

**”Olen pahoillani ja huolissaan” – Tutkimus persoonakongruenssista
olotilanilmausidiomeissa Suomi24-korpuksessa 2001–2017**

Markus Mattila, 22414

Pro gradu -avhandling i finska språket

Handledare: Leena Maria Heikkola

Fakulteten för humaniora, psykologi och

teologi

Åbo Akademi

2019

ÅBO AKADEMI – FAKULTETEN FÖR HUMANIORA, PSYKOLOGI OCH TEOLOGI

Abstrakt för avhandling pro gradu

Ämne: Finska språket	
Författare: Markus Mattila	
Arbetets titel: ”Olen pahoillani ja huolissaan” - En studie om personkongruens vid idiomatiska uttryck som uttrycker tillstånd med hjälp av Kielipankkis Suomi24-korpus	
Handledare: Leena Maria Heikkola	
<p>Abstrakt:</p> <p>Det sker hela tiden kontinuerlig och oavbruten förändring i mänskligt språk. Man kan försöka styra språkförändringen men det är omöjligt att stoppa eller ens kontrollera den fullt ut. I denna pro gradu-avhandling undersöker jag språkförändring i det finska språket med närmare fokus på idiom som beskriver tillstånd och där det enligt rekommendationer bör finnas personkongruens mellan satssubjektet och det obligatoriska possessivsuffixet som hör till nomendelen av idiomet.</p> <p>Konstruktionsgrammatikerna anser att grammatiken består av olika slags konstruktioner som är etablerade kombinationer av form och betydelse. Den konstruktionstypen som är den viktigaste för denna avhandling är idiomkonstruktionen. Idiom är konstruktioner som inte är genomskinliga till sin betydelse. Verbfraasideom som består av ett verb och ett komplement är den klart vanligaste idiomtypen i finskan.</p> <p>Kongruens betyder ett samband mellan ord och fraser i en konstruktion. De måste anpassa sig till varandra så att de bilda en grammatiskt enhetlig helhet. Man böjer olika komponenter i en konstruktion i kasus, numerus och/eller person och på det sättet ökar deras kohesion och visar att de hör ihop. Possessivsuffixet är ett suffix som placeras efter ett substantiv för att ange vem dess ägare är. Korpuslingvistik är en metod där man använder korpusar för att analysera språk. En korpus är en stor och systematiskt samlad kollektion texter eller tal som man har samlat för att studera olika fenomen inom språket och att göra lingvistisk analys.</p> <p>Denna studie är en korpusstudie och som material har jag valt korpusen Suomi24 virkkeit -korpus (2017H2) (beta) som är en av Kielipankkis (Språkbanken) korpusar. Korpusökningarna är gjorda med Kielipankkis webbaserade konkordansverktyg KORP och sökresultaten bearbetades med Microsoft Excel. Den statistiska analysen gjorde jag med SPSS. Avhandlingens syfte var att studera huruvida tre idiomkonstruktioners rekommendationsstridiga användning med tanke på possessivsuffix har ökat under de sju åren som Suomi24-korpusen omfattar. Andra forskningsfrågan var huruvida det finns skillnader på det fenomenet mellan olika idiom.</p> <p>Denna studie visar att icke-rekommendationsenlig användning av idiom av denna typ är på en relativt låg nivå. Det är rätt sällsynt att man använder dessa idiom utan korrekt böjt possessivsuffix. Fast korpusens 17 år är en mycket kort tid med tanke på språkförändring finns det statistiskt signifikanta förändringar i användningen av två idiomerna så det verkar ha skett små ökningarna beträffande detta. Dessutom finns det skillnader mellan idiom med tanke på hur ofta det korrekta possessivsuffixet blir ersatt av en rekommendationsstridig form.</p>	
Nyckelord: Idiom, personkongruens, possessivsuffix, korpusforskning	
Datum: 24.6.2019	Sidoantal: 66

Sisällys

Abstrakti	i
Sisällys.....	ii
1 Johdanto.....	1
2 Tutkimuksen teoreettinen tausta.....	2
2.1 Konstruktiot ja konstruktiokielioppi	3
2.2 Idiomit.....	7
2.3 Kongruenssi ja omistusliitteet.....	12
2.4 Kielen variaatio ja muutos	16
2.5 Korpustutkimuksista	20
2.5.1 Korpustutkimuksen työkaluja	22
2.5.2 Aiempaa korpustutkimusta idiomeista.....	24
3 Tutkimus.....	25
3.1 Aineisto ja metodit	25
3.2 Tutkimuksen toteutus ja tulokset	28
3.2.1 Pilottitutkimus idiomien yleisyydestä.....	28
3.2.2 Pääitutkimus kolmen idiomien persoonakongruenssista	31
3.2.2.1 Hakulausekkeet.....	33
3.2.2.2 Idiomien muodostamisessa käytetyt verbit.....	34
3.2.2.3 Hakutulosten läpikäynti	36
3.2.2.4 Tutkimuksen tulokset.....	40
4 Pohdinta.....	45
5 Loppupäätelmät	52
Lähteet	57
Svensk sammanfattning – Ruotsinkielinen tiivistelmä.....	60

1 Johdanto

Kuten kaikki muukin elämässä, ihmiskielet käyvät läpi jatkuvaa muutosprosessia tulematta koskaan valmiiksi. Ainoat kielet, jotka eivät muutu – näkökulmasta riippuen kehity tai rappeudu – ovat niin kutsutut kuolleet kielet, joilla ei ole äidinkieliä puhujia. Kielenmuutosta voidaan yrittää ohjata, mutta täysin hallita tai pysäyttää sitä ei voi ja erilaisia muutoksia tapahtuu koko ajan: osa nopeammin ja osa hitaammin.

Kielenmuutoksesta huolissaan olevia on ollut aina ja mitä luultavimmin tulee aina olemaan. Se, pitääkö siitä olla huolissaan vai innoissaan, on jossain määrin makuasia samaten kuin sekin, missä määrin kieltä pitää huoltaa ja normittaa. Toisaalta hyvin tiukka normittaminen ja säätely voi johtaa tilanteeseen, jossa kirjoitettu kieli ja puhuttu kieli muistuttavat toisiaan varsin vähän, mikä tekee kielen opiskelusta haastavaa myös sitä ensimmäisenä kielenään opiskeleville. Toisaalta valtakieliä opiskellaan ympäri maailman monesti ilman natiiviopettajia, mikä luo tarpeen selkeille säännöille ja ohjeistukselle.

Nykymaailmassa, kun tuotettu kieli on vastaanottajan käytettävissä käytännössä heti vaikka toisella puolella maapalloa, eri ilmiöt leviävät nopeasti. Toisaalta valtava informaatiotulva ja valtakielten jatkuva läsnäolo tuottavat vaikutteita pienempiin kieliin lainasanojen, mutta myös sanontojen, idiomien ja jossain määrin jopa uusien tai ainakin muuttuvien rakenteiden – esimerkiksi omistuksen ilmaiseminen suomessa – muodossa. Nuorisokielessä vilisee muualta omaksuttuja sanoja kuten vaikkapa *feilata*, *stalkata* tai *rageta*. Toisaalta heidän vanhempansa antavat sujuvasti *feedbackia* ja *tsekkaavat mailinsa serveriltä*, vaikka kaikille sanoille on myös kelpo vastineet suomen kielessä. Heidän vanhemmistaan tai isovanhemmistaankaan tuskin kukaan tosissaan puhuu enää näköradiosta, ja farmarihousutkin tunnetaan tänä päivänä pääasiassa farkkuina.

Tässä tutkielmassa tutkin kielen muutosta suomen kielen osalta keskittyen tarkemmin ottaen olotilanilmausidiomeihin ja niiden pakollisen omistusliitteen kongruenssiin lauseen subjektin kanssa esimerkiksi ilmauksessa *olla huolissaan*. Tutkimusmenetelmänä on korpustutkimus, joka on omiaan mahdollistamaan tällaisen hieman harvinaisemman ilmiön tutkimisen. Koska kielenmuutos tapahtuu kontrolloidumman ja stabiilimman kirjoitetun kielen sijaan pääasiassa puhekielessä, olen valinnut tutkittavakseni Kielipankin Suomi24-korpuksen, joka käsittää Suomi24-keskustelupalstalla vuosina

2001–2017 käydyt keskustelut: epäviralliset ja nimimerkin takaa käydyt keskustelut muistuttavat puhekieltä huomattavasti enemmän kuin tekstit virallisissa dokumenteissa, uutisissa tai kirjallisuudessa, jotka oikoluetaan ja joita viilataan joskus pitkäänkin ennen julkaisemista.

Tutkielma käsittelee siis omistusliitteellisten olotilanilmausidiomien mukautumista subjektin persoonaan virkkeissä, joissa subjekti esiintyy yksikön ensimmäisessä persoonassa. Omistusliitettä ja sen katoamista puhekielestä on käsitelty kielitieteellisissä tutkimuksissa ja kielen ammattilaisten artikkeleissa jo pitkään ja runsaastikin. Omistusliitteen leksikaalistumista kolmanteen persoonaan on niin ikään havainnoitu ja sen eri ilmenemismuodoista keskustellaan muun muassa Isossa suomen kieliopissa (2004), mutta ilmiön etenemisestä suomen kielessä on vielä niukalti tutkimusta. Tämä tutkimus pyrkii täyttämään tuota aukkoa suomen kielen tutkimuksessa.

Suosituksen mukaan yksikön ensimmäisessä – kuten myös toisessa sekä monikon ensimmäisessä ja toisessa persoonassa – esiintyvän subjektin kanssa olotilanilmausidiomien omistusliitteen pitää mukautua esimerkiksi *olen huolissani* vs. **olen huolissaan*. Valitsin tutkittaviksi idiomeiksi yksinkertaisen pilottitutkimuksen perusteella kolme yleistä omistusliitteen sisältävää olotilanilmausidiomia *olla huolissaan*, *olla pahoillaan* ja *olla innoissaan* ja kartoitin kongruoivan ja kongruoimattoman omistusliitteen välistä suhdetta, siinä tapahtuvaa muutosta ja mahdollisia eroja eri idiomien välillä. Pilottitutkimusta käsitellään luvussa 3.2.1 ja päätutkimusta luvusta 3.2.2 eteenpäin.

Lukujen tasolla tätä johdantoa seuraa ensiksi tutkimuksen teoreettinen tausta, jonka jälkeen käsitellään itse tutkimus. Niiden antia puntaroidaan pohdinnassa ja koko työ vedetään yhteen loppupäätelmissä. Lukujen sisäistä rakennetta ja alalukujen mahdollisia alalukuja alustetaan lukujen ja alalukujen alussa.

2 Tutkimuksen teoreettinen tausta

Ennen varsinaisen tutkimuksen esittelyä käsittelem teoreettisemmin konstruktiokielioppia ja konstruktioista lähemmin idiomikonstruktioita sekä tutkimukselle hyvin keskeisiä kieliopillisia rakenteita: omistusliitteitä ja kongruenssia. Näiden lisäksi teen katsauksen

kielen variaatioon ja muutokseen sekä esittelen lyhyesti korpustutkimusta kielentutkimuksen metodina.

Konstruktiokielioppi on valittu lähtökohdaksi, koska idiomeja lähestytään rakenne edellä ja siihen juuri konstruktiokielioppi tarjoaa erinomaiset eväät, kuten seuraavassa alaluvussa esitetään. Idiomikonstruktioita, omistusliitteitä ja kongruenssia pohjustetaan teoriaosiossa, koska ne ovat aivan tutkimuskysymysten ytimessä. Tutkimuskysymykset liittyvät tämän hetken tilanteen lisäksi myös tutkittavan ilmiön kehitykseen ja näin ollen kielen muutokseen tietyssä variantissa ja korpustutkimuksella kerätään data, jonka pohjalta tutkimuskysymyksiin vastataan.

2.1 Konstruktio ja konstruktiokielioppi

Tässä alaluvussa käsitellään konstruktiokielioppia lähtien liikkeelle käsitteen *konstruktio* määrittelemisestä ja erilaisten konstruktioesimerkkien esittämisestä. Konstruktiokielioppia ja sen historiaa esitellään myös lyhyesti ja vertaillaan kognitiivisen ja konstruktiokieliopin toisistaan poikkeavia lähestymistapoja ja niiden erilaisia vahvuuksia.

Adele Goldberg (viitattu Leino 2010: 91) muotoili vuonna 1998 klassisen määritelmänsä konstruktioista:

”C on konstruktio $\text{joss}_{\text{def}} C$ [jos ja vain jos C] on sellainen muodon ja funktion yhteenliittymä, jossa jokin muodon tai funktion piirre ei ole suoraan ennustettavissa C:n rakenneosista.”

Toisin sanoen konstruktioita ovat toisaalta pienimmillään kielen komponentit, joilla on oma muoto ja funktio. Toisaalta isomman konstruktion pitää – ollakseen konstruktio – olla enemmän kuin sen muodostavien pienempien konstruktioiden summa. Tämä jälkimmäinen konstruktion vaatimus sopii erinomaisesti myös tässä työssä pääasiallisesti käsittelemääni konstruktiotyyppiin eli idiomeihin.

Vuonna 2005 Goldberg (viitattu Leino 2010: 94) esitti konstruktiolle uuden, tarkennetun ja laajennetun määritelmän:

”Mikä tahansa kielellinen kokonaisuus katsotaan konstruktioksi, kunhan jokin sen muodon tai funktion piirre ei ole suoraan ennustettavissa sen rakenneosista tai

muista konstruktioista, joiden on todettu olevan olemassa. Lisäksi kokonaisuudet tallennetaan konstruktioina, vaikka ne olisivatkin täysin ennustettavissa, mikäli ne esiintyvät riittävän tiheästi.”

Tässä määritelmässä Goldberg siis lisää konstruktion ominaisuuksiin sen vakiintuneisuuden eli esiintymisen ”riittävän tiheästi”. Ominaisuus on tässä määritelmässä jopa niin tärkeä, että se sivuuttaa ennustamattomuuden vaatimuksen. Tämä laajentaa konstruktion tarkoittamien rakenteiden joukkoa huomattavasti merkitykseltään läpinäkyvien kielen ilmausten suuntaan.

Kotilainen ja Pekkarinen (2002: 584) määrittelevät konstruktiot konventionaalisiksi muodon ja merkityksen yhteenliittymiksi. Heidän mukaansa ”Ne ovat kokonaisuuksia, joita kielenkäyttäjät sekä käyttävät että tulkitsevat sellaisenaan rakentamatta niiden merkitystä osien merkitysten summana”. Konstruktioita voi olla hyvinkin erilaisia ”sijamorfeemeista ja sanoista aina lauseisiin ja sitäkin suurempiin rakenteisiin”. Nämä lauseet ja suuremmat rakenteet muodostuvat pienemmistä konstruktioista eli konstruktioita on usein sisäkkäin useitakin.

Leinonkin (2003: 64) mukaan konstruktio on kielen perusyksikkö, joka sisältää sekä syntaktisia että semanttisia ominaisuuksia, jotka ovat konstruktion merkityksen, käytön ja toisiin konstruktioihin yhdistämisen kannalta molemmat välttämättömiä ja keskenään tärkeydeltään tasa-arvoisia. Leino (2010: 94) listaa konstruktioihin kuuluviin rakenteisiin esimerkiksi ”morfeemit tai sanat, idiomit, osittain leksikaalisesti täytetyt sekä täysin yleiset lausekemallit”

Hoffmann ja Trousdale (2013: 1–2) mainitsevat hekin esimerkkinä konstruktioista tämän tutkimuksen kohteena olevan idiomikonstruktion kuten esimerkiksi *olla innoissaan*. Muita eri konstruktioita, joita Hoffmann ja Trousdale nostavat esimerkeikseen ovat muun muassa sana-, komparatiivi- ja resultatiivikonstruktio. Sanakonstruktio voi olla mikä tahansa sana kuten vaikkapa *hirvi*. Komparatiivikonstruktiossa käytetään adjektiivin komparatiivimuotoa kahden olion vertailuun kuten esimerkiksi *Janne on väsyneempi kuin Mikko*. Resultatiivikonstruktio kertoo tyypillisesti, miten X muuttaa Y:n Z:ksi tekemällä jotain: *Antero hioo lampunjalan sileäksi*.

Leino (2003: 64) kirjoittaa nykyaikaisen kielitieteen pioneerin Ferdinand de Saussuren ajatusmallista, jonka mukaan kaikki kielelliset merkit sisältävät tietyn merkityksen ja muodon. Niillä on toisin sanoen ”semanttinen ja fonologinen napa”, joita ”yhdistää symbolinen suhde”. Ajatus on peruslähtökohtana niin kognitiivisessa kuin konstruktiokieliopissakin: rakenne ja merkitys ovat saman kolikon eri puolet.

Myös Hoffmann ja Trousdale (2013: 1–2) nostavat esiin de Saussuren merkin ja miten se muodon ja merkityksen välisine yhteyksineen on konstruktiokieliopin ytimessä. Yli seitsemänkymmentä vuotta de Saussuren kuoleman jälkeen tutkijat alkoivat tutkia ajatusta, että hänen näkemyksensä muodon ja merkityksen yhteydestä voisi olla hyödyllinen myös laajempien kokonaisuuksien kuin vain sanojen tai morfeemien kuvaamisessa.

Konstruktio on siis Hoffmannin ja Trousdalen (2013: 1–2) mukaan määriteltävissä de Saussuren merkin laajennetuksi versioksi ja tätä ajatusta hyödyntäviä lähestymistapoja kieliopin tutkimukseen kutsutaan nyt konstruktiokieliopiksi. Sen sijaan, että leksikko ja syntaksi olisivat erillisiä tutkimuskohteita, Hoffmann ja Trousdale näkevät konstruktiot osana leksikaalis-syntaktista jatkumoa. Myös William Croft (2001 viitattu Leino 2010: 94) näkee, että ”leksikon ja kieliopin välillä ei ole laadullista rajaa”. Toisin sanoen konstruktiokieliopin keskeisiin ajatuksiin kuuluu se, ettei leksikkoa, morfologiaa ja muuta kielioppia ei ole tarkoituksenmukaista tai edes mahdollista erottaa toisistaan täydellisesti.

Hoffmannin ja Trousdalen (2013: 2) mukaan lähestymistapa, jossa syntaktisia rakenteita käsitellään muodon ja merkityksen liittona, voidaan jäljittää ainakin 1100-luvun keskiaikaisiin kielentutkijoihin. Niin ikään Noam Chomskyn 1950- ja 1960-luvuilla esittämissä malleissa käsiteltiin konstruktioita ja niitä määritteleviä sääntöjä muun muassa verbi–predikatiivi -konstruktioita, komparatiivikonstruktioita ja kausatiivikonstruktioita. Varsinaisen konstruktiokieliopin katsotaan kuitenkin alkaneen vasta 1980-luvulla muiden muassa Fillmoren ja Lakoffin julkaisujen myötä (ks. esim. Fillmore 1985 ja Lakoff 1987).

