiin sekä tuki- ja liikuntaelinsairauksiin.
- 22) Osallistujat satunnaistettiin kolmeen ryhmään, joista kahdessa *diabetesta* **seulottiin** kattavilla testeillä ja kolmas sai toimia vertailuryhmänä.

Aineistosta kävi ilmi, että tutkimus ja diabetes kulkevat eri konteksteissa yhdessä. Tutkimusta tuotiin esille niin sairauden nykyhetken kuin myös sen tulevaisuuden kannalta.

4.3 Yhteiskunta ja ympäristö

Kolmantena diskurssiprosodia aineistossa nousee esiin yhteiskunnalliset kollokaatit. Kollokaateista 36 prosenttia kuului tähän ryhmään, ja niistä suurin osa oli substantiiveja. Myös muutama verbi sekä adjektiivi oli mukana. Ryhmässä mukana olevat substantiivit kuvaavat kansansairauksia, jotka ovat yhteiskunnallisesti merkittäviä sekä yhteiskunnallisia tekijöitä sairastumiseen, yhteiskunnallista roolia tai esimerkiksi paikkaa.

Taulukko 4. Diskurssiprosodia *yhteiskunta ja ympäristö*.

muut sairaudet

kansantauti, kansansairaus;

suolistosairaus, sydäntauti, munuaistauti, sydänsairaus, allergia, verisuonitauti, verisuonisairaus, keliakia, astma, verenpainetauti, reuma, epilepsia, Parkinsonin tauti, kolesteroliongelma, sepelvaltimotauti, aivohalvaus, Alzheimerin tauti, syöpä, verenpaine

sairastumisen syy

ympäristömyrky, raskausaika, liikalihavuus, tupakointi

henkilöt

iäkäs, tytär, aikuinen, äiti

paikka ja määrä

maailma, eniten, vähän, puoli

taudin ehkäisy

välttyä, säästyä

sairauden ja oireiden kuvaus

huolestuttava

Ylen artikkeleissa on nähtävissä diabeteksen rinnastaminen muihin sairauksiin yhteyksissä, joissa puhutaan asian x aiheuttamista ongelmista tai seurauksista. Esimerkki 23 ilmentää diabeteksen paikkaa osana suomalaista geeniperimää substantiivilla *kansantauti*, joka tarkoittaa laajasti esiintyvää tautia (kielitoimiston sanakirja s.v. kansantauti). Esimerkeissä 24 ja 25 tyylikeinona sairauksien rinnastamisella (VISK § 1082) merkitään niiden yleisyyttä ja yhteyksiä muihin sairauksiin yhteiskunnan tautikirjossa.

- 23) Tutkimusten avulla pyritään selvittämään kuusamolaisten ja muidenkin suomalaisten geeniperimästä tavallisille **kansantaudeille**, kuten *diabetekselle* sekä sydän- ja verisuonitaudeille, altistavia muutoksia.
- 24) Maailman kantasolututkijat toivovat löytävänsä ihmisen erikoistumattomista soluista parannuskeinon muun muassa *diabetekseen*, **Alzheimerin tautiin** ja Parkinsonin tautiin.
- 25) Iän ohella eteisvärinälle altistavat kohonnut verenpaine, **sydänsairaudet**, *diabetes* ja liikapaino.

Myös erilaisia uskomuksia ja olettamuksia eri sairauksien syntyyn ja ehkäisyyn ilmeni. Esimerkeissä 26 ja 27 arvuutellaan sairastumisen syytä. Nämä erilaiset uskomukset, jotka myös ilmentävät yhteiskunnan ongelmia ja ajatusmalleja, vaikuttavat paljon lukijan mielikuvaan diabeteksestä. Substantiivit *allergia* ja *sepelvaltimotauti* ovat rinnastettuina yhdessä *diabetes*-termin kanssa. Sairastamisesta yleisesti nousee esimerkissä 28 hieman negatiivissävytteinen kuva käytetyn *kolesteroliongelma*-termin takia. Kokonaisuudessaan kaikki sairastaminen koetaan usein negatiivisena asiana ja yleisesti kaikki sairastamisen ympärillä liikkuva kielenkäyttö liittyy paranemiseen. *Ongelma*-termi kuitenkin luo tässä yhteydessä lukijalle jo valmiiksi kielteisessä yhteydessä negatiivisen käsityksen, vaikka *ongelma*-sanaan liittyy myös pyrkimys sen parantamiseen tai ratkaisemiseen (kielitoimiston sanakirja s.v. *ongelma*).