Konstruktiokieliopin mukaan ”kielioppi koostuu kieleen vakiintuneista muodon ja merkityksen yhteenliittymistä eikä esimerkiksi universaaleilla ja synnynnäisillä perusrakenteosilla tehtävistä operaatioista tai muutossäännöistä” (Leino 2010: 73). Konstruktiokieliopissa keskitytään siis konstruktioihin eli vakiintuneisiin kielenyksiköihin, jotka sisältävät sekä muodon että merkityksen tai hieman eri kannalta tarkastellen kielellisen rakenteen ja funktion.

Kotilainen (2007: 19) listaa konstruktioajattelun peruseriaatteet:

- ”1 Kielioppi koostuu konstruktioista.
- 2 Konstruktioissa yhdistyy tieto muodosta ja merkityksestä.
- 3 Kielioppi on järjestäytynyt.”

Kohta 1 asettaa siis konstruktioit kaiken kielen kuvauksen lähtökohdaksi. Ihmisten kielenkäyttöä eivät ohjaa kieliopilliset säännöt vaan tieto konstruktioista: pystyäkseen kommunikoimaan toisten kanssa kielenkäyttäjän pitää tietää, millaisella muodolla ilmaistaan esimerkiksi jonkun olevan rakennuksen sisällä tai mitä tarkoittaa, kun joku on pahoillaan. Mitä enemmän puhuja tuntee konstruktioita, sitä helpompi hänen on ilmaista itseään ja sitä monivivahteisemmin hän pystyy sen tekemään.

Kotilaisen (2007: 19–20) useita muita konstruktion määritelmiä mukailevan kohdan 2. mukaan konstruktio sisältää tiedon siitä, mitä sillä voidaan ilmaista ja miten sitä tulee käyttää suhteessa muihin konstruktioihin, jotta kielenkäyttäjä tulee ymmärretyksi toisten kielenkäyttäjien keskuudessa. Kohta 3. kertoo, että kieliopin muodostava valtava määrä konstruktioita on hierarkkisessa suhteessa toisiinsa. Asiaa voidaan selventää lausumalla, että jotkin konstruktioit ovat isompien konstruktioiden välttämättömiä osia, joita ilman isompi konstruktio ei täyttäisi tehtäväänsä. Toisaalta osa konstruktioista voidaan nähdä toisten konstruktioiden laajentumina ja kielioppi järjestyy myös ”erilaisten skemaattisuudeltaan poikkeavien konstruktioiden jatkumoksi” (Kotilainen 2007: 20).

Edellä keskustellun mukaisesti sekä konstruktiokieliopissa että kognitiivisessa kieliopissa lähtökohtana pidetään siis sitä, että muoto ja merkitys ovat keskenään erottamattomasti sidoksissa. Leinon (2003: 64–65) mukaan kumpaakaan ei voi täysin ymmärtää ymmärtämättä toista. Näin ollen toisen menestyksekkäs tutkiminen vaatii ainakin jonkin asteista perehtymistä myös toiseen.

Haddington ja Sivonen (2010: 10) kirjoittavat kognitiivisen kieliopin korostavan merkityksiä, joita kaikki muu palvelee. Se painottaa myös semantiikan ensyklopedisuutta eli ”kielellisen tiedon ja muun, maailmaa koskevan tiedon voi erottaa toisistaan vain keinotekoisesti” (Langacker 1987, viitattu Haddington ja Sivonen 2010: 10). Näin ollen kaikki puhujan tieto vaikuttaa sanojen merkityksiin eikä kielellistä merkitystä voi analysoida irrallaan siitä tai hänen kulttuuristaan. Toisin sanoen kognitiivinen kielioppi siis lähestyy kieltä ja sen käyttöä kognition eli kielenkäyttäjän tajunnan kautta.

Leino (2003: 65) kirjoittaa, että siinä missä kognitiivinen kielioppi tarjoaa erinomaiset työkalut kielen semanttisen puolen tarkasteluun, konstruktiokielioppi on omimmillaan rakennetta tutkittaessa ja kuvattaessa, vaikka täysin vain toiseen osioon keskittyminen onkin harvoin jos koskaan tarkoituksenmukaista. Leino (2003: 64) kuvaa konstruktiokielioppia ja kognitiivista kielioppia saman mallin eri puoliksi.

Croft (2001 viitattu Leino 2010: 94) näkee kuitenkin, että konstruktiokieliopin suosio perustuu siihen, että se mahdollistaa koko kieliopin kuvaamisen yhtenäisen mallin mukaan, mikä puolestaan mahdollistaa sen hyödyntämisen monipuolisempien ilmiöiden tutkimuksessa kuin perinteisemmät osajärjestelmistä koostuvat lähestymistavat. Tämän tutkimuksen päähuomio kiinnittyy tiettyjen idiomikonstruktioiden rakenteellisiin seikkoihin eikä niinkään niiden merkityksiin – vaikka niiden merkityksetkin ovat toki selkeästi sukua toisilleen ja sekin on näitä idiomeja yhdistävä tekijä. Näin ollen konstruktiokielioppi on tässä tutkimuksessa kognitiivista kielioppia toimivampi lähestymistapa.

2.2 Idiomit

Tämän tutkimuksen aiheeseen keskeisimmin liittyvä konstruktio tyyppi on jo edeltävässä alaluvussa lyhyesti käsitelty idiomikonstruktio. Tässä alaluvussa määritellään, mitä idiomit ovat ja millaisia eri rakenteita niillä voi olla. Sen lisäksi kartoitetaan erilaisia tapoja, joilla idiomeja kieleen tulee. Viimeiseksi käydään vielä läpi tämän tutkimuksen idiomien tyyppi ja rakenne.

Idiomit ovat konstruktioita, joita ihmiskielessä esiintyy viljalti niin jokapäiväisissä keskusteluissa, uutisissa, tieteellisissä tutkielmissa kuin kaunokirjallisuudessa ja elokuvissakin. Idiomit ovat semanttisesti äärimmäisen mielenkiintoisia, koska esimerkiksi idiomien *saada kuulla kunniansa* olisi helppo ilman tietoa ilmauksen idiomaattisesta merkityksestä kuvitella kertovan vastaanottajan kannalta myönteisestä tapahtumasta. ”Älä käy mun kimppuuni sanoi floristi” on sanaleikki, joka varmasti hämmentää vasta suomen kieltä opiskelevaa, vaikka tämä tunnistaa ja käyttääkin idiomeja sujuvasti äidinkielellään asiaa sen enempää miettimättä.

Nenonen (2002: 8) määrittelee väitöskirjassaan idiomien seuraavasti: ”idiomit ovat epäkompositionaalisia, syntaktisesti ja/tai leksikaalisesti jähmeitä sanaa pitempiä yksiköitä”. Epäkompositionaalisuudella tarkoitetaan sitä, että idiomien merkitystä ei voi täysin varmasti tietää, päätellä tai ennustaa sen osien merkitysten summasta. Näin ollen idiomit ovat myös konventionaalisia: ymmärtääkseen idiomien merkityksen kielenkäyttäjän pitää ainakin jossain määrin tuntea se entuudestaan. Syntaktisella tai leksikaalisella jähmeydellä tarkoitetaan sitä, että toimiakseen idiomikonstruktion pitää noudattaa siihen liittyviä, mutta kieliopista johtumattomia rajoituksia. Yleisesti ottaen idiomit ovat epäitsenäisiä lausekkeita eli ne eivät voi sellaisinaan ilman muita komponentteja muodostaa itsenäisiä lauseita.

Sinclairin (viitattu Nenonen 2002: 14) mukaan sanat esiintyvät teksteissä usein juuri tietyissä ryhmissä kuten kollokaatioissa ja idiomeissa eivätkä tyystin sattumanvaraisessa järjestyksessä. Hän määrittelee kollokaatiot usein yhdessä tavattaviksi kahden tai useamman sanan ryhmiksi. Nenonen (mt.) toteaaakin, että ”kaikki idiomit ovat jossain määrin leksikaalistuneita, mutta kaikki sanaa pitemmät leksikaalistumat eivät siis ole välttämättä idiomeja”. Esimerkkeinä hän mainitsee vakiintuneet kollokaatiot, fraasit ja kliseet, jotka esiintyvät tekstissä selvästi yhteen kuuluvina rykelminä, mutta jotka ovat semanttisesti läpinäkyviä eivätkä näin ollen täytä idiomien määritelmän epäkompositionaalisuusvaatimusta. Esimerkiksi Nenonen (mt.) nostaa semanttisesti läpinäkyvän lauseen *Onkos täällä kilttejä lapsia?*, joka kuitenkin selvästi on kielessä oma yksikkönsä.

Epäkompositionaalisia idiomeja ei siis voida tavattaessa prosessoida täysin samalla tavoin kuin merkitykseltään läpinäkyviä ilmauksia. Näin idiomit ovat kielenkäyttäjälle muita ilmauksia vaativampia, koska niitä kohdattaessa pitää mahdollisesti jossain määrin prosessoida myös niiden osasten kirjaimelliset merkitykset (Nenonen 2002: 14). Tämä kuitenkin tapahtuu yleensä asiaa sen enempää miettimättä ja äärettömän nopeasti, jos kuulija tai lukija vain tuntee ilmauksen ennalta.

Nenosen (2002: 55) mukaan suomen verbilausekeidiomi ”koostuu yleensä finiittiverbistä ja komplementista, joka on taivutettu nomini” kuten esimerkiksi *olla innoissaan*. Hän huomioi myös harvinaisemman muodon, joka muodostetaan subjektista ja finiittiverbistä esimerkiksi *nakki napsahtaa* ja *pipo kiristää*, mutta toteaa, että tämän tyyppiset idiomit ovat monissa kielissä ”hyvin harvinaisia tai jopa mahdottomia”. Verbistä ja komplementista rakentuvat verbilausekeidiomit ovat kuitenkin suomessa selvästi yleisin idiomityyppi. Oman ryhmänsä muodostavat fraasiverbit kuten *antaa ylen* ja *lyödä laimin*, joita voidaan käyttää myös käänteisesti yhdysverbeinä *ylenantaa* ja *laiminlyödä*. Nenosen (2002:55) mukaan idiomien ja fraasiverbien välinen raja on häilyvä ja hän päätyikin väitöskirjassaan käsittelemään fraasiverbejä muiden idiomien joukossa.

Nenosen (2002: 54) väitöskirjatutkimuksen mukaan noin viidesosa idiomeista muodostetaan *olla*-verbin kanssa ja puolet verbilausekeidiomeista sisältää jonkun kymmenestä yleisimmästä idiomeissa esiintyvistä verbistä, jotka ovat *olla*, *ottaa*, *saada*, *mennä*, *pitää*, *vetää*, *tulla*, *tehdä*, *käydä* ja *panna*. Kaiken kaikkiaan Nenonen identifioi noin 300 000 sanetta käsittäneestä korpuksestaan reilut 1 200 erilaista verbilausekeidiomia. Nenosen tutkimukseen palataan tarkemmin tämän työn alaluvussa 2.5.2, jossa sitä käsitellään korpustutkimuksen näkökulmasta. Lisäksi tämän työn pilottitutkimuksen idiomilista on koostettu Nenosen työssään keräämän listan pohjalta. Pilottitutkimus käydään läpi alaluvussa 3.2.1.

Mikone (2000: 25) hahmottelee idiomien kehitysprosessia. Hänen näkemyksensä mukaan idiomit kehittyvät metaforista, jotka vakiintuessaan käyttöön muuttuvat idiomeiksi. Yleistyessään idiomit konventionaalistuvat eli voivat sisältyä kielenkäyttäjien ilmausvarastoon, vaikkeivat nämä tuntisikaan historiaa niiden takana. Paljon käytetyt idiomit voivat Mikosen mukaan myös kulua käytössä, jolloin niiden käyttöalue laajenee

ja rakenne väljenee. Mikone kuvaa näkemyksensä idiomien syntyprosessista seuraavasti: ”LEKSEEMI > METAFORA > IDIOMI > YLEISTYNYT IDIOMI > KIELIOPILLINEN MORFEEMI” (mt).

Häkkinen (2000: 12) esittää kuitenkin Mikosen prosessikuvauksen yleistettävyydestä poikkeavan näkemyksen:

”Idiomit saattavat sisältää kuvauksia ainutkertaisista tapauksista, henkilöistä ja tilanteista, joihin viittaavia ilmauksia on yhteyksistään irrotettuina mahdotonta tulkita alkuperäisellä tavalla mutta jotka silti voivat toimia kielen aineksina niihin liittyneiden konventionaalisten merkitysten ansiosta.”

Esimerkiksi Häkkinen nostaa idiomien *vetää lyhyempi korsi*, joka hänen mukaansa kuuluu suomalaisten ymmärtämien ja käyttämien idiomien joukkoon, vaikkei vanha arvontamenetelmä olisikaan tuttu. Metaforaan perustuvana sitä kuitenkin on hankala pitää. Tämän tutkimuksen idiomit sen sijaan sisältävät kaikki abstraktin käsitteen – *huoli*, *paha* ja *into* – joten niiden historian voi hyvin olettaa palautuvan metaforiin, joissa joku on tavalla tai toisella huoliensa tai intojensa keskellä tai pahojensa äärellä.

Viertiö (2009) keskustelee Häkkisen mainitsemista henkilö pohjaisista idiomeista artikkelissaan idiomista *ottaa ritolat*, joka tarkoittaa paikalta poistumista ja usein hyvinkin ripeästi. Tämä idiomia juontaa juurensa Ville Ritolan voittoisasta loppukiristä Amsterdamissa 1928, kun hän jätti kilpakumppaninsa Paavo Nurmea myöten kuin seisomaan. Tapahtuma oli niin merkittävä, että *ottaa ritolat* -idiomia tavataan vielä nykyäänkin. Suomi24-korpuksen viimeisimmässä eli vuoden 2017 keskustelut sisältävässä alakorpuksessa idiomia tavataan lähes 50 kertaa esimerkiksi ”Soini ottaa ritolat puheenjohtajan pallilta” ja ”Kaverikin otti ritolat ja lähti huitsinnevadaan”.

Johonkin tiettyyn henkilöön ja hänen tekemiinsä perustuva sanonta voidaan muodostaa esimerkiksi verbeillä *tehdä*, *käydä* tai *ottaa* ja jollain tietyllä tavalla toimineen henkilön erisnimellä, jota kuitenkin käytetään yleisnimen tapaan esimerkiksi *tehdä vanhaset*. Viertiön (2009) mukaan useimmat vastaavat sanonnat ovat kuitenkin hyvin lyhytikäisiä eivätkä vakiintuneen *ottaa ritolat* -idiomin tapaan konventionaalistu idiomeiksi. Viertiö kirjoittaa, että lööpeissä usein esiintyvistä henkilöistä muodostetut kielikuvat tarkoittavat ”tänään yhtä ja huomenna jo toista”.

Muikku-Werner (2011) avaa historiaa *tulla maitojunalla* -idiomin taustalla: ” Alun perin kyseessä on ollut jatkosodan aikainen huoltojuna, jonka mukana epäonnistuneet upseerioppilaat palasivat perusyksikköön”. Idiomi siis perustuu oikeasti tapahtuneeseen tilanteeseen. Henkilö- ja tilannepohjaisen *rupesi lyyti kirjoittamaan* -idiomin päähenkilön henkilöllisyys on jäänyt unholaan, mutta se ei ole haitannut sen käyttöä. Ympäri Suomen monta monituista tulostinta on kyseisen idiomien inspiroimina nimetty Lyyteiksi. Omassa idiomissaan esiintyvä myyttinen H. Moilanen on niin ikään nykypuhujan tunnistamattomissa, mikä idiomien vähemmän mairittelevan luonteen vuoksi lienee hänen kannaltaan hyvä asia, jos idiomi ylipäättään perustuu oikeaan historialliseen henkilöön.

Idiomeja myös lainataan/lainautuu toisista kielistä ja kulttuureista sekä taiteesta ja Raamatusta. Muikku-Werner (2011) listaa esimerkiksi *Don Quijotesta* tutun *taistelu tuulimyllyjä vastaan* ja Raamatun *heittää ensimmäinen kivi*. Monesti idiomeja myös kotoutetaan kuten englannin *storm in a teacup* suomalaisessa muodossaan *myrsky vesilasissa*: teen juominen ja siihen käytetty välineistö ei Suomessa ole ollut yhtä merkittävässä asemassa kuin lainatun idiomien alkuperäisessä kotimaassa.

Tämän tutkimuksen käsittelemät idiomit *olla huolissaan*, *olla pahoillaan* ja *olla innoissaan* ovat kaikki merkitykseltään olotilanilmausidiomeja, joilla kuvataan lauseen subjektin henkistä tilaa. Ne ovat verbilausekeidiomeja eli ne sisältävät verbin ja sitä määrittelevän komponentin eli näitten idiomien tapauksessa nominiosan. Tavallisimmassa muodossaan nämä idiomit muodostetaan siis *olla*-verbiä käyttämällä. Sen sijasta saatetaan joissain tapauksissa käyttää myös muuta verbiä, mutta oleminen on kuitenkin idiomien merkityksessä läsnä, vaikkei sitä erikseen lausuttaisi tai kirjoitettaisikaan.

Idiomien nominiosista kaksi *huolissaan* ja *innoissaan* ovat substantiivipohjaisia ja *pahoillaan* on adjektiivipohjainen. Toteuttaakseen funktionsa ja välittääkseen merkityksensä ne seuraavat edellä keskustellun mukaisesti kieliopista johtumattomia rajoituksia eli ne esiintyvät paikallissijoissa – kaksi ensimmäistä inessiivissä ja kolmas adessiivissä – ja sisältävät monikon tunnuksen *i*. Lisäksi kaikkiin näihin idiomeihin sisältyy myös pakollisena komponenttina omistusliite, jonka kielenhuollon suositusten

mukaan pitää mukautua lauseen subjektin mukaan oikeaan muotoonsa. Tästä muotoutumisesta keskustellaan lisää seuraavassa alaluvussa.

2.3 Kongruenssi ja omistusliitteet

Tässä alaluvussa perehdytään yleisesti kongruenssin ja omistusliitteen käsitteisiin, minkä jälkeen pohditaan niiden esiintymistä yhdessä. Tämän jälkeen niistä keskustellaan myös idiomien osana ja osana adverbiaaleja. Niin ikään tutkimuksen keskeisenä teemana olevaa kielenmuutosta sivutaan myös keskittyen sen omistusliitteitä koskeviin vaikutuksiin sekä historiallisesta näkökulmasta että tulevaisuuteen katsoen.