- 26) Runsaalla täyssäilykkeiden syönnillä saattaa olla yhteys ykköstyypin *diabeteksen* ja **allergioiden** lisääntymiseen.
- 27) Suuren nitriitti-, samoin kuin nitraattimäärien on epäilty lisäävän mm. *diabeteksen*, **sepelvaltimotaudin** ja syövän syntyä, mutta asiasta ei ole tieteellistä varmuutta.
- 28) Professori Hannu Mykkänen sanoo lehdelle, että vehnäleipä ei sovi etenkään niille, jotka kärsivät kakkostyypin *diabetekselle* altistavasta keskivartalolihavuudesta tai **kolesteroliongelmista**.

Sairastumisen syyt ja siihen johtavia tekijöitä esitellään esimerkeissä 29 ja 30 substantiiveilla *ympäristömyrkky* ja *liikalihavuus*. Yhteiskunnassa olevien ulkoisten tekijöiden tai yksilön henkilökohtaisten tapojen seurauksena sairastuminen nähdään myös yhteiskunnallisena ongelmakohtana.

- 29) Ympäristössä pitkään säilyvät **ympäristömyrkyt** saattavat altistaa aikuistyyppin *diabeteksen* kehittymiselle, hiljattain julkaistu suomalaistutkimus osoittaa.
- 30) Etelä-Pohjanmaalla erikoissairaanhoidoa kuormittavat varsinkin kansantaudit kuten *diabetes*, sydänsairaudet ja myös **liikalihavuus**.

Henkilöt ja ihmiset ovat merkittävä osa yhteiskuntaa ja sen toimintaa. Esimerkeissä 31, 32 ja 33 kuvataan henkilörooleja sairastumisessa. Esimerkissä 31 *iäkäs*-kollokaatti adjektiivina substantiivikäytössä viittaa ikäryhmään toimien lauseessa subjektina (VISK § 626). Myös esimerkeissä 32 ja 33 kollokaatit merkitsevät iästä syntyvää yhteiskuntaeroa ja sukulaisuussuhdetta.

- 31) *Diabetesta* sairastavat **iäkkäät** saattavat olla tavallista alttiimpia dementialle ja sen lievemmillä muodoilla, osoittaa tuore tutkimus.
- 32) Tuula-Maria Rintalan väitöstutkimuksen mukaan **aikuisen** *diabetes* vaikuttaa perheen elämänrytmiin ja voi aiheuttaa myös pelkoja.

- 33) Mikko Linnolahti alkoi tutkia verensokerin tasaamisen mahdollisuuksia vuonna 2012, varmistaakseen **tyttärensä** ykköstyypin *diabeteksen* turvallisen hoidon.

Paikan ja määrän adverbit (VISK § 646) ilmaisevat esimerkeissä 34, 35 ja 36 diabeteksen määrää ja laajuutta yhteiskunnassa.

- 34) Hän on yksi todella suuresta joukosta sillä *diabetesta* sairastaa **maailmassa** 366 miljoonaa ihmistä.
- 35) Jostain syystä Suomessa on maailman **eniten** lasten ykköstyypin *diabetekseen* sairastuneita väkilukuun suhteutettuna.
- 36) Arvioiden mukaan jo lähes **puoli** miljoonaa suomalaista sairastaa tyyppin kaksi *diabetesta*.

Yhteiskunnan arvoja ja ajatusmalleja ilmentävät esimerkit 37, 38 ja 39. Negatiivisen sävyn omaavat mentaaliset verbit ilmaisevat sairauden ja sairastumisen aikaansaamia tunteita (VISK § 445). Sairastuminen kuvataan huonona, lopullisena asiana ja siltä säästyminen hyvänä asiana. Esimerkki 39 kuvaa yleisesti sairastamisen aikaansaamaa tunnetta VA-partisiipilla muodostetulla adjektiivilla.