Iso suomen kieliopin (§ 1267) mukaan ”Samaan konstruktion kuuluvien sanojen ja lausekkeiden kieliopillista yhdenmukaisuutta ja toisiinsa mukautumista kutsutaan KONGRUENSSIKSI.” Tuo mukautuminen tarkoittaa käytännössä sitä, että edellä mainittuja konstruktion osasia taivutetaan siten, että niiden muodot ovat yhteneviä sijan, luvun ja/tai persoonan osalta. Kongruenssi liittyy konstruktion osat tiiviimmin toisiinsa ja lisää sen sisäistä koheesiota. Tämän tutkimuksen aiheena oleva persoonakongruenssi idiomeissa ilmenee siten, että idiomien verbiosa taipuu kulloisenkin persoonan mukaiseen muotoon ja sen nominiosa saa niin ikään persoonan mukaisen omistusliitteen: esimerkiksi *olen huolissani, olet huolissasi, hän on huolissaan* ja niin edelleen.

Omistusliitteet eli possessiivisuffiksit ovat ISK:n (§ 95) mukaan ”nomineihin ja muihin nominaalisiin muotoihin kiinnittyviä tunnuksia, jotka ilmaisevat persoonaa”. Ensimmäisen persoonan omistusliitteet ovat *-ni* yksikössä ja *-mme* monikossa ja toisen persoonan vastaavasti *-si* ja *-nne* esimerkiksi *kirjani, kirjamme, kirjasi* ja *kirjanne*. Kolmannessa persoonassa ei tehdä eroa yksikön ja monikon välillä vaan molemmissa mahdolliset suffiksit ovat *-nsA* ja *-Vn* esimerkiksi *hänen/heidän kirjansa* ja *hänen/heidän talossaan*. Omistusliitteen paikka sanassa on sijapäätteen jäljessä, mutta mahdollisen liitepartikkelin edellä esimerkiksi *kirjassaniko* tai *taloannekin*.

Markkula (2016) kirjoittaa omistusliitteen kuuluvan kiinteästi ”moniin tavan, tilan, asennon tai asenteen ilmauksiin” kuten esimerkiksi olotilanilmausidiomeihin. Hänen mukaansa yleiskielessä tuollaisissa ilmauksissa tekijän persoona vaikuttaa omistusliitteen muotoon ja hän nostaaakin yhdeksi esimerkikseen tämän tutkimuksen kohteisiin kuuluvan

idiomin: ”Olin huolissani sinusta.” Kongruoimatonta omistusliitettä koskien Markkula arvelee, että moni mieltää kolmannen persoonan omistusliitteen -Vn tämän tyyppisissä ilmauksissa omistusta ilmaisevan liitteen sijaan päätteeksi, jota käytetään adverbien muodostamisessa ja jättää siksi sen taivuttamatta. Vielä useampi ei varmastikaan juuri pohdi asiaa kieltä käyttäessään.

Omistusliitettä käytetään myös ISK:n mukaan sijapäätteiden ohella tilaa ilmaisevien adverbien muodostamisessa esimerkiksi ”Mutisin hämilläni, etten ollut tuntenut häntä ilman silmälaseja” (§ 96). Esimerkki osoittaa, että olotilanilmausidiomien ja tilaa ilmaisevien adverbien suhde on usein varsin häilyvä: tässäkin tapauksessa asia voitaisiin ilmaista myös: ”Olin hämilläni ja mutisin, etten ollut tuntenut...” Samoin voidaan pohtia omistusliitteen sisältävien idiomien *olla hämillään ja mennä hämilleen* eroja ja yhteyksiä, mutta ne jäävät tämän tutkimuksen rajauksen ulkopuolelle.

ISK mainitsee, että omistusliite ei välttämättä puhekielessä aina kongruoi subjektin kanssa ja antaa esimerkkinä tällaisesta tapauksesta lauseen ”*Mä tein sen kyllä ihan tahallaan*”. Osa tämän tyyppisistä muodoista on leksikaalistunut täydellisesti siten, että ainoastaan kolmannen persoonan omistusliite on mahdollinen. ISK listaa esimerkkeinä adverbit ”*aikaisintaan, edelleen, erikseen, kahdestaan, uudelleen, yhteensä*.” Kieliopin mukaan ei siis ole mahdollista sanoa *Lähdimme kaupunkiin *kahdestamme* tai *luin koko kirjan *uudelleni*. Jälkimmäinen muoto on mahdollinen, mutta se tarkoittaa eri asiaa, kuin *luin koko kirjan uudelleen*.

Taivutusmuotoisia adverbeja käsittelevässä pykälässä 386 ISK listaa lisää adverbeja, jotka tavataan vaihtelemattomasti vain kolmannessa persoonassa: ”*konsanaan, kerrassaan, harvakseltaan, yleensä*”. Subjektin mukaan kongruoiviin muotoihin ISK lukee tässä pykälässä kaksi tämän tutkimuksen aiheena olevista ilmauksista, *huolissaan* ja *pahoillaan*, mutta mainitsee myös puhekielisen, *olen huolissaan* -muotoisen käytön.

ISK (§ 682) listaa muodon *pahoillaan* olotilaa ilmaisevana adjektiivipohjaisena adverbiaalina. Tässä mielessä se eroaa tutkimuksen muista kohdekonstruktioista, jotka ovat substantiivipohjaisia. Pykälässä 1445 ISK käsittelee pakollista possessiivisuffiksia, joka esiintyy vain muodollisesti refleksiivisessä muodossa, toisin sanoen muodossa *olla*

innoissaan, muttei koskaan muodossa **hänen innoissaan*. ISK mainitsee jälleen, että yleiskielessä omistusliite noudattaa korrelaatin eli yleensä subjektin persoonaa, mutta että puhekielessä tavataan kolmanteen persoonaan leksikaalistuneita muotoja.

Markkula (2016) kirjoittaa, että omistusliitteen persoonakongruenssia noudatetaan yleiskieltä vähemmän puhekielen lisäksi myös sosiaalisessa mediassa ja nettikeskusteluissa käytettävässä kielessä. Hänen mukaansa ”Kirjoittaminen yhteisöpalveluissa voi olla niin ripeää, että on joutuisinta käyttää yhtä liitettä kaikkialla.” Lausuma herättää kysymyksen siitä, onko persoonakongruenssin hylkääminen tietoinen valinta ajan säästämiseksi vai onko kyse siitä, ettei sitä osata tai tiedetä käyttää. Toki varmasti jokaiselle suomalaisen peruskoulun käyneelle on teroitettu kyllästymiseen asti, että esimerkiksi ilmaus *me mennään* ei ole hyvää yleiskieltä, mutta silti sitä käytetään niin paljon, että suosituksenmukainen (*me*) *menemme* ei puhekielessä enää kaikissa yhteyksissä oikein kuulosta luontevalta.

Myös kielitoimiston ohjepankin mukaan ”Yleiskielessä kuuluu käyttää omistusliitettä”. Vaikka ohjepankki huomioi ISK:n tapaan osittain kivettyneet 3. persoonan omistusliitteellä muodostettavat muodot kuten *mielellään* ja *kallellaan*, se kertoo, että ”Yleiskielen suositus on kuitenkin persoonaan mukautuminen”. Joko osittain tai kokonaan leksikaalistuneista muodoista ohjepankissa listataan *kallellaan*, *erillään*, *hajallaan*, *irrallaan*, *aikanaan*, *harvakseltaan* ja *hiljalleen*. Lisäksi ainakin *mielellään* ja *tavallaan* mainitaan muotoina, jotka normaalisti mukautuvat subjektin persoonaan mukaan, mutta erityistapauksissa eivät. Tutkimuksen kohdeilmauksia *olla huolissaan*, *olla pahoillaan* ja *olla innoissaan* Kielitoimisto ei listaa leksikaalistuneiden joukkoon vaan suosittaa omistusliitteen vaihtelua persoonaan mukaan.

Omistusliitteen alkuperästä Hakulinen (1979 viitattu Remes 2009: 96) kirjoittaa, että ”omistusliitteisen taivutuksen juuret ulottuvat todennäköisesti kantauraliin asti” eli niin kauas suomen kielen historiallisiin esimuotoihin kuin kielitieteellisesti on mahdollista katsoa. Laanestin (1975 viitattu Remes 2009:96) mukaan suomessa käytetään itämerensuomalaisten kielten joukossa eniten omistusliitteitä. Muut pohjoisen ryhmän kielet käyttävät omistusliitteitä selvästi vähemmän ja eteläisen ryhmän kielissä omistusliitteistä on jäljellä vain rudimentteja, joita tavataan lähinnä adverbeissa.

Monissa eurooppalaisissa kielissä possessiivisuffiksin sijaan käytetään possessiivipronomineja kuten esimerkiksi *palloni: my ball, min boll, mein Ball* ja niin edelleen. Suomen merkittävimmissä sukulaiskielessä virossa omistusta ilmaistaan tänä päivänä germaanistyyppisellä possessiivipronominilla ilman omistusliitettä esimerkiksi *minu/mu pall*. Näin selvästi suomesta poikkeavaa rakennetta selittänee ainakin osittain saksalaisten ritarikuntien keskiaikaisista valloituksista alkanut ja vuosisatoja kestänyt voimakas saksan kielen vaikutus, johon liittyi aina 1800-luvulle kestänyt ankara maaorjuus (ks. esim. Zetterberg 2017: 33 alkaen). Remeksen (2009: 99–100) mukaan viron kirjakielessä omistusliitteistä viestivät vain rudimentit, mutta kansankielessä omistusliitteitä tavataan joissakin murteellisissa adverbeissa, sanonnoissa ja kansanrunoissa. Näihin historiallisiin pohdintoihin palataan kielen variaatiota ja muutosta käsittelevässä alaluvussa 2.4.

Paunonen (1995: 501) kirjoittaa, että omistusliitteiden käytön väheneminen on ollut yksi näkyvimmistä muutoksista puhutussa suomen kielessä viitaten siis pääasiassa *minun pallo* -tyyppisiin muotoihin eikä niinkään suoranaisesti tämän tutkimuksen aiheena oleviin omistusliitteellisiin idiomeihin. Hän lainaa Y. Lumion kirjoitusta Uuden Suomen yleisönosastossa vuonna 1966, jossa tämä kritisoi omistusliitteettömiä muotoja: ”Tämä tällainen kuulostaa lapsen puheelta tai jonkinlaiselta hällävälä-tyyliin puhutulta kotikieleltä”. Lumio oli huolissaan ennen kaikkea televisiosta kuultavissa olevasta kielestä ja siitä, miten se tarjoaa katsojille ”väärää kielellisiä malleja. (mt.)

Paunonen (1995: 502) nimeää urheilutoimittaja Jari Porttilan, joka 1980-luvun puolivälissä oli edesauttamassa omistusliitteettömien tuloa myös sanomalehtikieleen osin tavoitellessaan puhekielistä sävyä, mutta käytti sitä myös yleiskielisessä rekisterissä. Paunonen kirjoittaa omistusliitteettömien muotojen löytäneen tiensä kirjoitusvuoteen mennessä jo Helsingin Sanomien pääkirjoituksiin asti (mt). Hän (1995:514) ennustaakin Helsingin puhekieltä käsittelevässä tutkimuksessaan, että yksikön ensimmäisessä persoonassa omistusliitteellinen muoto katoaisi kyseisestä suomen kielen variantista 2040-luvun aikana.

Korhonen (2016) arvioi kuitenkin, että vaikka omistusliite joissain yhteyksissä ”tuntuu menettäneen persoonavoimansa”, se ei ole katoamassa yleiskielisestä asiatekstistä. Edellä mainitun kaltaisista yhteyksistä hän nostaa esimerkiksi juuri *olla huolissaan* -tyyppiset ilmaukset, jotka vaikuttavat ainakin jossain määrin olevan leksikaalistumassa. Korhonen toteaa kuitenkin ykskantaan, että mukautumaton omistusliite ”ei ole yleiskielen mukainen” (mt).

2.4 Kielen variaatio ja muutos

Tässä alaluvussa käsitellään kielen variaatiota ja kuinka se on avain kielenmuutokseen. Muutosta tapahtuu niin kielensisäisten kuin kielenulkoistenkin syiden vuoksi. Mitä ne ovat olleet aikaisemmin ja mitä ne ovat nyt, kuuluvat myös tämän alaluvun valottamiin asioihin. Kielenmuutoksen luonnetta myönteisenä vai kielteisenä ilmiönä pohditaan myös ja lopuksi käsitellään jo hieman seuraavaa alalukua alustuen kielenmuutoksen tutkimista.

Mustanoja (2011: 51) esittää kolme keskeistä huomiota kielen muutosta käsittelevästä teoriasta:

- 1) ”Elävä kieli varioi ja muuttuu.”
- 2) ”Kaikessa kielen variaatiossa on kielenmuutoksen potentiaali.”
- 3) ”Kuitenkin vain osa kielen variaatiosta johtaa kielenmuutokseen.”

Mustanojan (2011: 57) mukaan variaationtutkijoiden keskuudessa vallitsee konsensus, että kieli muuttuu koko ajan ja muutos vaikuttaa kielen kaikkiin osa-alueisiin. Muutos voi olla ”kiivasta ja näkyvää”, mutta toisaalta myös ”erittäin hidasta ja vaikeasti havaittavaa”. Kielen muutosta voidaan tutkia jo tapahtuneiden muutosten osalta nykytilanteen pohjalta. Käynnissä olevien muutosten tutkimuksessa arvioidaan ja ennustetaan aikaisemman tilanteen ja nykyhetken perusteella myös tulevaisuudessa tapahtuvia muutoksia kuten esimerkiksi Paunosen edellä sivutussa tutkimuksessa omistusliitteistä.

Ojutkangas, Larjavaara, Miestamo ja Ylikoski (2016: 204) jakavat variaation kahteen päätyyppiin: yhtäältä alueelliseen tai sosiaaliseen ja toisaalta tilanteeseen vaihteluun. Ensimmäisessä kieltä käytetään eri tavoin eri puhujien ja puhujaryhmien keskuudessa ja toisessa yksittäisen puhujan kielenkäyttö vaihtelee tilanteesta toiseen. Kieltä siis käytetään eri tavoin eri puolilla Suomea ja eri yhteiskuntaluokissa, mutta yksittäisenkin

puhujan kieli kuulostaa erilaiselta lapsuudenystävien kanssa käydyin keskustelun tiimellyksessä ja työhaastattelukysymyksiin vastatessa.

Ojutkangas ym. (2016: 225) kirjoittavat kielen sisäisistä ja kielenulkoisista syistä tapahtuvasta muutoksesta. Kielen sisällä muutosta aiheuttaa tahto käyttää kieltä yhtäältä taloudellisesti, mutta toisaalta tehokkaasti. Nämä pyrkimykset voisi kiteyttää pyrkimykseen käyttää kieltä tarkoituksenmukaisesti: puhuja ei halua nähdä enempää vaivaa kielenkäytössään kuin on tarpeellista, mutta kielenkäytöllä pitää pystyä suorittamaan tietyt tehtävät. Ojutkangas ym. viittaavat taloudellisuustekijöihin syinä siihen, että kielestä karsiutuu epäsäännöllisiä muotoja äänne muutokset tapahtuvat lähinnä assimilaatioiden kautta. Toisaalta tehokkuusvaatimus hidastaa tai ohjaa kielenmuutosta toiseen suuntaan: kielen pitää säilyä käyttökelpoisena eri tehtäviinsä. Kielenulkoisia syitä heidän mukaansa ovat muiden muassa kielikontaktit, joiden tuloksena sanoja ja ilmauksia lainataan, ja kielenulkoisen maailman muutokset, jotka synnyttävät tarvetta uusille termeille tai ilmauksille.

Mielikäinen (1982: 291–292) kirjoitti vuonna 1982 *Virittäjä*-lehdessä:

”Puhekielen nopea muuttuminen ei ole suomen kielessä uutta, mutta entisiin murroskausiin verrattuna nykytilanne on kahdessa mielessä poikkeuksellinen. Ensinnäkin suomalaisten kontaktiverkko on liikenteen ja jatkuvan maassamuuton vuoksi täysin muuttunut.”

Mielikäisen kirjoituksen jälkeen ”liikenteen ja jatkuvan maassamuuton” lisäksi kielenmuutosta edistävien seikkojen lista on kasvanut valtavasti ensin uusien TV- ja radiokanavien sekä lehtitarjonnan monipuolistumisen myötä puhumattakaan maailmanlaajuisen tietoverkon ja siellä keskustelemisen ja julkaisemisen helppouden vaikutuksesta. Englannin kieltä on vaikeaa, jos ei suorastaan mahdotonta, olla kohtaamatta päivittäin. Etenkin populaarikulttuurin, mutta myös tieteellisen tutkimuksen sarjoilla sen merkitys muitten kielten kehitykseen eritoten lainasanojen, mutta myös idiomaattisten ilmausten ja jopa erilaisten rakenteiden osalta on huomattava. Kuten edellä esitetty *storm in a teacup* -esimerkki osoittaa, ilmiö ei suinkaan ole uusi.

Korhonen ja Lappalainen (2013) kirjoittavat, että ”Viime vuosien aikana tekstimaailma on räjähdysmäisesti kirjavoitunut”. Uusien kommunikaatiovälineiden läpimurto ja nopea yleistymisen sekä näiden mukanaan tuomat uudenlaiset tarpeet kommunikoida ja

julkaista tietoa ovat synnyttäneet kokonaan uusia tekstilajeja ja -tyylejä. Aikaisempaa suurempi osuus julkaistavista teksteistä on lähes kaikkien saatavilla ja entistä useampi kirjoittaa tekstejä ja tavalla tai toisella myös julkaisee niitä.

Monet näistä uusista tekstilajeista, kuten esimerkiksi juuri tämän tutkielman aiheena olevat keskustelupalstoilla julkaistavat kirjoitukset tai eri pikaviestimillä nopeasti lähetettävät viestit, muistuttavat puhekieltä. Toisaalta kirjoitetun kielen muodot kuten LOL tai OMG, jotka lienevät alun perin muodostuneet tarpeesta säästää aikaa tai merkkejä kirjoitetussa viestinnässä, vaikuttavat puhekieleen mahdollisesti enemmän kuin koskaan ennen eli kielenmuutoksen tapahtuminen pääasiassa puheessa ei enää olekaan aivan niin yksiselitteistä.

Korhonen ja Lappalainen (2013) kartoittivat kyselytutkimuksella mielipiteitä kielen käytänteistä. Kyselyyn vastanneista, joiksi kyselyn toteutustavasta johtuen valikoitui kielestä kiinnostuneita ihmisiä, puolet oli huolissaan suomen kielen muuttumisesta, mutta puolet taas ei. Suomen kieltä uhkaavina tekijöinä pidettiin yleisesti median välittämiä huonoja kielenkäytön malleja, kaikkialla läsnä olevan englannin ylivaltaa ja välinpitämättömiä asenteita kielen tilaa ja yleiskielen normeja kohtaan. Toisaalta kielen muutokseen vailla suurempaa huolta suhtautuvat kokivat, että ”muutos ei ole merkki kielen huonontumisesta” (mt).