- 37) Runsaasti kahvia juovien on useissa tutkimuksissa nähty **välttyvän** muun muassa *diabetekselta* muita todennäköisemmin, mutta kahvilla saattaa olla suotuisia vaikutuksia myös suolistosyöpää sairastaviin.
- 38) Paljon magnesiumia ravinnostaan saavat saattavat **säästyä** aikuistyypin *diabetekselta* muita todennäköisemmin.
- 39) Ongelma on kasvava, sillä *diabetes* yleistyy lasten ja nuorten parissa **huolestuttavaa** vauhtia.

Tässä diskurssiprosodiassa yhdistyvät diabeteksen kuvaaminen yhteiskunnallisella tasolla sekä siihen johtavat yhteiskunnalliset seikat. Diabeteksen laajuus sairautena tässä yhteiskunnassa tulee hyvin vahvasti esiin konkreettisista sekä abstrakteista kollokaateista, ja se myös välittyy merkittävänä seikkana lukijalle.

5 PÄÄTÄNTÖ

5.1 Yhteenveto ja pohdinta

Tässä tutkimuksessa olen käsitellyt Ylen verkkouutiskorpuksesta saamiani kollokaatteja sekä niihin liittyviä diskurssiprosodioita. Ensimmäinen tutkimuskysymykseni liittyi *diabetes*-termin saamiin kollokaatteihin, joita AntConc-ohjelmistosta tuli 368. Näistä tarkastelin tarkemmin 100 merkitsevintä poistaen kuitenkin väleistä merkityksettömiä sanoja. Kollokaateista ilmeni, että uutisoinnissa diabeteksesta puhutaan todella neutraalisti lääketieteen ja tutkimuksen yhteydessä lähinnä rinnastaen diabetesta muihin sairauksiin ja esittelemällä sairauden ominaisuuksia. Toinen tutkimuskysymys keskittyi diskurssiprosodioiden muodostamiseen, joita saatujen kollokaattien pohjalta rakentui kolme. Nämä diskurssiprosodiat osoittivat, että uutisointi diabeteksesta on asiallista ja tosifaktoihin pohjautuvaa. Toki joukkoon mahtui pohtivaa ja arvailevaakin uutisointia, joka osaltaan saattaa vaikuttaa lukijoiden ajatuksiin sairauden ehkäisemisestä. Tutkimuskysymyksillä ja saaduilla tuloksilla voin todeta, ettei diabeteksesta uutisoidessa Ylen artikkeleissa ilmene yllättäviä seikkoja. Kielelliset ilmaisut ovat tyyllillisesti asiallisia, informoivia ja kattavia selvennyksiä sairaudesta ja sen muodoista. Bednarekin tutkimusartikkelissa (2020) tuodaan esille, kuinka tietyn henkilöryhmän diabeteksen sairastamisesta uutisoidaan todella negatiivisesti. Negatiivisesta uutisoinnista lukijoille syntyy stereotyyppisiä sekä stigmoja tietystä sairaudesta ja sitä sairastavista. Toisaalta esille tuodaan mahdollisuus tietoisuuden sekä tutkimuksen lisäämisestä. (Bednarek 2020: 18.) Näitä samoja seikkoja pyörii varmasti myös suomalaisessa uutisoinnissa liittyen yleisemmällä tasolla suomalaisten diabeteksen sairastamiseen. Vaikka negatiivista sävyä ei suuresti tuloksissani ilmennytkään, saattaa silti esimerkiksi moniin muihin sairauksiin yhdistäminen

pelästyttää lukijan. Kuitenkin tuloksista saatu tieto kertoo lukijalle myös, kuinka diabeteksen parantamisen ja ehkäisemisen eteen tehdään töitä. Vaikka teksti on suurimmalta osin neutraalia, tuodaan myös ikäviä turhautumista ja haastavuutta kuvaavia seikkoja esiin. Lisäksi esimerkeistä nousee erinäisiä sairastumiseen liittyviä epäilyjä, joilla ei välttämättä ole tieteellistä varmuutta. Nämä edellä mainitut seikat voivat vaikuttaa vahvasti uutisten lukijoiden kuvaan ja käsityksiin diabeteksestä sairautena. Sairaus ja sairastaminen kattaa elämässä monta osa-aluetta ja myös se näkyy läpi tulosten. Asia koskettaa monia ja tulee esiin monissa eri asiayhteyksissä. On myös otettava huomioon, että uutistekstin tavoitteena on olla neutraalia, varmaa ja spesifiä (Saukkonen 2012: 249). Siinä mielessä tutkimukseni tulokset ovat siis positiivisia, sillä tulokset ovat uutistekstin genren mukaisia. Tutkimukseni tavoitteena oli selvittää, kuinka diabeteksestä puhutaan ja uutisoidaan Ylen verkkouutisissa ja millaista kuvaa tehdyt kielelliset valinnat välittävät lukijalle. Muodostuneet diskurssiprosodiat antavat kattavasti informaatiota lukijalle eri aihepiireistä, ja diabeteksestä muodostuu selvä käsitys. Tutkimuskysymykseni sekä tekemäni haku Korpista olivat melko yksinkertaiset. Aineiston kanssa ei ollut suurempia ongelmia, vaikka toki analyysia tehdessä kyseessä on aina tutkijan omat huomiot ja johtopäätökset ja joku toinen saattaisi tehdä samasta yhteydestä eriävän tulkinnan.

Saamistani tuloksista on todettavissa, että diabeteksestä puhuttaessa Ylen verkkouutisissa kielenkäyttö on melko neutraalia, eikä diskurssiprosodioissa ilmene vahvoja sävyjä, kuten esimerkiksi Jantusen tutkimuksissa (2018) on ilmennyt. Tämä tietysti johtuu suoraviivaisesti siitä, että tässä tutkimuksessa kyseessä on kansallinen uutissivusto, jossa tekstien kuuluukin olla asiallisia ilman persoonallisia sävyjä. Jantusen tutkimuksessa on noussut esiin vahvasti esimerkiksi vihapuheen ja henkilöiden diskurssiprosodiat seksuaalivähemmistöistä puhuttaessa (Jantunen 2018b: 37), jollaisia taas ei tässä yhteydessä noussut esiin. Tekstistä ei näe yksittäisten henkilöiden mielipidettä tai kantaaottavaa uutisointia, vaan asiallinen tyyli kantaa läpi esimerkkien.

Diabeteksen rinnastaminen jatkuvasti muihin sairauksiin on aineiston keskiössä, joka puoltaa sen paikkaa merkittävänä tekijänä osana suomalaista sekä kansainvälistä tautikirjoa. Tutkimus ja sen tavoitteet onnistuivat, ja tuloksista ilmeni selkeästi eri diskurssiprosodiat. Vaikka vahvoja, yllättäviäkin seikkoja ei ilmennyt, saatiin silti tutkimuksessa selville diabeteksen uutisoinnissa käytettäviä ilmaisuja. Näin ollen tutkimus onnistui, eikä ongelmia suuremmin ilmennyt. Tulosten yleistettävyyttä ja hyödynnettävyyttä tarkastellessa voidaan todeta, että vaikka korpuslingvistiikka määrällisesti koetaan helposti yleistettävänä ja tulokset sopivina vertailukohtina muille tutkimuksille, on tämä tutkimus vain katsaus pieneen suomalaiseen uutisaineistoon. Jos tutkimuksessani olisi ollut suurempi korpus tai siitä olisi tutkittu kollokaatteja laajemmin, eikä vain esimerkiksi sataa merkitsevintä, olisivat tulokset

voineet olla erilaiset. Kieli on hetkellistä ja tietyssä kontekstissa käytettyä, eikä siis varmasti ikinä tule olemaan kokonaisuudessaan yleistettävissä kaikkeen kielenkäyttöön.