Vastaajien näkemysten mukaan kielen muuttuminen on luonnollista ja elävän kielen merkki. Monet kokevat myös, että ihmiset osaavat itse käyttää asianmukaista kieltä eri tilanteissa ja että epävirallisissa yhteyksissä kuten keskustelupalstoilla käytetty puhekielenomainen kieli ei välttämättä tarkoita, että ihmiset eivät virallisissa yhteyksissä kykenisi tuottamaan hyvää ja normienmukaista yleiskieltä. Liian normittamisen nähdään myös hankaloittavan ihmisten viestimistä toisilleen, mikä kuitenkin on kielen pääasiallinen tarkoitus. Eri piirteiden sopivuutta asiatekstiin kartoitettaessa lähes yhdeksän kymmenestä piti kuitenkin omistusliitteen poisjättämistä liian puhekielisenä. (Korhonen ja Lappalainen 2013.)

Ferdinand de Saussure sanoi, että aika muuttaa kaikkia asioita eikä ole mitään syytä, miksi kieli yksin olisi muutokselle immuuni (Aitchison 2001: 4). Silti hänen mukaansa monet

älykkäät ihmiset tuomitsevat ja pahoittavat mielensä muuttuvasta kielestä pitäen muutoksen syynä tarpeetonta huolimattomuutta, laiskuutta tai oppimattomuutta. Aitchison (2001: 5–6) pohtiikin, voiko näitä muutoksesta huolissaan olevia verrata ihmisiin, jotka yrittävät pysäyttää aaltojen liikkeen tai auringon kierron. Mutta toisaalta: entä jos he ovatkin oikeassa ja kielen muutos johtuu suurelta osin siitä, ettei kieltä huolleta ja ylläpidetä riittävästi? Ehkä kielen muuttumista ei tarvitse hyväksyä väistämättömänä tai vaikka se olisikin väistämätöntä, onko mahdollista hidastaa sitä ja onko tuo hidastaminen toivottavaa? Olisiko mahdollista lajitella hyvät ja huonot muutokset ja pyrkiä estämään jälkimäisiä? (mt)

Aitchison (2001: 6) esittää kielenmuutokseen liittyen kolme mahdollista ajatusmallia, joiden voidaan nähdä liittyvän ihmiskieleen kokonaisuutena tai mihin tahansa yksittäiseen kieleen. 1800-luvulla ajateltiin yleisesti, että eurooppalaiset kielet rappeutuvat pikkuhiljaa. Todisteena tästä pidettiin vanhojen päätteiden katoamista. Toinen vaihtoehto on, että kielet kehittyvät kohti tehokkaampaa muotoa: ne kielet, jotka pysyvät parhaiten mukautumaan uusiin vaatimuksiin jäävät eloon ja valtaavat alaa toisilta. Jespersen (viitattu Aitchison 2001: 7) koki jo 1922 erilaisten päätteiden katoamisen olevan hyvä asia kielen kannalta, koska näin väärinkäsitysten mahdollisuus pienenee. Kolmas näkemys on, että kieli pysyy ajan kuluessa kehittymisen tai rappeutumisen kannalta pääosin muuttumattomassa tilassa, koska nuo vastavoimat jarruttavat toisiaan ja vaikka kieli muuttuu, muutos ei sinänsä vie kieltä parempaan tai huonompaan suuntaan (Aitchison 2001: 7). Samaa ajatusta sivuttiin edellä, kun keskusteltiin kielen taloudellisuus- ja tehokkuusaspekteista.

Markkula (2016) kirjoittaa, että kielelle on luonnollista yksinkertaistua ja helposti muodostettavat muodot syrjäyttävät käytössä monimutkaisempia sääntöjä seuraavia. Kangasmaa-Minnin (1986 viitattu Remes 2009:100–101) mukaan kielestä katoavat yleensä tarpeettomat osat. Suomen omistusliitteen käytön kehitys vaikuttaa hänen mukaansa kulkevan kehässä, sillä omistusliite on alun perin kehittynyt substantiivin jäljessä olleesta persoonapronominista.

Vielä tällä hetkellä persoonapronominin genetiivi sanaparissa *minun palloni* on yleiskielen suositusten mukaan useissa tapauksissa tarpeeton, koska pallon omistaja käy

ilmi myös pelkästä omistusliitteestä. Kuitenkin puhekielisessä muodossa *minun/mun pallo* tuo edellä tarpeettomana pidetty osa vaikuttaa olevan syrjäyttämässä omistusliitteen (vrt. Paunonen alaluvussa 2.3). Suomen puhekielen kehitys esimerkiksi *minun/mun pallo* näyttää siis kielen lähenevän viron tapaa ilmaista omistusta (Remes 2009: 96-97), jota käsiteltiin luvussa 2.3, mutta myös esimerkiksi englannin ja ruotsin vastaavia.

Schilling-Estesin (2006: 17) mukaan kielen variaation ja muutoksen malleja on mahdollista tunnistaa ja tutkia vasta, kun on kerätty riittävä määrä havaintoja huomattavalta määrältä puhujia. Siinä missä historiallisten lingvistien työ saattoi tapahtua pääosin kirjastossa, antropologisten lingvistien alkuperäiskansojen parissa tai teoreettisten lingvistien tutkimus- ja neuvotteluhuoneissaan, sociolingvistien piti hakeutua ihmisten pariin keräämään tavallisten ihmisten jokapäiväistä kieltä.

Labovin esittämän havainnoijan paradoksin mukaan sociolingvistien tavoitteena on havainnoida, miten ihmiset käyttävät kieltä, kun heitä ei havainnoida (Feagin 2006: 20). Tässä mielessä korpustutkimus – jota tarkastellaan lähemmin seuraavassa alaluvussa – on hyvä metodi, sillä korpuksen tekstit on kirjoitettu tietämättä, että niitä käytetään tutkimusmateriaalina. Toisaalta tämänkin tutkimuksen korpus on kirjoitettua kieltä, joka on parhaassakin tapauksessa edellä esitetyn mukaisesti vain puhekielen esittämistä kirjoitusmerkein, mikä toisilla muistuttaa puhekieltä enemmän ja toisilla vähemmän.

2.5 Korpustutkimuksista

Tässä teorialuvun viimeisessä alaluvussa käsitellään tämän tutkielman pääasiallista tutkimusmenetelmää korpustutkimusta. Käsittely alkaa kielentutkimukseen sopivien tekstien kriteereistä, josta jatketaan korpuksen ja korpustutkimusten määrittelyyn ja luonteeseen. Ennen tämän alaluvun omia alalukuja käsitellään vielä korpustutkimuksen toimintatapoja ja sovellusalueita.

Schneider (2006: 71) listaa millaiset tekstit ovat käyttökelpoisia kielentutkimukselle. Vaikka yksittäisten tekstien tasolla niiden käyttökelpoisuus voi vaihdella, hän listaa yleisellä tasolla muutamia vaatimuksia tutkimukseen soveltuville teksteille. Schneiderin mukaan tekstien tulee olla niin lähellä puhuttua kieltä ja eritoten kansankieltä kuin mahdollista. Jo tämä vaatimus rajaa pois suuren osan virallisista ja kaunokirjallisista

teksteistä. Tekstien tulisi myös olla peräisin eri lähteistä eli useilta eri kirjoittajilta, jotka edustavat eri sukupuolia, ikäryhmiä, yhteiskuntaluokkia, ja niin edelleen. Myös eri tyylilliset tasot ovat suositeltavia.

Edellä mainittujen seikkojen lisäksi tekstien pitää käsittää eri muotoja tutkittavasta ilmiöstä eli kattaa käytössä olevat muunnelmat mahdollisista tavoista ilmaista sama asia. Koska kvantifiointi on yksi kielen muutosta kartoittavien tutkimusten tärkeimmistä metodeista, tekstikorpuksen pitää täyttää tietyt laajuusvaatimukset. Vaikka ei sinänsä olekaan olemassa tarkkoja sanamääriä, kuinka suuria aineistojen pitää olla, korpuksen pitää kuitenkin sisältää riittävän suuri tapausfrekvenssi tutkittavan ilmiön eri muunnelmista. Korpuksen on myös hyvä olla niin laaja, että se mahdollistaa useiden lingvististen ilmiöiden kvantitatiivisen analysoimisen (Schneider 2006: 71).

Weisser (2016: 23) määrittelee korpuksen puhuttujen tai kirjoitettujen tekstien kokoelmaksi, joka on kerätty ennalta määriteltyjä kriteerejä noudattaen tarkoituksena tarjota materiaalia kielentutkimukseen. Tärkeää Weisserin määritelmässä on, että korpus on kerätty systemaattisesti ja kerääminen on suoritettu, jotta voitaisiin tutkia kielenkäytön ilmiöitä. Vaikka jotkut ensimmäisistä korpuksista olivatkin vain luettelokortteja tai papereita, tänä päivänä lähes kaikki käytössä olevat korpuksot ovat digitaalisessa muodossa, ja niitä tutkitaan tietokoneita hyödyntäen. Se, että tietojenkäsittelytehon, jonka säilyttämiseen ennen vaadittiin oma huone, voi nykyään kantaa pienessä laukussa tai jopa taskussa, avaa ennennäkemättömiä mahdollisuuksia myös kielentutkimukselle.

Myös Kennedyn (1998: 3) mukaan korpuksen koostamisen järjestelmällisyys ja kerääminen varta vasten kielitieteellistä tutkimusta ajatellen ovat korpuksen tärkeitä tuntomerkkejä, mutta hän toteaa kuitenkin, että myös muita tekstikokoelmia on mahdollista käyttää korpustutkimuksen materiaalina. Korpustutkimuksen pääasiallisina vahvuuksina Kennedy (1998: 7–8) luettelee mahdollisuuden tutkia kielen luonnetta, kuvata sen rakennetta ja selvittää sen käyttötapoja ja -tilanteita. Korpustutkimuksessa vastauksia tutkijoiden kysymyksiin etsitään tekemällä hakuja korpuksista eikä esimerkiksi havainnoimalla tai haastattelemalla.

Myös McEneryn ja Hardien (2012: 1–2) mukaan korpustutkimusta tehdään nykyään käytännössä yksinomaan tietokoneita hyödyntäen, sillä korpuukset ovat kooltaan niin suuria, ettei niitten tutkiminen muuten olisi ajankäytöllisesti tarkoituksenmukaista. He nostavat olennaiseksi asiaksi mahdollisuuden valita juuri sellaisia korpuksia, jotka parhaiten mahdollistavat kulloisenkin tutkimusongelman selvittämisen. Tietokoneiden käyttö myös vähentää inhimillisen virheen mahdollisuutta esimerkiksi eri ilmiöiden esiintymisfrekvenssejä laskettaessa ja helpottaa eri konstruktioiden käyttötarkoitusten ja kontekstien kartoittamista.

McEnery ja Hardie (2012: 1) toteavat, että korpustutkimus poikkeaa valtaosasta muista kielentutkimuksen osa-alueista, koska se ei keskity mihinkään yhteen kielen aspektiin vaan siinä pääosaa esittävät proseduurit ja metodit, joilla voidaan lähestyä miltei mitä tahansa kielitieteellistä tutkimusaihetta. Tästä syystä McEneryn ja Hardien mukaan korpustutkimuksella on potentiaalia mullistaa koko joukko kielentutkimuksen lähestymistapoja ja sen avulla voidaan hioa ja uudelleen määritellä kielitieteellisiä teorioita ja jopa kehittää uusia.

2.5.1 Korpustutkimuksen työkaluja

Tässä alaluvussa perehdytään lyhyesti korpustutkimuksen työkaluihin. Pääasiallinen paino on tämänkin tutkimuksen toteutuksessa käytetyillä ja sen onnistumisen suhteen korvaamattoman arvokkailla hakutyökaluilla. Konkordanssihakutyökalujen toiminnan kannalta kriittisiä ja lukuisia uusia ominaisuuksia mahdollistavia annotointityökaluja sivutaan myös ja luodaan vielä nopea katsaus korpustyökalujen tulevaisuuteen, joissa ohjelmat voivat entistä paremmin palvella sekä konkordanssi- että annotointitarpeita.

McEneryn ja Hardien (2012: 25) mukaan korpustutkimus antaa kielentutkijoille mahdollisuuksia lähestyä hyvin monenlaisia tutkimuskysymyksiä. Korpuksia luotaavien hakutyökalujen kehitys ja korpuksien tarkemmat analysointi- ja annotointityökalut mahdollistavat yhä useampien tutkimustyyppien toteuttamisen korpustutkimuksena. Jopa siinä määrin, että McEnery ja Hardie esittävät, että vaikka korpustutkimuksella tai muilla metodeilla lähestyttävien tutkimuskysymysten joukot ovat osittain yhtenevät, tutkimuskysymyksiä, joiden ratkomisessa tutkijat hyötyvät korpustutkimuksista, on selvästi enemmän kuin sellaisia, joissa korpustutkimusta ei voi käyttää.

Korpustutkimuksessa lähestytään kieltä konkordanssihakujen avulla. Weisser (2016: 80) määrittelee konkordanssihaun analyysitekniikaksi, jolla tutkija voi tutkia erilaisten sanojen tai muiden konstruktioiden esiintymistä ja käyttöä aidoissa tilanteissa. Kielenkäyttäjätyypit ja nämä aidot tilanteet, joissa he kieltä käyttävät voivat poiketa eri korpuksissa toisistaan, mikä mahdollistaa erilaiset vertailut esimerkiksi natiivien ja ei-natiivien tai aloittelevien ja edistyneitten oppijoiden välillä. Myös historiallinen kehitys on tärkeä tutkimuskohde, jossa eri aikojen kieltä sisältäviä korpuksia hyödynnetään. Weisser nostaa esimerkiksi korpusten käytöstä myös sanakirjatyön, jossa korpustutkimusta hyödynnetään yleisimpien, sopivimpien tai edustavimpien esimerkkien etsinnässä tai sanojen eri käyttötarkoitusten kartoituksessa.

McEnery ja Hardie (2012: 35) nimeävät konkordanssihakutyökalut korpuslingvistin tärkeimmiksi välineiksi. Ne mahdollistavat spesifien merkkijonojen – sanojen, sanan osien tai fraasien – hakemisen korpuksista. Tulokset esitetään tyypillisesti yksi osuma rivillä, joka sisältää myös kontekstin, jossa haettu konstruktio esiintyy. Konkordanssihaku- ja muut korpustyökalut ovat McEneryn ja Hardien mukaan hyvin keskeisessä roolissa, jos korpustutkimuksilla lähestyttävien tutkimuskysymystyyppien määrää halutaan merkittävästi lisätä. Ne ovat korpuslingvistin tärkeimpiä työkaluja ja näin ollen ne ja niiden kehitys myös määrittelevät ja rajaavat, mitä korpustutkimuksella voidaan saavuttaa.

Tieteen termipankin (www.tieteentermipankki.fi) määritelmä annotoinnille on ”Tekstin varustaminen kieliopillisilla tunnuksilla”. McEnery ja Hardie (2012: 30) kirjoittavat, että korpuksissa annotointi tarkoittaa lingvistisen tiedon koodaamista korpuksen teksteihin siten, että se on systemaattisesti ja tarkasti haettavissa myöhemmin. Tämä koodaaminen eli annotointi tapahtuu tarkoitukseen varta vasten kehitetyillä annotointityökaluilla. He listaavat annotoinnin kolme lähestymistapaa: automaattinen annotointi; automaattinen annotointi, jota seuraa manuaalinen tarkastus ja korjaus sekä manuaalinen annotointi. Mikään näistä tavoista ei McEneryn ja Hardien mukaan ole virheetöntä. Korpusten kokojen kasvaessa automaattinen annotointi on välttämättömyys, mutta jatkuvan annotointityökalujen kehitystyön ansiosta sen tulokset jättävät koko ajan vähemmän ja

vähemmän manuaalista korjattavaa. Tulevaisuudessa myös yhdistelmätyökalut, joilla käyttäjät voivat sekä tehdä hakuja että parantaa annotointia tulevat yleistymään (mt: 34).

2.5.2 Aiempaa korpustutkimusta idiomeista

Tämä alaluku käsittelee esimerkkinä yhtä korpustutkimusta: Marja Nenosen vuoden 2002 väitöskirjaa *Idiomit ja leksikko*. Tutkimusmetodin lisäksi yhtymäkohtana tähän tutkimukseen on kiinnostus idiomeihin mukaan lukien olotilaa ilmaisevat omistusliitteelliset verbilausekeidiomit. Alaluvussa käsitellään myös Nenosen tutkimuksen ja tämän tutkimuksen tärkeimpiä eroja.

Nenonen (2002: 1) pyrki väitöskirjassaan muodostamaan yleisnäkemyksen suomen kielen idiomeista ja kartoittamaan, ”miten idiomit sijoittuvat kielisysteemiin, etenkin suhteessa morfologiaan ja syntaksiin”. Hän (2002: 46) valitsi projektiaan varten kymmenen 1990-luvulla julkaistua nuorille suunnattua kirjaa, joista hän koosti noin 300 000 sanetta käsittävän korpuksen. Analyysi tuotti korpuksesta idiomin noin sataa sanetta kohti eli yhteensä reilut 3 000 idiomiosumaa, joista lähes 2 000 oli verbilausekeidiomeja.

Idiomeista noin joka viides tuotti enemmän kuin yhden hakuosuman: yhteensä eri verbilausekeidiomeja korpuksessa on 1 282 (Nenonen 2002: 54). Nenosen tutkimus poikkesi hyvin oleellisesti tästä tutkimuksesta siinä mielessä, että korpus oli koottu kaunokirjallisista teoksista ja hyvin rajallisesta joukosta kirjoittajia eli tutkimuksen tässä kuvatus osion ei ollut tarkoituskaan tutkia kansankieltä. Nenosen väitöskirjatutkimus käsitti myös muita korpuksia – esimerkiksi Karjalaisen ja Turun Sanomien korpuksia sekä Suomen kielen tekstipankkia – hyödyntäviä osia.

Nenosen (2002: 61) korpuksessa noin joka viides kaksisanainen verbillinen idiomi oli vain monikossa esiintyvä. Normaalista tavallisempi monikko vaikuttaa hänen mukaansa olevan ulkopaikallissijoissa lukuun ottamatta illatiivia, jossa yksikkö on yleisempi. Omistusliitteellisiä verbillisiä idiomeja korpuksessa oli 156 keskenään erilaista. Nenosen havainnon mukaan omistusliite kongruoi useimmiten pääsanana kanssa, mutta korpus käsitti myös kolmanteen persoonaan kivettyneitä verbi-idiomeja. Inessiivissä ja adessiivissä esiintyvät – etenkin olla verbin kanssa muodostettavat – verbilliset idiomit

ilmaisevat useimmiten henkistä tilaa kuten esimerkiksi *olla huolissaan*, *olla pahoillaan* tai *olla innoissaan*. Nenonen (2002: 62) kirjoittaakin, että ”suomessa monikko ja possessiivisuffiksi yhdessä paikallissijojen kanssa muodostavatkin eräänlaisen kiteytyneen olotilailmausten muotin”.