5.2 Mahdollisuudet jatkotutkimuksiin

Korpusavusteisen diskurssintutkimuksen kenttä on verrattain varsin tuore, joten uudelle tutkimukselle on tilaa. Jantunen (2018) sekä Johansson, Jantunen, Heimo, Ahonen sekä Laippala (2018) ovat tutkimusartikkeleissaan käsitelleet Suomi24-keskustelupalstan kielellisiä ilmiöitä. Artikkelit pohjaavat vahvasti tavallisten henkilöiden tekemiin kielellisiin valintoihin vuorovaikutteisessa mediakeskustelussa. Kielenkäyttö keskustelupalstoilla on arkista ja epämuodollista, mikä eroaa suuresti uutisartikkeleiden virallisesta kielenkäytöstä. Mielekästä olisikin tehdä tutkimusta vertaillen tuloksia tästä tekemästani katsauksesta Ylen uutisartikkeleihin sekä Suomi24-keskustelupalstan analyysin tuloksista. Kaikki diskurssianalyysi ja sen jatkeena myös korpusavusteinen diskurssintutkimus on kilpailevaa. Tiettyjä diskurssiryhmien ominaisuuksia on mahdollista paljastaa ja arvioida ainoastaan vertailemalla niitä toisiinsa. (Partington, Duguid & Taylor 2013: 12.) Tämän tutkimuksen tuloksia ja menetelmiä voisi siis hyvin käyttää esimerkiksi toisen tutkimuksen rinnalla vertailuun. Keskustelupalstan kielenkäyttöä olisi mielenkiintoista tutkia, sillä sellaisessa kontekstissa tästä tutkimuksesta puuttuvat sävyerot tulisivat varmasti esiin. Tavallisten ihmisten kielenkäyttö aiheesta informaatioketjun toisessa päässä tulisi näin tutuksi ja kielipiirteiden välillä voisi tehdä vertailua. Myös sairauden uutisointiin keskittyvää tutkimusta voisi jatkaa muiden kansansairauksien pariin, jolloin saataisiin kattavampi käsitys perussairauksien uutisoinnista ja kansalaisille informoinnista. Myös tämänkaltaista tutkimusta voisi vertailla tekemääni tutkimukseen ja näin rakentaa kattavaa kielikäsitystä uutisten kielenkäytöstä. Lisäksi esimerkiksi Bednarekin (2020) tutkimuksen kaltaista rajausta voisi soveltaa suomalaiseen yhteiskuntaan ja käsitellä esimerkiksi tiettyjen ihmisryhmien sairastamiseen keskittyvää uutisointia.

LÄHTEET

- Ainiala, Terhi & Jantunen, Jarmo Harri 2019: Korpusonomastinen tutkimus slanginimistä Hesa ja Stadi digitaalisissa diskursseissa. – *Sananjalka* 61 (61) s. 57–79.
- Annunen, Anni 2021: *Seksuaalivähemmistöt Suomen Ja Venäjän Mediassa: Korpusavusteinen Diskurssintutkimus Seksuaalivähemmistöjen Käsittelystä Uutismedioissa*. Tampere University.
- Baker, Paul 2010: *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Bednarek, Monika 2020: *Invisible or high-risk: Computer-assisted discourse analysis of references to Aboriginal and Torres Strait Islander people(s) and issues in a newspaper corpus about diabetes*. PLoS ONE 15(6): e0234486.– <https://doi.org/10.1371/journal.pone.0234486>
- Diabetes Care, American diabetes association 2021: *About Diabetes Care*. – <https://care.diabetesjournals.org/content/about-diabetes-care> 16.3.2021.
- Diabetesliitto 2021: *Yleistä diabeteksesta*. – https://www.diabetes.fi/diabetes/yleista_diabeteksesta 1.4.2021.
- Fairclough, Norman; Blom, Virpi & Hazard, Kaarina 1997: *Miten media puhuu*. Tampere: Vastapaino.
- Fairclough, Norman 2003: *Analysing discourse: Textual analysis for social research*. London, New York: Routledge.
- Heikkinen, Vesa 2012: Diskurssi. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käsikirja* s. 94–99. Helsinki: Gaudeamus.
- Hoosen, Sarah & Collins, Anthony 2004: Sex, Sexuality and Sickness: Discourses of Gender and HIV/ AIDS among Kwazulu-Natal Women. *South African Journal of Psychology*. 34(3):487-505.
- Jantunen, Jarmo Harri 2009: "Minulla on aivan paljon rahaa": Fraseologiset yksiköt suomen kielen opetuksessa. – *Virittäjä* 113 (3) s. 356–381.
- Jantunen, Jarmo Harri 2018a: Homot ja heterot Suomi24:ssä: analyysi digitaalisista diskursseista. *Puhe ja kieli*, 38 (1), 3–22. doi:10.23997/pk.65488
- 2018b: Korpusavusteinen diskurssintutkimus (CADS): analyysiesimerkki homouden ja heterouden digitaalisista diskursseista. *AFinLA-E: Soveltavan Kielitieteen Tutkimuksia* 11 (2018) s. 20–44.
- Johansson, Marjut; Jantunen, Jarmo H.; Heimo, Anne; Ahonen, Mirka & Laippala, Veronika 2018: Verkkokeskustelujen kansa: Korpusavusteinen diskurssianalyysi Suomi24-keskustelupalstasta. – *Sananjalka* 60 (60) s. 96–117.