Tämän tutkimuksen kannalta Nenosen tutkimus on merkityksellinen siksi, että hän tekee siinä ansiokkaan katsauksen suomen kielen idiomeihin mukaan lukien omistusliitteelliset olotilanilmausidiomit, mutta ennen kaikkea siksi, että tämän tutkimuksen idiomien valintaa tukeneen pilottitutkimuksen idiomit poimittiin Nenosen keräämältä idiomilistalta. Nenosen tutkimuksen ja tämän työn erona on se, että tämä on pro gradu - tutkielmana luonnollisesti väitöskirjaa kapea-alaisempi, mutta myös se, että tämän tutkimuksen tavoitteena on selvittää idiominkäytössä ja etenkin tietyissä omistusliitteellisissä idiomeissa ajan kuluessa tapahtuvaa muutosta, josta tutkimustietoa on vielä niukanlaisesti.

3. Tutkimus

Tässä luvussa esitetään tutkimuksen aineisto ja metodit sekä siihen sisältyneet tutkimukset – pilotti- ja päätutkimus – tuloksineen ja analyyseineen. Ensimmäisessä alaluvussa käsitellään tutkimuksen aineisto ja metodit. Sitä seuraa alaluku 3.2, jossa käydään läpi työn pilotti- ja päätutkimukset. Päätutkimuksen alaluvuissa esitellään sen hakulausekkeet, yhteenveto idiomien muodostamisessa käytetyistä verbeistä, esimerkkejä hakutulosten perkaamisprosessista ja tutkimuksen varsinaiset tulokset.

3.1 Aineisto ja metodit

Tässä tutkimuksessa selvitetään subjektin ja idiomien nominiosan omistusliitteen persoonakongruenssia kolmen idiomien – *olla huolissaan*, *olla pahoillaan* ja *olla innoissaan* – käytössä ja minkälaisia muutoksia siinä on mahdollisesti tapahtunut tietyn aikajakson kuluessa. Kyseessä on korpustutkimus ja sen aineistona toimii Kielipankin Suomi24 virkkeet -korpus (2017H2) (beta), joka käsittää Suomi24-keskustelupalstalla vuosina 2001–2017 julkaistut viestit. Korpusajot on tehty Kielipankin selainpohjaisella KORP-konkordanssihakuohjelmalla, jolla korpuksesta haetaan dataa sen morfosyntaktisen jäsenyyksen pohjalta. Hakulausekkeet on kuvattu tarkemmin alaluvussa 3.2.2.1.

KORP-ohjelmasta konkordanssihakujen raakadata vietiin Microsoft Exceliin tulosten perkausta, järjestelyä, laskentaa, analysointia ja muokkausta varten. Tutkimusraportin taulukot on toteutettu Exceliä käyttäen kuten myös datan muokkaus tilastollisen analyysin vaatimaan muotoon. Tilastolliset analyysit on tehty IBM SPSS-ohjelmalla ristiintaulukointia hyödyntäen.

Tutkimus sisältää pilottitutkimuksen, jonka perusteella tutkittavat idiomit valittiin sekä päätutkimuksen, josta saatua dataa analysoimalla pyritään saamaan vastauksia tutkimuskysymyksiin. Pilottitutkimuksen korpusajot tehtiin 25.3.2019 ja päätutkimuksen ajot 19.–21.4.2019. Ajojen päivämäärät ovat sikäli merkityksellisiä, että Kielipankin ilmoituksen mukaan ”korpus on beetatestausvaiheessa, joten siihen voi tulla muutoksia ilman ilmoitusta”. Tämä tarkoittaa sitä, että tutkimusta toistettaessa tulokset saattavat poiketa jossain määrin aikaisemmista, mutta tiedot muutoksista, mikäli niitä tapahtuu, dokumentoidaan varmasti korpusta ylläpitävän Kielipankin toimesta. Tutkimusaineisto käsittää kaikki Suomi24-korpuksen tämän version alakorpukset, jotka käsittävät kukin yhden kokonaisen vuoden viestit vuosien 2001 ja 2017 välillä.

Suomi24 on sen omistaman ja ylläpitämän Aller Median sivujen mukaan ”Suomen suurin keskustelualusta 2,3 miljoonalla kuukausittaisella käyttäjällään”, mutta luku kertoo palstan tämänhetkisen käyttäjämäärän, joka on tutkimusvuosina korpuksen sane- ja virkemääristä päätellen vaihdellut. Sukupuolijakauma – niin ikään tämän hetken tilanne – on suhteellisen tasainen 52 % naisia ja 48 % miehiä. Ikäryhmistä aktiivisimpia käyttäjiä ovat 35–44-vuotiaat, joita on reilu viidennes kaikista käyttäjistä. Kaiken kaikkiaan Suomi24-korpus käsittää yli neljä miljardia sanetta ja yli 350 000 virkettä. Aktiivisimmillaan palstalla on oltu vuosien 2005–2009 välillä, jolloin vuosittainen sanemäärä oli yli 300 000 ja virkkeitä kirjoitettiin vähintään 26 000 vuodessa. Keskustelupalsta ja näin ollen myös korpus on nimimerkkien käytöstä johtuen lähestymiseltään anonyymi, joten korpus ei mahdollista tulosten vertailua eri taustamuuttujien kuten sukupuoli tai asuinlääni suhteen.

Taulukko 1. Suomi24-korpuksen vuosittaiset sane ja virkemäärät.

vuosi	Saneet	Virkkeet
2001	35 649 363	2 937 056
2002	65 453 459	5 431 137
2003	110 659 660	9 241 164
2004	203 900 194	16 901 072
2005	314 690 348	26 337 748
2006	400 735 051	33 943 563
2007	379 315 696	32 119 129
2008	355 877 731	30 219 927
2009	332 817 298	28 183 689
2010	283 986 287	23 583 371
2011	254 803 891	21 401 802
2012	264 995 748	22 498 189
2013	247 978 970	21 164 246
2014	265 570 125	23 000 358
2015	223 278 292	20 298 522
2016	225 236 906	20 196 732
2017	167 716 831	14 527 297
Yht	4 132 665 850	351 985 002

Koska Suomi24-keskustelupalstalle on eri vuosina kirjoitettu eri määrä viestejä ja virkkeitä ja näin ollen alakorpuksien sanamäärät vaihtelevat suuresti, eri idiomien suosituksenvastaista käyttöä ei vertailla absoluuttisina käyttökertoina vaan sen prosentuaalisena osuutena kyseisen idiomien kokonaiskäytöstä. Tässä työssä kartoitetaan idiomien käyttöä ainoastaan yksikön ensimmäisen persoonan osalta.

Tämän työn tutkimuskysymykset ovat:

- Kuinka yleistä on kongruoimattoman omistusliitteen käyttö olotilanilmausidiomeissa yksikön ensimmäisessä persoonassa?
- Onko kongruoimattoman eli suosituksenvastaisen muodon käytön osuudessa idiomien kokonaiskäytöstä tapahtunut muutoksia Suomi24-korpuksen käsittämän ajanjakson puitteissa?
- Onko tutkimuksen idiomien välillä eroja siinä, miten yleistä suosituksenvastainen käyttö on?

Hypoteesini tutkielmalle ovat, että kongruoimattoman eli kolmanteen persoonaan kivettyneen omistusliitteen käyttö yksikön ensimmäisessä persoonassa on lisääntyvä

ilmiö, mutta että eri idiomeissa suosituksenvastainen kielenkäyttö on eri tasoilla eli mukautuvat omistusliitteet kiinnittyvät tiiviimmin osaan idiomeista ja heikommin toisiin. Testaan näitä hypoteeseja tämän tutkimuksen kolmen idiomien suhteen kolme aikajaksoa käsittävällä korpustutkimuksella, jonka tulosten tilastolliset merkitsevyydet selvitetään ristiintaulukoimalla ja Pearsonin χ^2 -testillä. Tutkimukseen valittujen kolmen idiomien valintaprosessi käydään tarkemmin läpi alaluvussa 3.2.1.

3.2 Tutkimuksen toteutus ja tulokset

Tässä luvussa esittelen tutkimuksen toteutusta käytännössä. Pilottitutkimuksessa, jota käsittelen seuraavassa alaluvussa, kartoitin eri idiomien yleisyyttä Suomi24-korpuksessa löytääkseni päätutkimukseen tarkoituksenmukaiset, riittävän suuren frekvenssin omaavat idiomit. Päätutkimuksessa selvitin kolmen idiomien persoonakongruenssin suhteen suosituksenvastaista käyttöä edellä mainitulla keskustelupalstalla ja korpuksessa. Päätutkimuksen toteutusta ja tuloksia käsittelen alaluvusta 3.2.2 alkaen.

3.2.1 Pilottitutkimus idiomien yleisyydestä

Toteutin idiomien yleisyyttä selvittävän pilottitutkimuksen Suomi24-korpusta hyödyntäen. Valitsin Nenosen (2002: 149–164) listaamista idiomeista 25 yleistä olla-verbin kanssa muodostettavaa olotilanilmausidiomia, jotka esiintyvät paikallissijoissa ja sisältävät omistusliitteen. Koska valtaosa ehdot täyttävistä idiomeista esiintyy Nenosen aineistossa vain kerran, oletus yleisyydestä perustui tässä vaiheessa tutkijan intuitioon ja tämä olikin tärkeä syy siihen, että idiomien todellinen yleisyys varmistettiin pilottitutkimuksella ennen varsinaista tutkimusta. Viittä lukuun ottamatta pilotoidut idiomit esiintyvät monikossa. Tekstissä esimerkkinä toimii idiomi *olla huolissaan*, joka esiintyy inessiivissä ja joka välittääkseen idiomaattisen merkityksensä vaatii monikollisen muodon (vrt. **olla huoleessaan*) sekä omistusliitteen (vrt. **olla huolissa*).

Pilottitutkimuksen ajot käsittivät hakuosumat yksikön ensimmäisen persoonan omistusliitteellä (esim. *huolissani*) sekä kolmansien persoonien ja passiivin käyttämät muodot (esim. *huolissaan*), jotka sisältävät tutkimusta varten etsittävän ohjeistuksenvastaisen käytön eli esiintymät, joissa omistusliite ei kongruoi subjektin kanssa. Idiomit järjestin yleisyysjärjestykseen näiden kahden haun yhteenlaskettujen tulosten mukaan. Haut muissa persoonissa olisivat toki antaneet tietoa idiomien

kokonaiskäytöstä, mutta tämän tutkimuksen keskittyessä yksikön ensimmäiseen persoonaan ja siihen liittyvään persoonakongruenssiin, haut näitä kahta muotoa käyttäen muodostivat tarkoituksenmukaisimman kokonaisuuden.

Eniten osumia tuottivat haut muodoilla *tosissani/tosissaan*, *huolissani/huolissaan*, *pahoillani/pahoillaan* ja *innoissani/innoissaan*. Seuraavaksi yleisimmät idiomin osat *peloissani/peloissaan*, *kahuissani/kahuissaan*, *päissäni/päissään* ja *varuillani/varuillaan* tuottivat kukin alle puolet neljänneksi käytetyimmän idiomikomponentin *innoissani/innoissaan* osumamäärästä. Pilottitutkimuksen tarkat tulokset on listattu taulukossa 2, jossa päätutkimukseen valikoituneet idiomit on korostettu harmaalla taustalla.

Taulukko 2. Nenosen listan idiomien jälkimäisen osan frekvenssejä Suomi24-korpuksessa. Tutkimukseen valitut idiomit on korostettu taulukossa harmaalla taustalla.

Hakusana	3. persoona	Yks. 1.	Yht.
tosissaan	214410	28600	243010
huolissaan	78818	38189	117007
pahoillaan	11167	50536	61703
innoissaan	29971	9645	39616
peloissaan	12803	5009	17812
kauhuissaan	10578	5574	16152
päissään	12881	1456	14337
varuillaan	11542	1554	13096
ymmällään	4195	4636	8831
hämillään	4731	3853	8584
rähmällään	7479	135	7614
raivoissaan	5699	1864	7563
hyvillään	3805	2576	6381
onnessaan	3952	1108	5060
tolaltaan	3800	1229	5029
toheissaan	3543	966	4509
läpällään	2909	726	3635
harmissaan	2159	1448	3607
riemuissaan	2678	499	3177
haltioissaan	2281	683	2964
tajuissaan	2259	557	2816
kummissaan	1903	604	2507
täpinöissään	1308	544	1852
suunniltaan	1390	299	1689
käärmeissään/ kärmeissään	1487	151	1638

Kärkinelikosta tutkimukseen valitsin kolme viimeistä, sillä tosissaan olemisen lisäksi Suomi24-korpuksessa tosissaan ainakin otetaan, halutaan, ihmetellään, mietitään, väännetään, kaivataan, uskotaan, epäillään, säikähdetään ja niin edelleen. Esimerkiksi jäykkää mutteria saattaa pitää vääntää tosissaan, vaikka olisikin muuten leikkimielellä liikkeellä. Sanaa *tosissaan* käytetään myös merkityksessä *tosiaan*, *todella* tai *oikeasti* kuten korpuksessa esimerkiksi ”Onko se tosissaan näin?” ja ”no tosissaan houkuttelevan tarjouksen ne siitä on itsellenikin antaneet”. Jälkimmäinen esimerkki myös valaisee *tosissaan*-sanon monimerkityksisyyden tuomaa epävarmuutta: on mahdollista, että kirjoittaja tarkoittaa ”tosissaan houkuttelevan” eli ”todella/tosiaankin houkuttelevan tarjouksen” tai että ”[ovat] antaneet [tarjouksen] tosissaan”. Tämän lisäksi se esiintyy

erilaisissa ilmauksissa kuten esimerkiksi vaihdettaessa leikinlaskusta vakavampiin asioihin ”no mutta siis ihan tosissaan”.

Näin ollen *tosissani/tosissaan* tuotti enemmän hakuosumia kuin kärkinelikon kolme muuta idiomiä yhteensä, mikä olisi myös tehnyt osumien manuaalisesta perkaamisesta kohtuuttoman työlään ja aikaa vievän prosessin. Edellä mainittujen paikallissijan ja omistusliitteen lisäksi kaikki tutkimukseen valikoituneet idiomit esiintyvät monikollisina, kuten edellä esitettiin *olla huolissaan* -idiomia esimerkkinä käyttäen.

3.2.2 Päättökäytännön kolmen idiomin persoonakongruenssista

Tässä alaluvussa esitellään tutkielman päätulokset ja miten niihin on päädytty. Ensinnä tarkastellaan lyhyesti tutkittujen idiomien nominiosien perusmuotojen etymologiaa ja idiomeissa esiintyvien muotojen määritelmiä. Sen jälkeen kerrataan vielä niiden suosituksenmukaiset muodot, joita siis myös tässä tutkimuksessa haetaan korpukselta. Tämän jälkeen listataan vielä muutamia käytännön perusasioita tutkimuksesta ennen siirtymistä alalukuihin.

Tämä alaluku käsittää neljä omaa alalukua, joista ensimmäisessä määritellään ja listataan korpushakujen hakulausekkeet. Alaluvussa 3.2.2.2 kartoitetaan tutkittujen idiomien muodostamisessa käytetyt verbit ja verbimuodot suosituksenvastaisen omistusliitteen sisältävien hakuosumien osalta. Alaluvussa 3.2.2.3 valaistaan konkordanssihaun tulosten manuaalista läpikäyntiä ja sitä, minkälaisia virheellisiä hakutuloksia haku tuotti. Alaluku sisältää myös esimerkkejä syistä, jotka näitä virheellisiä osumia aiheuttivat. Alaluvussa 3.2.2.4 esitetään tutkimuksen varsinaiset tulokset ja vastaukset tutkimuskysymyksiin.

Tutkittaviksi idiomeiksi valikoituivat siis viime alaluvussa kuvatun pilottitutkimuksen perusteella *olla huolissaan*, *olla pahoillaan* ja *olla innoissaan*. Näistä *Nykysuomen etymologisen sanakirjan* (Häkkinen 2013) mukaan sanat *huoli* ja *paha* ovat kuuluneet suomen kirjakielen jo Agricolan ajasta lähtien ja *into* tavataan ensimmäistä kertaa Henrik Gabriel Porthanin lisäyksenä Daniel Jusleniuksen sanakirjaan 1770-luvulla. *Huoli*-sanon sanavartalo on balttilainen laina, jolle on vastineet kaikissa itämerensuomalaisissa kielissä. *Paha*-sanon alkumuotoa pidetään kantaskandinaavisena, josta se on niin ikään laajalti levinnyt itämerensuomalaisiin kieliin. Sanon *into* alkuperän

etymologinen sanakirja ilmoittaa kantagermaaniseksi. Myös sillä on vastineita lähisukukielissä, vaikkakaan ei yhtä laajasti kuin edeltävillä sanoilla.

Kielitoimiston sanakirja (KTS 2018) määrittelee sanan *huolissaan* seuraavasti: ”huolestuneena, hädissään, peloissaan” ja antaa käyttöesimerkin ”*Olen huolissani hänestä, hänen vuokseen.*” *Pahoillaan* tarkoittaa sanakirjan mukaan ”pahoilla mielin, suruissaan” ja sitä voidaan käyttää esimerkiksi seuraavasti: ”*Hän on kovin pahoillaan siitä, mitä sanoit. Älä ole pahoillasi! Tulin kovin pahoilleni kuullessani asiasta.*” Sanakirja huomioi myös idiomien yksikön ensimmäisessä persoonassa saaman toisen, spesifimmän tarkoituksen: ”*Olen pahoillani* [= pahoittelen, pyydän anteeksi], *että soitan näin myöhään*”.

KTS ohjaa *innoissaan*-sanan haulla hakijan sanan *into* hakusana-artikkeliin, joka sisältää *innoissaan*-sanan määritelmän: ”innon vallassa, innostuneena, innostuksissaan” ja esimerkin: ”*Olin asiasta hyvin innoissani*”. Kaikkien kolmen sanan siis kerrotaan sanakirjassa kuvaavan erilaisia olotiloja. Kaikista kolmesta sanasta Kielitoimiston sanakirja antaa myös lisätiedon ”omistusliite persoonittain vaihteleva”. Mielenkiintoista on, että kaikkien sanojen määrittelyssä käytetään niille läheistä sukua olevia sanoja tai ilmauksia kuten *huolestuneena*, *pahoilla mielin* sekä *innon vallassa* ja *innostuneena*.

Tutkittavista idiomeista siis *olla huolissaan* ja *olla innoissaan* muodostetaan inessiiviä ja *olla pahoillaan* adessiiviä käyttäen. Ne kaikki vaativat monikollisen muodon välittääkseen idiomaattisen merkityksensä. Yhtään niistä ei mainita ISK:n tai Kielitoimiston ohjepankin listoilla, joissa listataan leksikaalistuneita eli tässä tapauksessa omistusliitteen persoonakongruenssin menettäneitä muotoja. Toisin sanoen suositus on, että näissä idiomeissa omistusliite mukautuu lauseen subjektin mukaan.

Tämän työn päätutkimus kartoittaa siis edellä mainittujen idiomien persoonakongruenssia yksikön ensimmäisen persoonan osalta: missä määrin suosituksenvastaista käyttöä on esiintynyt Suomi24-keskustelupalstan viesteissä ja onko määrässä tapahtunut muutosta korpuksen kattaman 17 vuoden ajanjakson aikana. Tilastollista analyysiä varten tulokset jaettiin kolmeen aikajaksoon, joista ensimmäinen (2001–2006) ja viimeinen (2012–2017) käsittävät kuusi ja keskimääräinen (2007–2011) viisi vuotta. Tällainen jako on perusteltu

siksi, että keskimäinen jakso käsittää vuodet, jolloin keskustelupalstalle on kirjoitettu aktiivisimmin eli tutkittavaa materiaalia on sane- ja virkemäärillä mitattuna eniten.

Vuotta pitemmät vertailuaikajaksot tuovat myös aikajaksojen tasajakoa (kolme 5v. 8kk -mittaista jaksoa) korkeammat määrät esiintymiä ensimmäiseen ja viimeiseen jaksoon, mikä on tärkeää tilastollisen analyysin kannalta. Tulosten esittäminen on myös yksinkertaisempaa, kun jaksot vaihtuvat kalenterivuositain. Ennen kaikkea tämä vuosijako tuottaa suhteellisen tasakokoiset määrät saneita jokaiseen aikajaksoon, sillä ihan ensimmäisinä vuosina palstalla ei vielä ollut yhtä paljon käyttäjiä kuin huippuvuosinaan ja viimeisessä aikajaksossa sen suosio on jo vilkkaimmista vuosista hieman hiipunut.

Sane- ja virkemäärien vaihdellessa vuodesta toiseen jokaisesta idiomista piti selvittää sekä suosituksenvastainen että suosituksenmukainen käyttö, jotta olisi mahdollista laskea suosituksenvastaisen käytön osuus kokonaiskäytöstä. Edellä mainittu osuus kertoo tapahtuneesta muutoksesta enemmän kuin vain pelkkä suosituksenvastaisen käytön esiintymien määrä, joka saattaa nousta kirjoitettujen virkkeiden määrän noustessa, vaikka sen osuus idiomin kaikesta käytöstä vähenisi.

3.2.2.1 Hakulausekkeet

Suosituksenvastainen käyttö kartoitettiin hakemalla korpuksessa virkkeitä, joissa oli yksikön ensimmäisessä persoonassa taivutettu verbi sekä pääsana perusmuodossaan esimerkiksi *huolissaan*. Ensimmäinen hakurajaus oli ”Morfologinen analyysi alkaa PRS_Sg1”, joka tuottaa tuloksenaan esimerkiksi *olen*, *olin* ja *olisin* mahdollisine liitepartikkeleineen ja murteellisine muotoineen.

Tutkimukseen valittuja idiomeja käytetään siis korpuksessa myös ilman *olla*-verbiä esimerkiksi ”vähän huolissaan kysyin viestillä miespuoliselta työkaverilta” tai ”Menin ihan innoissaan ostamaan tuon kirjan”, mutta ei samassa laajuudessa kuin siitä pois rajattua *olla tosissaan* -idiomia. Ajatuksena toki myös esimerkkien kaltaisissa tapauksissa on, että tekijä on huolissaan tai innoissaan, vaikkei sitä *olla*-verbillä erikseen tuodakaan julki. Tämän tyyppinen esiintyminen vaikuttaa kuitenkin olevan huomattavasti harvinaisempaa kuin *tosissaan*-konstruktiolla. Hakulauseke ”Morfologinen analyysi

alkaa PRS_Sg1” tuottaa siis myös kaikki muut yksikön ensimmäisessä taivutetut verbit eli kattaa yllä esitettyjen esimerkkien kaltaiset tapaukset. Eri idiomien muodostamisessa käytettyjä verbejä käsitellään seuraavassa alaluvussa.

Verbi ja idiomien jälkimmäinen komponentti eivät välttämättä ole lauseessa aivan peräkkäin eli hakulausekkeisiin tarvitaan idiomien jäsenten väliin määritelmä ”sana on <mikä tahansa sana>” sekä ”toista 0–9 kertaa”. Tämä mahdollistaa siis 0–9 muun sanan esiintymisen idiomien osien välissä. Idiomien jälkimmäiseenkin osaan saattaa liittyä esimerkiksi liitepartikkeli *-ko*, *-kin* tai *-kaan* eli hakulausekkeisiin pitää sisällyttää sen suhteen ehto ”sana alkaa *innoissaan*”, jossa *innoissaan* korvataan aina kulloisellakin hakusanalla.

Suosituksen mukainen käyttö eli kongruoivan omistusliitteen sisältävät muodot kartoitettiin yksinkertaisella haulla käyttäen ainoana hakulausekkeena esimerkiksi ”sana alkaa *pahoillani*”, jossa *pahoillani* korvataan vuorollaan kunkin sanan vastaavalla muodolla. Suosituksenvastaisiakin muotoja korpuksesta löytyy, mutta ei siinä määrin, että olisi perusteltua käydä yksi kerrallaan läpi edellä kuvattujen hakujen tuottamat lähes satatuhatta riviä tuloksia. Lienee myös perusteltua olettaa, että esimerkiksi muodot *olet* tai *olette huolissani*, joita 17 vuoden keskustelut käsittävästä korpuksesta löytyy yhteensä kolme kappaletta, selittyvät ennemmin näppäilyvirheillä tai tekstin kesken jääneellä editoinnilla kuin sillä, että joku oikeasti muotoja aktiivisesti käyttäisi. Toki lienee myös mahdollista, että kyseessä on mukaelma ilmaisusta *olet ajatuksissani*, mutta siinäkin tapauksessa olematon frekvenssi ei motivoi toimenpiteitä

3.2.2.2 Idiomien muodostamisessa käytetyt verbit

Vaikka *olla huolissaan* -tyypin idiomi voidaan siis muodostaa myös käyttämällä *olla*-verbin sijasta jotain muuta verbiä, suosituksenvastaisista muodoista Suomi24-korpuksessa yli puolet on muodostettu *olla*-verbin preesensmuodoilla *olen*, *oon*, *olenkin*, *olenko* tai *oonko* ja lähes neljäsosa konditionaalilla *olisin*, *oisin*, *olisinkin* tai *olisinko*. Kaiken kaikkiaan *olla*-verbin preesens, konditionaali, imperfekti ja perfekti kattavat yli 90 prosenttia *olla huolissaan* -tyypin idiomeista. Huolissaan ollaan pääosin myönteisissä lauseissa, sillä kieltomuoto esiintyy alle prosentissa tapauksista. On tärkeää huomioida, että tämän tyyppiset analyysit kaikista kolmesta tutkitusta idiomista on tehty ainoastaan

niistä esiintymistä, joissa tavattiin tutkimuksessa etsittyä suosituksenvastaista käyttöä. Tarkempi jakauma tuloksista on esitetty taulukossa 3.

Taulukko 3. *Olla huolissaan* -tyypin idiomien muodostamisessa käytetyt verbimuodot.

... huolissaan	kpl	%
Olen/oon/olenkin/olenko/oonko	773	53,5
Olisin/oisin/olisinkin/olisinko	352	24,3
Olin/olinkin	141	9,8
Olen ollut/oon ollut/oon ollu	54	3,7
Alan olla/alan oleen/alan olemaan/rupean olla	29	2,0
Pitäisikö/pitäiskö/tarvitseeko/tuleeko minun/mun olla	9	0,6
En/enkä/etten ole/oo	8	0,6
Saan olla	5	0,3
Muut muodot/verbit	75	5,2

Olla pahoillaan -tyypin idiomia muodostetaan niin ikään pääosin *olla*-verbin kanssa: useammin kuin yhdeksän kertaa kymmenestä se on idiomien verbinä preesens-, konditionaali- tai imperfektimuodoissaan. Preesens dominoi kuitenkin tämän idiomien käyttöä vielä *olla huolissaan* -idiomia voimakkaammin: *olen* tai *oon* -muotoja käytetään useammin kuin neljässä tapauksessa viidestä. Kielteisesti pahoillaan olemisesta kerrotaan vain alle kahdessa prosentissa tapauksista. *Alan olla* esiintyy aineistossa kaksi kertaa eli vajaan prosentin osuudella. Muu-ryhmän muodot ovat yksittäisiä esiintymiä ja niitä on yhteensä vajaat kuusi prosenttia. Tarkemmat tulokset alla taulukossa 4.

Taulukko 4. *Olla pahoillaan* -tyypin idiomien muodostamisessa käytetyt verbimuodot.

... pahoillaan	kpl	%
Olen/oon	187	84,6
Olisin/olisinko	8	3,6
Olin	7	3,2
En/enkä ole/oo	3	1,4
Alan olla	2	0,9
Muut	13	5,9

Tutkimuksen kolmas idiomi, huomattavasti vähemmän hakuosumia tuottanut *olla innoissaan*, poikkeaa muutenkin kahdesta muusta. Sen muodostamisessa *olla*-verbiä käytetään vain noin kahdessa tapauksessa kolmesta: imperfektissä innoissaan olemisesta kerrotaan reilussa 30 prosentissa tapauksista ja preesensissä hieman tuota harvemmin. Kun ensimmäisissä kahdessa idiomissa muut kuin *olla*-verbilliset muodot muodostivat

selvästi alle kymmenesosan tapauksista, *olla innoissaan* -tyypin idiomeissa pelkästään *odottaa*-verbin preesens- ja imperfektimuotojen kanssa rakentui jo tuota suurempi osuus.

Kyse on kuitenkin edelleen *olla innoissaan* -idiomista, sillä kertoja on innoissaan odottaessaan, mennessään, kertoessaan, lähtiessään ja niin edelleen, vaikka itse *olla*-verbi jätetäänkin usein näissä tapauksissa sujuvampaa ilmaisua tavoitellessa pois. Poikkeuksellisen suuri muut-ryhmä koostuu alle prosentin osuuden saavuttavista – pääasiassa yksittäisistä – verbeistä kuten *katsoin, kuuntelen, kirjoitin, luin, sanoin, tekstasin* ja niin edelleen. Taulukkomuodossa tulokset esitetään alla taulukossa 5.

Taulukko 5. *Olla innoissaan* -tyypin idiomien muodostamisessa käytetyt verbimuodot.

... innoissaan	kpl	%
Olin/olinpa/olinhan	78	31,2
Olen/oon/oonkin	72	28,8
Odotan/ootan/odottelen/oottelen/ootankin	19	7,6
Olisin/oisin	8	3,2
Odotin/ootin	7	2,8
Olen/oon ollut/ollu	6	2,4
Menin/menen	6	2,4
Kerroin	4	1,6
Lähdin/lähden	4	1,6
Muut	47	18,4

3.2.2.3 Hakutulosten läpikäynti

.Suosituksenmukainen käyttö muodosti valtaosan kokonaiskäytöstä kaikkien tutkittujen idiomien kohdalla ja – kuten alaluvussa 3.2.2.1 esitettiin – sitä kartoittavien hakujen käyminen läpi virheosumien karsimiseksi ei ollut tarpeellista. Suositusvastaista käyttöä kartoittavien hakujen tulokset sitä vastoin oli käytävä läpi rivi kerrallaan, jotta löydettäisiin todelliset tapaukset ja voitaisiin karsia pois erilaiset virheelliset osumat. Esimerkiksi ilmaus *juuri näin* tuotti virheellisen osuman, kun KORP käsitteli sitä *nähdä*-verbin yksikön ensimmäisen persoonan imperfektimuotona. Samanlaisia ongelmia aiheuttivat *päin vastoin*, joka palautui hausta mitä ilmeisimmin *vastoa*-verbin imperfekti muotona, ja *aika ajoin*, jonka KORP luokitteli *ajaa*-verbin vastaavaksi muodoksi.

KORPin aineistot on analysoitu automaattisesti ja näin ollen haut tuottavat sangen koomisiakin tuloksia, kun esimerkiksi verbien *malittaa, aitoa, murtaa ja kertoa* muotoja

tunnistetaan virheellisesti lauseista ”Maltan johto on huolissaan...”, ”... jos kaikki uskovaiset olisivat aidon huolissaan...”, ”Armeija syrjäytti Muslimiveljeskuntaan kuuluneen presidentti Mohammed Mursin, jonka valtaannoususta maan kristityt ovat olleet huolissaan” ja ”ollaani kaksin kerroin pahoillaan ja pyytelemässä anteeksi”.

Suosituksenvastainen käyttö voi olla myös epäsuoraa esitystä, jossa kerrotaan jonkun toisen sanomisista. Yksi käyttäjä kertoo: ”Halonen sanoi tänään uutisissa, että 'minä olen huolissaan'”. Toinen kirjoittaa: ”Kepu J. Sipilä: olen "huolissaan siitä"”. Vaikkei tämä suoranaisesti kerrokaan tutkimuksen aiheena olevasta Suomi24-keskustelupalstan käyttäjien kielenkäytöstä palstalle kirjoittaessaan, voidaan havaita, että puhekielessä suosituksenvastaista käyttöä voidaan tavata jopa tasavallan korkeimmalla johdolla aina entistä presidenttiä ja pääministeriä myöten.

Virheellisen osuman aiheutti myös muuan käyttäjä, joka opasti kanssakäyttäjiään *olla huolissaan* -ilmauksen persoonakongruenssissa: ”Olen huolissani, olet huolissasi, hän on huolissaan, olemme huolissamme jne ...”. Puheenvuoron aloittava *Olen* ja sitä myöhemmin virkkeessä seuraava *huolissaan* täyttävät hakuehdot, vaikka kyseisessä viestissä ei suosituksenvastaista käyttöä olekaan. Tällaiset virheosumat olivat sangen yleisiä johtuen yksikön ensimmäisessä persoonassa haetun verbin ja *huolissaan*-sanan välissä sallittujen sanojen määrästä: usein kirjoittaja kertoi omista havainnoistaan tai aavistuksistaan, miten joku tai jotkut muut olivat huolissaan jostakin, itsestään, jostakusta tai joistakuista muista. Toisaalta oli välttämätöntä hyväksyä korkea määrä sanoja idiomien verbin ja pääsanan väliin, sillä vaikka lause ”Minä olisin tuolloisesta pyynnöstä enemmän kuin innoissaan” olisi päätynyt hakutuloksiin, esimerkiksi lause ”Olin kans tuolloin 8kk paikkeilla yhtäkkiä hirveen huolissaan...” olisi määrittelyksellä, joka sallii 0–5 muuta sanaa idiomien osien väliin, jäänyt tulosten ulkopuolelle, vaikka siinä selkeästi itsestään kertova kirjoittaja on huolissaan vailla kongruoivaa omistusliitettä.

Populaarikulttuuri tuottaa myös useita virheosumia dataan. Yksi keskustelija miettii: ”Taidan lähettää kysymyksen sinne TV-ohjelmaan; "Pitäisikö olla huolissaan"?” Kyseinen ohjelma (Pitääkö olla huolissaan? MTV3 kanavalla syksystä 2016) on löytänyt katsojia myös Suomi24-keskustelijoiden joukosta ja keskustelu siitä tuottaa osumia tämän tutkimuksen hakulausekkeilla, vaikkei valtaosassa niistä suosituksenvastaista

käyttöä esiinnykään. Kuitenkin mitä tahansa käyttäjä kertoo ohjelman tiimoilta yksikön ensimmäisessä persoonassa tehneensä, se tuottaa rivin hakutuloksiin.

Huolissaan olemisen tarpeellisuuden aprikoiminen ei kuitenkaan alkanut syksyllä 2016, kun edellä mainittu ohjelma lanseerattiin, vaan Suomi24-keskustelijat olivat kirjoittaneet aiheesta jo huomattavasti aikaisemmin. Hakutuloksiin aiheesta päätyi kahta päätyyppiä eli nollapersoonarakenteinen ”tarkoitin sitä, että pitääkö olla huolissaan” ja ”Aloin tässä pohtia, pitäisikö sittenkin olla huolissaan” sekä omakohtaisesti ensimmäisessä persoonassa pohdittava ”kirjoitin tänne, koska mietin että pitäiskö mun olla huolissaan” tai ”haluaisin nyt vielä vain kysyä että olisiko minun syytä olla huolissaan”. Näistä kahden ensimmäisen esimerkin nollapersoonana, yksikön kolmannen persoonan verbimuoto ja omistusliite seuraavat suosituksia, mutta kahden viimeisen genetiivisubjektit vaatisivat suositusta seurataksaan yksikön ensimmäisen persoonan omistusliitteet esimerkiksi ”olisiko minun syytä olla huolissani”.

Mielenkiintoisia tapauksia tutkijalle tuotti myös pronomini *itse*. Virkkeessä ”Olin huolissani hänestä vaikka olisi pitänyt olla huolissaan itsestä sillä ero pitkitty kohdallani” ensimmäisen sivulauseen nollapersoonarakenne aiheuttaa sen, että sanan *huolissaan* omistusliitteen taivutus seuraa yksikön kolmannen persoonan mallia, mutta sanan *itse* viitatessa päälauseen tekijään, se kaipaisi ensimmäisen persoonan omistusliitettä. Jos nollapersoonan sijaan olisi käytetty ensimmäisen persoonan genetiivisubjektia, sekä *huolissaan-* että *itse-*sanojen olisi pitänyt kongruoida subjektin kanssa: ”...vaikka minun olisi pitänyt olla huolissani itsestäni...”. Alkuperäisessä muodossaan virke ei kuitenkaan sisällä suosituksenvastaista käyttöä tämän tutkimuksen aiheen eli idiomien persoonakongruenssin osalta, vaikka yksi omistusliite (itseni) jääkin uupumaan.

Pronomini *itse* on mielenkiintoisessa osassa myös virkkeessä: ”Tunteen puolella taas tiedän olevani temperamenttinen, jotenkin voisin normaalisti uskoa itseni olevan äärimmäisen pettynyt ja pahoillaan, mutta aikuisena ihmisenä ajattelen liian järkevästi”. Käyttäjä viittaa ensimmäisessä sivulauseessa itseensä kolmannessa persoonassa, joten sinänsä muoto *pahoillaan* seuraa suosituksia, mutta varsin epätavallinen lauseen muotoilu varmasti on verrattuna muotoon ”voisin normaalisti uskoa olevani äärimmäisen pettynyt

ja pahoillani”. Virkkeen alun *tiedän olevani* -rakenne ei toistu muodossa *voisin uskoa olevani* vaan harvinaisemmassa ”voisin uskoa itseni olevan”.