- Kielipankki 2020: Työkälyt. – <https://www.kielipankki.fi/tyokalut/> 26.2.2021.
- Kielitoimiston sanakirja. 2020. Helsinki: Kotimaisten kielten keskuksen verkkojulkaisuja 35. URN:NBN:fi:kotus-201433. <URL <https://www.kielitoimistonsanakirja.fi>>. Päivitettyjulkaisu. Päivitetty 4.5.2021.
- Kunnallisalan kehittämissäätö 2019: *Yle ja STT luotettavimmat uutisoijat*. – <https://kaks.fi/uutiset/yle-ja-stt-luotettavimmat-uutisoijat/> 21.1.2021
- Lounela, Mikko & Heikkinen, Vesa 2012: Korpus. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käsikirja* s. 120–127. Helsinki: Gaudeamus.
- Luodonpää-Manni, Milla; Hamunen, Markus; Konstenius, Reetta Alexandra; Miestamo, Matti; Nikanne, Urpo; Sinnemäki, Kaius & Ainiala, Terhi 2020: *Kielitutkimuksen menetelmiä: I-IV*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Luoma, Elina 2015: Kerro kaikki mitä tiedät diabeteksestä. – *Inspis*. – https://www.diabetes.fi/inspis/diabeteksen_hoitaminen_ruoka/kerro_kaikki_mita_tiedat_diabeteksesta
- McGannon, Kerry; Berry, Tanya; Rodgers, Wendy & Spence, John 2016: Breast cancer representations in Canadian news media: a critical discourse analysis of meanings and the implications for identity. *Qualitative Research in Psychology*. 13:2, 188–207, DOI: 10.1080/14780887.2016.1145774
- Nevalainen, Terttu & Rissanen, Matti 2012: Englanninkielisten standardikorpusten tekstilajeista: Brownin ja Helsingin korpusperheet. Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käsikirja* s. 320–324. Helsinki: Gaudeamus.
- Paju, Päivi 2006: *Megamässäilyllä Elintapasairaaksi: Vastuun Ja Syyllisyyden Rakentuminen Tyypin 2 Diabeteksen Ehkäisyä Käsittelevissä Sanomalehtiutisissa*. Pro gradu -tutkielma. Tampereen yliopisto.
- Partington, Alan; Duguid, Alison & Taylor, Charlotte 2013: *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam: John Benjamins Publishing Company.
- Pietikäinen, Sari & Mäntynen, Anne 2019: *Uusi Kurssi Kohti Diskurssia*. Tampere: Vastapaino.
- Pöllänen, Satu & Vartiainen, Marika 2006: *Tyypin 2 Diabetes Sanomalehti Keski-suomalaisessa Vuonna 2004*. Pro gradu -tutkielma. Jyväskylän yliopiston terveystieteiden laitos.
- Saukkonen, Pauli 2012: Tekstilajien systematiikkaa. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käsikirja* s. 240–254. Helsinki: Gaudeamus.
- Sinclair, John 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, Michael 2001: *Words and phrases: Corpus studies of lexical semantics*. Malden, MA: Blackwell Publishers.
- Tyrkkö, Jukka & Taavitsainen, Irma 2012: Laajojen tekstimassojen luokitteluperusteita. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käsikirja* s. 309–319. Helsinki: Gaudeamus.

- VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. – <http://scripta.kotus.fi/visk> 23.3.2021.
- World Health Organization 2021: *Diabetes*. – https://www.who.int/health-topics/diabetes#tab=tab_1 1.4.2021.
- Yle 2020: *Usein kysyttyä: Yle pähkinänkuoressa*. – <https://yle.fi/aihe/artikkeli/yleisradio/yle-pahkinankuoressa> 21.1.2021.
- Yleisradio. *Ylen suomenkielinen uutisarkisto 2011–2018, sekoitettu, Korp* [tekstikorpus]. Kielipankki. –<http://urn.fi/urn:nbn:fi:lb-2019121004>.
- Yule, George: 1996: *The study of language* (2nd ed.). Cambridge: Cambridge University Press.

LIITTEET

LIITE 1 SATA MERKITSEVINTÄ KOLLOKAATTIA TERMILLE *DIABETES*

kollokaatti	frekvenssi	MI-luku
sairastava	22	4.15062
lapsuusikä	22	4.13612
nuoruustyyppi	57	4.11726
nuoruusikä	13	4.05520
iäkäs	14	4.02461
ympäristömyrkky	10	3.99862
ehkäisy	61	3.99272
kakkostyyppi	376	3.96918
esiaste	22	3.96620
aikuistyyppi	216	3.96397
tytär	12	3.93972
seuloa	12	3.93972
ykköstyyppi	203	3.92417
tyyppi	335	3.87983
rokote	17	3.85726
care	34	3.85726
aikuisikä	65	3.84108
ykkönen	10	3.82869
huolestuttava	10	3.82869
sairastaa	324	3.81687
suolistosairaus	11	3.81419
puhjeta	80	3.80879
herkkä	13	3.79217
keliakia	18	3.75915
sokeritauti	19	3.75469
raskausaika	19	3.75469
seuraus	20	3.75069
yleistyä	66	3.74380
synty	25	3.73558
astma	50	3.73558
verenpainetauti	83	3.72980
sairastavi	33	3.72108
välttyä	11	3.67669
varsinainen	12	3.67669
tyypillinen	10	3.67669
säästyä	11	3.67669

sydäntauti	30	3.67669
silmä	12	3.67669
reuma	23	3.67669
paljastaa	11	3.67669
munuaistauti	11	3.67669
joku (poistettu)	11	3.67669
heikentää	14	3.67669
epilepsia	20	3.67669
ennustaa	19	3.67669
ennaltaehkäisy	16	3.67669
eniten	12	3.67669
aikuinen	47	3.64632
sydänsairaus	42	3.64274
kansantauti	28	3.62606
liikalihavuus	23	3.61529
allergia	40	3.60630
liitännäissairaus	18	3.59869
kehittyä	68	3.59423
verisuonitauti	78	3.58705
eli (poistettu)	85	3.57829
ehkä (poistettu)	14	3.57715
lehti	40	3.57235
Parkinsonin tauti	13	3.56977
kolesteroliongelma	13	3.56977
hoito	202	3.56672
sepelvaltimotauti	37	3.56421
johtua	37	3.56421
verisuonisairaus	48	3.56121
vaikkapa (poistettu)	12	3.56121
vaikea	12	3.56121
pitkäaikaissairaus	12	3.56121
perimä	12	3.56121
äiti	46	3.55639
hoitaa	91	3.55513
lisäsairaus	11	3.55116
paitsi (poistettu)	10	3.53919
potea	19	3.53230
sairastua	362	3.52548
erityisesti (poistettu)	25	3.51319
altistaa	63	3.50422
todennäköinen	38	3.49835

aivohalvaus	22	3.49226
liittyä	108	3.48906
riski	282	3.48532
vähän	14	3.48404
tuntea	14	3.48404
Alzheimerin tauti	28	3.48404
geeni	20	3.47505
puoli	25	3.46256
lisääntyä	62	3.46096
syöpä	112	3.45062
julkaista	29	3.44721
vaatia	17	3.44222
verenpaine	101	3.44008
merkitys	16	3.42876
maailma	48	3.42876
kansansairaus	16	3.42876
löytää	21	3.42515
mahdollinen	26	3.42293
tausta	20	3.41365
parantaa	30	3.41365
ilmetä	10	3.41365
hiiri	10	3.41365
tupakointi	39	3.40750