Suomi24-korpuksessa löytyy monenlaisia malleja omistusliitteen käytöstä tutkitun kaltaisten idiomien yhteydessä. Osa käyttäjistä jättää johdonmukaisesti tämän tyyppiset idiomit taivuttamatta: ”Ensin olin huolissaan, sitten raivoissaan”, ”Vauvan iho on aiemmin ollut hyvä ja nyt olenkin ihmeissään ja huolissaan näppyloistä” ja ”Minä ainakin koin asian näin vaikka huolissaan ja mietteissään olinkin”. Joskus viesteissä osa idiomeista kongruoi ja osa ei: ”Enemmän olen ollut peloissani ja huolissaan asioista” ja ”Olen murheissani ja huolissaan”.

Mielenkiintoisia tapauksia ovat myös sellaiset, joissa kongruenssin malli vaihtuu kesken virkkeen samaa idiomia käytettäessä eli yksittäinen virke voi sisältää sekä suosituksenmukaista että suosituksenvastaista käyttöä. Käyttäjä voi vaihtaa aikamuodon vaihtuessa myös idiomin taivutusta: ”Olen huolissani edelleen olen ollut jo pitkään huolissaan kuten välittämästään ihmisestä ollaan”. Samoin käy esimerkin käyttäjälle, kun ilmaus vaihtuu myönteisestä kielteiseksi: ”Enemmänkin olen huolissani kaiken maailman arvaamattomista sählämeistä, tai siis en huolissaan heistä vaan siitä, että he tulevat tänne.” tai kun verbi muuttuu kysyväksi: ”vaivaako mua vaan sairaus nimeltä äitiys ja olen turhankin huolissani vai olenko syystä huolissaan?”

Sanapari *olisin huolissaan* tuottaa eri muodoissaan yhteensä 100 etsityn kaltaista, suosituksenvastaista osumaa, mutta myös virheellisiä osumia, joissa ei esiinny haettua muotoa. Esimerkiksi ”... ennemmin sinuna olisin tuosta huolissani jos jostain nyt ylipäätään pitää huolissaan olla” sisältää suosituksenmukaisen muodon *olisin huolissani*, mutta palautuu hakuosumana jäljempänä olevan *huolissaan*-muodon takia. Sama tilanne kohdataan virkkeessä: ”Olen huolissani siitä, että muslimit ovat huolissaan ja katsovat tulleensa uhatuiksi”.

KORPin tarjoama mahdollisuus tutkia koko virkettä täytyy tämän tyyppistä tutkimusongelmaa selvittäessä käyttää hyväksi. Pelkkiä hakusanoja ja niiden väliin sijoitettavia sanoja tutkimalla esimerkiksi ”leikin pennun kanssa niin innoissaan” vaikuttaisi sisältävän suosituksenvastaisen muodon *leikin innoissaan*. Koko virkettä

”Eilen sitten kun leikin pennun kanssa niin innoissaan sitten puras mua kädestä ja kunnolla niin että sattui ja jälkikin tuli” tutkiessa onkin helppo päätyä johtopäätökseen, että innoissaan olikin pentu eikä tarinan kertoja. Sama tilanne on lauseessa ”voisin tunnista toiseen tuijottaa innoissaan mattoa repivää onnellista hauvaa”, jossa vasta idiomien jälkimmäisen komponentin jälkeinen osa virkkeestä ilmaisee, ettei innoissaan ole kertoja vaan mattoa repivä hauva. Myös lause ”Olen aina ollut hyvin huolissaan olevaa sorttia” tuottaisi ilman kontekstia muodossa ”Olen aina ollut hyvin huolissaan” osuman väärään *olen huolissaan* -sarakeeseen.

Pohdittavaa tutkijalle aiheutti myös Suomi24-aineistossa 15 kertaa esiintyvä muoto *huolissaani*, joka on mahdollisesti näppäilyvirheestä johtuva hybridi yksikön ensimmäisen ja kolmansien persoonien sekä passiivin muodoista *huolissani* ja *huolissaan*. Muoto esiintyy 13 kertaa yksikön ensimmäisessä persoonassa käytetyn verbin yhteydessä ja kaksi kertaa kyse on passiivista. Nämä muodot olen jättänyt pois laskuista suosituksenvastaisia muotoja laskiessani eli olettanut kirjoittajien pyrkineen kielessä yleisempään, norminmukaiseen muotoon. Samoin olen menetellyt muotojen *pahoillaani* ja *innoissaani* kanssa, joita korpuksesta hauissa palautui 16 ja kuusi kertaa.

3.2.2.4 Tutkimuksen tulokset

Tämän pro gradu -tutkielman tavoitteena oli tutkia omistusliitteen persoonakongruenssia eri olotilanilmausidiomien käytössä sekä siinä mahdollisesti ajan kuluessa tapahtuvaa muutosta. Lisäksi oli tarkoitus selvittää, miten eri idiomien tulokset vertautuvat toisiinsa. Tutkimusmetodinä oli korpustutkimus Kielipankin Suomi24 virkkeet (2017H2) (beta) -korpuksesta, joka käsittää vuosien 2001–2017 Suomi24-keskustelupalstalla käydyt keskustelut vuodenmittaisiksi alakorpuksiksi jaettuina.

Jaoin alakorpuksat kolmeen suhteellisen samankokoiseen jaksoon kuitenkin noudattaen kronologista järjestystä eli vuosiin, 2001–2006, 2007–2011 sekä 2012–2017. Kukin jakso sisältää yli miljardi sanetta. Ensimmäinen jakso käsittää lähes ja kaksi viimeistä kumpikin reilusti yli sata miljoonaa virkettä. Kokonaisuudessaan korpus on yli neljän miljardin saneen ja reilun 350 miljoonan virkkeen laajuinen. Vertailujaksojen sane- ja virkemäärät on esitetty taulukossa 6.

Taulukko 6. Suomi24-korpuksen saneet ja virkkeet.

Vuodet	Saneet	Virkkeet
2001–2006	1 131 088 075	94 791 740
2007–2011	1 606 800 903	135 507 918
2012–2017	1 394 776 872	121 685 344
Yhteensä	4 132 665 850	351 985 002

Edellä kattavasti kuvatun pilottitutkimuksen perusteella valitsin tutkittavaksi kolme idiomia, joiden frekvenssien oli realistista odottaa olevan niin suuria, että päätutkimuksesta olisi mahdollista saada järkeviä tuloksia: *olla huolissaan*, *olla pahoillaan* ja *olla innoissaan*.

Taulukossa 7 esitetään tutkimuksen idiomien suosituksenvastaisen ja suosituksenmukaisen käytön suhde yksikön ensimmäisessä persoonassa vertailujaksoittain ja kokonaisuutena koko Suomi24-korpuksessa. Sarakkeessa *Ei kongruoi* nähdään *olen huolissaan* -tyypin varmennettujen hakuosumien lukumäärä, sarakkeessa *Kongruoi* esitetään suosituksenmukaisesti kongruoivat hakuosumat ja seuraavassa sarakkeessa idiomien kokonaiskäyttö valitussa persoonassa. Viimeisessä sarakkeessa on laskettuna idiomien suosituksenvastaisen käytön prosenttiosuus kokonaiskäytöstä.

Olla huolissaan -idiomin suosituksenvastaisen käytön prosenttiosuus sen kokonaiskäytöstä oli kaikkina vertailujaksoina muiden tutkittujen idiomien vastaavia korkeampaa ja *olla pahoillaan* -idiomin vastaava osuus oli vastaavasti kaikilla jaksoilla matalin eli noin puoli prosenttia tai alle. *Olla innoissaan* -idiomin suosituksenvastainen käyttö, jonka prosenttiosuus vaikutti puhtaan aritmeettisesti hienoisesti kasvavan aikaisemmista vertailujaksoista myöhäisempiin siirryttäessä, ei viimeiselläkään jaksolla noussut *olla huolissaan* -idiomin ensimmäisen jakson tasolle. Kaikilla tutkituilla idiomeilla eniten suosituksenvastaisia käyttökertoja absoluuttisina lukumäärinä oli keskimmaisella tutkimusjaksolla, kun palsta oli suosituimmillaan, mutta prosentuaalisesti suosituksenvastaisen käytön osuus oli kaikilla korkeimmillaan viimeisellä jaksolla.

Taulukko 7. Idiomien Y1-persoonakongruenssi Suomi24-korpuksessa.

Olla huolissaan				
Aikajakso	Ei kongruoi	Kongruoi	Kokonaiskäyttö	Ei kongr. %
2001–2006	289	9 710	9 999	2,89
2007–2011	615	15 788	16 403	3,75
2012–2017	542	12 532	13 074	4,15
Yhteensä	1 446	38 030	39 476	3,66
Olla pahoillaan				
Aikajakso	Ei kongruoi	Kongruoi	Kokonaiskäyttö	Ei kongr. %
2001–2006	52	17 003	17 055	0,30
2007–2011	95	19 598	19 693	0,48
2012–2017	74	14 278	14 352	0,52
Yhteensä	221	50 879	51 100	0,43
Olla innoissaan				
Aikajakso	Ei kongruoi	Kongruoi	Kokonaiskäyttö	Ei kongr. %
2001–2006	71	3 113	3 184	2,23
2007–2011	102	4 014	4 116	2,48
2012–2017	77	2 610	2 687	2,87
Yhteensä	250	9 737	9 987	2,50

Korpusajoista saatujen tulosten ja niistä laskettujen prosentiosuuksien muutoksen tilastollista merkitsevyyttä tutkittiin SPSS-tilasto-ohjelman ristiintaulukointitoimintoja hyödyntäen. Analyysin tulosten tulkitsemisessä on hyödynnetty Tähtisen, Laakkosen ja Brobergin (2011) kirjaa *Tilastollisen aineiston käsittelyn ja tulkinnan perusteita*.

Ristiintaulukoinnista saadun Pearsonin khiin neliö -arvon perusteella voidaan todeta, että *olla huolissaan* idiomien suosituksenvastaisen käytön prosentiosuudessa nähdään tilastollisesti merkitsevä muutos ensimmäisen vertailujakson ja kahden viimeisen vertailujakson välillä, mutta toisen ja kolmannen jakson välinen muutos ei ole tilastollisesti merkitsevä. Molemmilla aikajakson 2001–2006 jälkeisillä jaksoilla suosituksenvastainen käyttö on siis kuitenkin ensimmäistä jaksoa yleisempää.

Vaikka kolmannen jakson suosituksenvastaisen käytön prosentuaalinen osuus on laskennallisesti toisen jakson vastaavaa korkeampi, ristiintaulukoinnin perusteella näiden jaksojen välillä ei ole tilastollisesti merkitsevää eroa. Näin ollen ei voida sanoa, että muutoksessa olisi havaittavissa trendi kohti suosituksenvastaisen käytön lisääntymistä, mutta koska muutos ei ole merkitsevä myöskään vähenevään suuntaan, idiomien suosituksenvastaisen käytön osuus näyttäisi vakiintuneen ensimmäisen jakson tasoa

korkeammaksi. *Olla huolissaan* -idiomin testituloksen numeeriset arvot ovat $\chi^2(2)=25,895$, $p<0,001$. Ne osoittavat tilastollisesti merkitsevän eron aikajaksojen välillä, mutta Cramérin V-arvo 0,026 kertoo, että ero on suhteellisen pieni.

Idiomi *olla pahoillaan* on poikkeuksellinen siinä mielessä, että pilottitutkimuksessa sen yksikön ensimmäisen persoonan muoto *pahoillani* tuotti näistä muodoista suurimman määrän osumia ja samaan aikaan *pahoillaan*-muodon osumia oli vähemmän kuin toisten idiomien vastaavia. Sen suosituksenvastaisen käytön osuus on selvästi vertailuryhmän pienin. Idiomin *olla pahoillaan* kohdalla nähdään tilastollinen merkitsevyys samojen aikajaksojen välillä kuin *olla huolissaan* -idiomillakin eli suosituksenvastainen käyttö näyttää lisääntyneen ensimmäisen vertailujakson jälkeen, muttei enää tilastollisesti merkitsevästi toisen ja kolmannen jakson välillä. Sen khiin neliö -arvo $\chi^2(2)=9,890$, $p=0,007$ sekä Cramérin V-arvo 0,014 kertovat ensimmäisen idiomin tapaan tilastollisesti merkitsevän eron olemassaolosta, mutta myös sen pienuudesta.

Tutkimuksen kolmas idiomi *olla innoissaan* on kokonaiskäytöltään selvästi toisia pienempi. Sen kokonaiskäyttö on viidesosa *olla pahoillaan* ja noin neljäsosa *olla huolissaan* -idiomien vastaavista. Sen osalta tilastollisesti merkitseviä muutoksia ei vertailujaksojen välillä nähdä eli käytön ei voida sanoa lisääntyneen, muttei vähentyneenkään. Tämä johtunee ainakin osittain suhteellisen pienestä frekvenssistä ja osittain pienestä muutoksesta suosituksenvastaisen käytön osuudessa. Tilastollisen analyysin numeeriset arvot *olla innoissaan* -idiomin osalta ovat khiin neliö $\chi^2(2)=2,431$, $p=0,297$ ja Cramérin V-arvo 0,016.

Persoonakongruenssin puuttumista eli suosituksenvastaisen käytön yleisyyttä eri idiomien kesken vertailtiin niin ikään SPSS-ohjelman ristiintaulukointityökaluilla. Tilastolliset analyysit tehtiin niin kokonaistasolla kuin tutkimuksen kolmen vertailujaksonkin puitteissa. Tuloksista havaitaan, että kahden viimeisen vertailujakson puitteissa ja koko korpuksen tasolla kaikkien kolmen idiomin suosituksenvastaisen käytön osuudet eroavat tilastollisesti merkitsevästi toisistaan. Toisin sanoen toisen ja kolmannen aikajakson kuluessa *olla huolissaan* -idiomia käytetään suosituksenvastaisesti eniten ja *olla pahoillaan* -idiomia vähiten, mutta merkitsevät erot nähdään myös

kummankin edellä mainitun ja niiden suosituksenvastaisen käytön osuudessa sijoittuvan *olla innoissaan* -idiomin välillä.

Ensimmäisen vertailujakson tuloksissa *olla huolissaan* ja *olla innoissaan* -idiomien välillä ei ole tilastollisesti merkitsevää eroa, mutta niiden ja *olla pahoillaan* -idiomin välillä on. Näin ollen voidaan sanoa, että *olla pahoillaan* -idiomi on säilyttänyt lauseen subjektin kanssa kongruoivan omistusliitteen tiiviimmin kuin toiset tutkitut idiomit, mutta myös *olla innoissaan* muodostetaan kahdella viimeisellä vertailujaksolla kolmesta useammin suosituksenmukaisen omistusliitteen kanssa kuin näistä kolmesta idiomista frekvenssiltään suurin *olla huolissaan*. Kokonaistasolla kaikkien idiomien välillä nähdään siinäkin tilastollisesti merkitsevä ero, mutta se on suhteellisen pieni kuten muidenkin aikajaksojen kohdalla. Tämän vertailun tilastolliset arvot on esitetty taulukossa 8.

Taulukko 8. Idiomien jaksovertailujen tilastolliset arvot.

Aikajakso	$\chi^2(2)$	p-arvo	Cramérin V
2001–2006	333,426	p<0,001	0,105
2007–2011	487,615	p<0,001	0,110
2012–2017	405,147	p<0,001	0,112
Koko korpus	1264,088	p<0,001	0,112

Koska kyseessä on vain kolme idiomiä käsittävä tutkimus ja koska korpuksen muiden idiomien käyttöfrekvenssit ovat huomattavasti tutkittuja alhaisemmat, ei liene tarkoituksenmukaista vetää pitkälle meneviä johtopäätöksiä siitä, miten käyttöfrekvenssi vaikuttaa suosituksenvastaisen käytön yleisyyteen. Yhtä vähän voi yleistää tuloksia erottelemaan sisä- ja ulkopaikallissijoilla muodostettavia olotilaa ilmaisevia verbilausekeidiomeja. On tärkeää huomata, että kaikilla tutkituilla idiomeilla suosituksenvastainen käyttö muodostaa alle viiden prosentin osuuden idiomien kokonaiskäytöstä yksikön ensimmäisessä persoonassa. *Olla pahoillaan* -idiomilla se on kaikilla vertailujaksoilla selvästi alle yhden prosentin eli varsin harvinainen ilmiö, vaikka idiomiä käytetäänkin varsin paljon yksikön ensimmäisessä persoonassa tai ehkä juuri siksi.

4 Pohdinta

Tutkimuksen tavoitteena oli selvittää omistusliitteen persoonakongruenssia eri olotilanilmausidiomien käytössä eli miten yleinen ilmiö kongruenssittomuus on ja onko suosituksenmukainen persoonakongruenssi vähenemässä. Lisäksi oli tarkoitus selvittää, onko eri idiomien välillä eroja siinä, miten hyvin suosituksenmukainen omistusliite on säilynyt niiden käytössä. Tutkittaviksi valikoituneet idiomit *olla huolissaan*, *olla pahoillaan* ja *olla innoissaan* ovat kaikki sukulaiskonstruktioita niin syktaktisesti kuin semanttisestikin.

Tutkimusmetodiksi valittiin korpustutkimus ja tarjolla olleista korpuksista Suomi24-keskustelupalstan viestit tietyltä aikaväliltä käsittävä Kielipankin korpus, koska kielen muutos tapahtuu pääasiassa puhekielessä ja keskustelupalstalla käytettävä kieli on kirjoitettunakin monien palstan käyttäjien kohdalla pitkälti puhekielenomaista. Suomi24-korpus tarjosi myös laajuutensa puolesta potentiaalia löytää materiaalia tällaisesta kuitenkin suhteellisen harvinaisesta ilmiöstä.

Tutkimuksessa selvitettiin suosituksenvastaisen käytön prosenttiosuus kokonaiskäytöstä kolmen frekvenssiltään yleisimpiin olotilanilmausidiomeihin kuuluvan konstruktion osalta sekä analysoitiin tilastollisesti niissä tapahtuneet muutokset sekä eri idiomien väliset erot. Tutkimuksen tulokset kertovat tietysti edellä mainituista seikoista ainoastaan kyseisellä keskustelupalstalla korpuksen materiaalin tuottamisajankohtana ja johtopäätökset, joita tuloksista voidaan vetää käyttäjien puhekieliseen ilmaisuun ovat rajallisia, mutta jotain viitettä ne siitä mitä luultavimmin antavat.

Tutkimuksen keskeiset havainnot olivat, että tutkituista idiomeista kahden useimmin tavattavan osalta suosituksenvastaisessa eli persoonan mukaan kongruoimattoman omistusliitteen käytössä nähtiin osittain tilastollisesti merkitsevää kasvua, mutta se oli kuitenkin varsin heikkoa ja suosituksenvastainen käyttö ylipäätään oli suhteellisen harvinaista. Tämän lisäksi havaittiin, että *olla pahoillaan* -idiomin käytössä suosituksenmukainen omistusliite on ainakin yksikön ensimmäisessä persoonassa säilynyt toisia idiomeja kiinteämmin. Pilottitutkimuksen mukaan käyttäjät kertoivatkin todellisilla esiintymiskerroilla mitattuna huomattavasti useammin omasta pahoillaan olostaan kuin jonkun toisen tai joidenkin toisten.

Vaikka käsitys tutkittavan ilmiön yleisyydestä perustui ainoastaan tutkijan intuitioon, hypoteesina oli, että se on kasvava ilmiö. Tästä tilastollisessa analyysissä löydettiin hieman todisteita, mutta odotuksena oli, että ilmiö yleistyisi ainakin siinä määrin nopeammin, että kehityksessä nähtäisiin ainakin jonkin idiomin osalta tilastollisesti merkitsevä ero kaikkien aikajaksojen välillä. Tämä ei kuitenkaan toteutunut, sillä käytetyimmälläkään idiomilla suosituksenvastaisen käytön kasvu ei ollut tilastollisesti merkitsevä kahden viimeisen jakson välillä.

Toisen tutkimuskysymyksen suhteen oletamus oli, että ilmiön suhteen on idiomien kesken eroja ja loogisesti lähinnä tai ainakin toisia tutkittuja idiomeja selvästi useammin yksikön ensimmäisessä persoonassa käytettävä idiomi *olla pahoillaan* olikin menettänyt suosituksenmukaista omistusliitettään vähiten. Tutkittavan ilmiön harvinaisuus ylipäättään oli jossain määrin yllätys. Johtuuko tämä ilmiön todellisesta harvinaisuudesta kielenkäytössä yleisesti vaiko siitä, että tämän puhekielenomaisen tekstilajin puhekielenomaisuus vaihtelee voimakkaasti käyttäjästä toiseen, on myös mielenkiintoinen, mutta tässä yhteydessä avoimeksi jäävä kysymys.

Moni kirjoittaja tuottaa keskustelupalstoilla hyvin huoliteltua yleiskieltä, vaikka käyttäisikin suosituksenvastaisia muotoja puhuessaan ja yksilön puhuttuakin kieli vaihtelee eri rekisterien välillä riippuen siitä, kenen kanssa ja missä yhteydessä hän kommunikoi. Voitaneen todeta, että vaikka yksilö käyttäisi ainoastaan moitteetonta yleiskieltä kirjoittaessaan keskustelupalstalle, on mahdollista, että hän saattaa käyttää vähemmän huoliteltua kieltä puhuessaan.

Jos yksilö käyttää suosituksenvastaisia muotoja kirjoittaessaan, on vaikea kuvitella, ettei niitä esiintyisi hänen puhuessaan. Tämä ei tarkoita sitä, etteikö kyseinen yksilö kykenisi halutessaan tuottamaan myös täysin korrektia puhekieltä, mutta yleisellä tasolla kirjoitetussa kielessä esiintyvä yleiskielen ohjeistuksesta poikkeava ilmiö, esiintyy hyvin todennäköisesti myös puheessa: onhan kirjoitetun kielen tehtävä pääosin kuvata puhuttua kieltä, jota on totuttu pitämään kielen pääasiallisena muotona suhteellisen nuoreen kirjoitettuun muotoon verrattuna. Kielentutkimuksen konsensuksen mukaan kielen muutokset tapahtuvat pääosin ensin juuri puhutussa kielessä, josta osa sitten siirtyy

aikanaan kirjoitettuun kieleen, josta vielä pienempi osa lopulta hyväksytään kielenhuoltoelimien siunauksella yleiskieleen.

Tutkimusasetelmaan ja -menetelmään liittyvistä rajoituksista ehkä merkittävin oli se, ettei Suomi24-aineisto valitettavasti mahdollista perehtymistä kirjoittajiin nimimerkkien takana. Näin ollen ei ole mahdollista tutkia kielenkäyttöä esimerkiksi eri ikäryhmissä tai asuinmaakunnan mukaan vaan tarkastelua voi tehdä ainoastaan kokonaistasolla. Toki tämä seikka oli tutkijan tiedossa jo tutkimusta suunnitellessa. Se, että *olla huolissaan* -idiomin suosituksenvastaisesta käytöstä reilu viidesosa tapahtuu keskustelupalstan aihealueella ”Yhteiskunta” tai suosituksenvastaisesti pahoillaan ollaan kolme kertaa kymmenestä ”Suhteet”-aihealueella, on sinänsä mukavaa tietää, mutta tuo tieto ei välttämättä ole kovinkaan hyödyllistä kielen muutosta pohtiessa.

Taustatietojen puute johtuu tietysti osittain siitä, että tekstejä ei alun perin ole kerätty korpuksen luomista varten vaan yksityishenkilöt ovat kirjoittaneet niitä keskustelupalstalle anonymiteetin takaa. Toisaalta tämän tutkimuksen ilmiöiden frekvenssit olivat sen verran matalia, ettei ainakaan tilastollisesta analyysistä olisi ollut kovinkaan todennäköistä löytää merkitseviä eroja edes perinteisten sukupuolten välillä eli vain kahteen ryhmään jaoteltuna.

Minkälaisista aiheista keskusteltaessa ollaan asioista huolissaan, pahoillaan tai innoissaan ja minkälaiset ihmiset niistä aiheista käytäviin keskusteluihin osallistuvat? Ensimmäiseen kysymykseen olisi Suomi24-korpukselta saadun datan perusteella mahdollista vastata, sillä se kertoo kustakin hakuosumasta aihealueen kuten esimerkiksi *ajoneuvot ja liikenne*, *harrastukset* tai *muoti ja kauneus*, sekä siitä vielä tarkemman aiheen kuten viimeksi mainitun alta valittavat *hiukset*, *kengät* tai *meikkaus*. Tämä sinänsä mielenkiintoinen aspekti jää kuitenkin tämän tutkimuksen rajauksen ulkopuolelle. Minkälaiset ihmiset näistä keskustelevat, on kysymys, johon korpusdata ei tarjoa käyttäjien nimimerkkejä tarkempia vastauksia. Se, käytetäänkö suosituksenvastaisia muotoja enemmän muutenkin puhekielenomaisempien viestien osana vai esiintyykö niitä yhtä lailla palstan lähellä yleiskieltä olevissa teksteissä, jää tämän tutkimuksen rajauksen ulkopuolelle, mutta on toki mielenkiintoinen aspekti jatkotutkimukselle.

Muita mielenkiintoisia kysymyksiä, joihin ei valitettavasti saada vastauksia datasta ovat esimerkiksi minkälainen koulutustausta eri keskusteluihin osallistuvilla ihmisillä on ja vaikuttaako ihmisen kiihtyneisyyden aste siihen, miten huoliteltua kieltä hän käyttää? Entä aiheuttaako keskustelupalstan anonymiteetti huolimattomampaa kielenkäyttöä, kun keskustellaan useimmiten tuntemattomana tuntemattomien kanssa? Miten Suomi24-keskustelupalstan suosio eri ihmisryhmissä on kehittynyt vuosien aikana ja miten uusien innovaatioiden ja alustojen lanseeraus ja tietyissä tapauksissa räjähdysmäinen suosion kasvu on vaikuttanut Suomi24:n käyttöön?

Onko palstan tyypillinen käyttäjäprofiili muuttunut kuluneiden vuosien aikana ja onko sillä merkitystä tämän tutkimuksen tuloksiin? Toisin sanoen onko mahdollista, että tutkimustulos osoittaa käyttäjäkunnan muutoksen takia suosituksenvastaisten muotojen käytön hienoista kasvua, vaikka vastaavaa kehitystä ei palstan ulkopuolella olisikaan tapahtunut?

Tutkimuksen idiomit ovat selkeästi vakiintuneita muodon ja merkityksen yhteenliittymiä, kuten konstruktion määritelmässä (esim. Goldberg 1998, 2005; Hoffmann ja Trousdale 2013) sanotaan. Ne vaativat tietyn rakenteen välittääkseen tietyn merkityksen. Tutkimuksen kohteena olevat kolme konstruktioita vaativat kaikki jälkimmäisen komponentin esiintymisen tietyssä paikallissijassa sekä monikossa ja niihin pitää liittyä omistusliite.

Vaikka suosituksenmukainen subjektin kanssa kongruoiva omistusliitteellinen muoto osittain onkin korvautunut persoonan mukaan mukautumattomalla muodolla, jälkimmäisessäkin muodossa omistusliite on välttämätön, jotta kyse edelleen olisi samasta konstruktioista, joka edelleen välittäisi sille kuuluvan merkityksen. Kaikkien tutkimuksen idiomien käyttöfrekvenssit kertovat siitä, että ne täyttävät myös Goldbergin (2005) päivitetyn määritelmän vakiintuneisuudesta.

Edistyneet kielenkäyttäjät pystyvät prosessoimaan ja käyttämään näitä konstruktioita niiden komponenttien osamerkityksiä miettimättä. Idiomien tunnusmerkit (esim. Nenonen 2002) ne täyttävät olemalla konventionaalisia, epäkompositionaalisia, syntaktisesti ja leksikaalisesti jähmeitä ja sanaa pitempiä yksiköitä. Niiden merkitys ei suoraan selviä

osien summana vaan ymmärtääkseen mitä niillä viestitään kielenkäyttäjän pitää tuntea ne ennalta. Jähmeys toteutuu siinä mielessä, että niiden muoto on tarkoin määritelty eikä niiden jälkimmäisinä osina toimivia sanoja voi vaihtaa toisiin aiheuttamatta vähintään vivahde-eroja merkityksiin. Sanaa pitempiä yksiköitä ne ovat, koska ne koostuvat kahdesta sanasta, joista *olla* toki toisinaan saattaa jäädä pois puhuttaessa tai kirjoitettaessa, mutta on kuitenkin läsnä idiomien merkityksessä. Kaikkien kolmen on helppo nähdä kehittyneen metaforista kuten Mikone (2000) idiomien syntyprosessia kuvaa ja kuten tämän tutkimuksen alaluvussa 2.2 esitettiin.

Kaikkien kolmen tutkitun idiomien kohdalla ohjeistus on selkeästi se, että omistusliitteen tulee kongruoida eli mukautua lauseen subjektin mukaan (esim. ISK §96, §1267, Kielitoimiston ohjepankki, Korhonen 2016). Se, että vain yhdellä idiomilla ja vain yhtenä vuotena seitsemäntoista vuotta käsittävässä korpuksessa suosituksenvastaisen käytön osuus ylittää viiteen prosenttiin, ja se, että tiukimmin suositusta seuraavan idiomien suosituksenvastainen käyttö ei minään vuonna ylitä yhden prosentin osuutta, kertoo, että suositusta noudatetaan palstan kirjoittajien keskuudessa varsin hyvin.

Tässä tapauksessa suosituksenvastainenkin käyttö sisältää omistusliitteen, vaikka se ei oikeassa persoonassa esiinnykään. Näin ollen on vaikeaa nähdä omistusliitteen kokonaan katoavan tutkituista idiomikonstruktioista tai voidaan ennustaa, että omistusliitteen kadotessa katoaa koko idiomiksi, sillä idiomien merkitys ei välity ilman, että se seuraa kaikkia sen muotoon liittyviä kieliopista riippumattomia sääntöjä. Se, kokevatko kielenkäyttäjät leksikaalistuneen omistusliitteen adverbinmuodostuksessa käytettäväksi päätteeksi kuten Markkula (2016) ehdottaa, jää muiden tutkimusten selvitettäväksi.

Yksi Suomi24-korpuksen hyvistä puolista on havainnoijan paradoksia (Feagin 2006) ja tätä tutkimusta ajatellen on se, kirjoittajat eivät ole kirjoittaneet tekstejä tietoisina siitä, että ne tulevat analysoitaviksi tai ainakin he vuonna 2015 tapahtuneen Suomi24-korpuksen julkaisun jälkeenkin ovat voineet luottaa siihen, että kirjoituksen säilyvät anonymeinä. Tämän tutkimuksen resurssit huomioiden korpus on niin lähellä puhekieltä kuin mahdollista kuten Scheiderin (2006) suosituksissa mainittiin. Niin ikään korpuksen suuri koko on myös sen ehdoton vahvuus ja se mahdollistaa Schneiderin mukaan

suhteellisen harvinaistenkin ilmiöiden, jollaiselta tämän tutkimuksen aihekin vaikuttaa, tutkimuksen.

Huonoja puolia Suomi24-korpuksessa on tietysti se, että kirjoittajista ei nimimerkkien lisäksi tiedetä mitään eli tuloksia voidaan tarkastella ainoastaan kokonaistasolla. Myös olettamus siitä, että tekstien kirjoittajat käsittävät molempia sukupuolia, eri ikäryhmiä eri koulutustasoilta on vain olettamus. Palstan käyttäjistä on kerätty taustatietoja, mutta niitä ei voi yhdistää yksittäisiin viesteihin ja osa käyttäjistä saattaa vieraila palstalla vain lukemassa toisten tuottamia tekstejä.

Kuitenkin Suomi24-keskustelupalstan käyttäjät ja näin ollen todennäköisesti myös Suomi24-korpuksen tekstien kirjoittajat edustavat eri sukupuolia, ikäryhmiä ja varmasti jossain määrin myös eri ammattiryhmiä, tulo- ja koulutustasoja, poliittisia kantoja ja niin edelleen. Siinä mielessä korpus on hyvä työkalu kielitieteelliseen tutkimukseen. Ryhmä, joka varmuudella kokonaan puuttuu, on tietokonetta ja internetiä syystä tai toisesta käyttämättömät henkilöt. Korpus täyttää myös systemaattisen kokoamisen vaatimuksen (esim. Weisser 2016), koska se sisältää kaikki palstalla tietyn ajanjakson kuluessa julkaistut viestit.

Suuresta volyymistään huolimatta tekstit ovat kielenmuutoksen mittapuulla suhteellisen lyhyeltä aikajaksolta. Tekstit ovat myös kaikki samaa tyyppiä eli puheenvuoroja yksittäisellä keskustelupalstalla. Tältä osin Schneiderin (2006) suositus useista eri tekstilähteistä ei toteudu. Tämä luonnollisesti rajoittaa johtopäätösten tekemisen koskemaan ainoastaan kyseisellä keskustelupalstalla ilmoitetun aikajakson aikana julkaistuja puheenvuoroja.

Hakutulosten analysoimisessa KORPin tuottama konteksti, eli virke ennen ja jälkeen varsinaisen hakutuloksen, on monesti välttämätön, jotta puhekielenomaisen ja usein hyvin puutteellisesti välimerkkejä sisältävän kielen tavoitelluista muodoista voi olla varma. Vaikeimmissa tapauksissa alkuperäiseen tekstiin Suomi24-palstalla voi päästä käsiksi Google-haun avulla, mutta täyttä varmuutta ei jokaisesta osumasta välttämättä saa, vaikka näkisi koko alkuperäisen viestinkin.

Automaattinen annotointi tuottaa alaluvussa 3.2.2.3 kuvatun kaltaisia tilanteita, jotka ilman hakutulosten manuaalista läpikäyntiä johtaisivat lukuisiin virheellisiin hakuosumiin datassa. Laajempaa idiomien kartoittamista yleispätevämmillä hakulausekkeilla hankaloitti myös se, että esimerkiksi *huolissani* ja *innoissani* on KORPissa analysoitu substantiivien ja *pahoillani* adjektiivin taivutusmuodoiksi, mutta muodot *huolissaan*, *innoissaan* ja *pahoillaan* puolestaan adverbeiksi. Jos *huolissani*-tyypin sanan sijasta olisi määritelty haettavaksi substantiivi monikon paikallissijassa omistusliitteellä olisi hakutuloksiin osunut esimerkiksi teksti "olen maininnut joissain kirjoitelmissani", joka ei sisällä idiomia. Tämä osoitti todeksi McEneryn ja Hardien (2012) huomion siitä, että mikään annotointitapa ei tuota virheitä täysin eliminoivaa lopputulosta.

Vaikkei tässä aivan appelsiineja ja omenoita verratakaan, on syytä kuitenkin muistaa jo muutamaan kertaan mainittu seikka, että vaikka keskustelupalstan kieli on monen käyttäjän kohdalla lähellä puhekieltä, se ei kuitenkaan ole kertakäyttöistä puhekieltä. Moni käyttäjä käyttää nimimerkin takaakin huoliteltua yleiskieltä tilanteessa, jossa – toisin kuin puhutussa kielessä – hänen tuottamansa kieli ei tallennu vain kuulijan muistiin vaan lähes kenen tahansa, missä tahansa ja milloin tahansa luettavaksi.

Se, että idiomien persoonittain mukautuva omistusliite ei ole tämän enempää kadonnut epävirallisesta yhteydestä, jota keskustelupalsta kiistämättä edustaa, on merkki siitä, ettei se kovin nopeasti tule katoamaan virallisesta yleiskielestäkään. Samaan suuntaan viittaisi myös edellä käsitelty Korhosen ja Lappalaisen (2013) tutkimustulos, jonka mukaan valtaosan mielestä minkäänlaiset omistusliitteettömät muodot eivät kuulu asiatekstiin.

Tämän tutkimuksen tulosten reliabiliteetti on hyvä, kun pidetään mielessä, että ne kertovat vain Suomi24-keskustelupalstalla vuosina 2001–2017 tapahtuneesta kielenkäytöstä ja -muutoksesta edellä mainittuine rajoituksineen. Suomi24-korpuksen vahvuus on se, että se käsittää kaikki palstalle kirjoitetun viestit tarkkaan määriteltynä ajanjaksona eli tutkimuksen korpusajot ovat kartoittaneet tutkittujen muotojen kaikki löydettävissä olevat esiintymät. Toinen merkittävä vahvuus on korpuksen koko, joka mahdollistaa suhteellisen harvinaistenkin ilmiöiden osalta riittävät frekvenssit määrälliseen ja tilastollisia analyysejä hyödyntävään tutkimukseen. Tilastollinen analyysi

on tehty tunnetulla ja laajasti käytetyllä työkalulla, mikä sekin lisää tutkimuksen luotettavuutta.

5 Loppupäätelmät

Paunonen ennusti vuonna 1995, että omistusliite katoaa Helsingin puhekielestä 2040-luvulla eli noin 50 vuoden kuluessa siitä hetkestä. Samassa artikkelissa hän lainasi vuoden 1966 yleisönosastokirjoitusta, jossa kirjoittaja oli huolissaan omistusliitteen käytöstä. Tämä kasvattaa prosessin aikajänteen noin 80 vuoteen. Huomioitavaa on, että prosessi oli siis jo vuonna 1966 sellaisessa vauhdissa, että siitä oltiin huolissaan lehtien palstoilla, mikä viittaa siihen, että se oli jo tuolloin ollut käynnissä jo hyvän tovin.

Vaikka maassamuutto ja liikenne ynnä muut vielä modernimmat ilmiöt ovat monissa tapauksissa kiihdyttäneet kielen muutosprosesseja, voidaan silti todeta, että Suomi24-korpuksen kattama 17 vuoden aikajakso on kielenmuutoksen kannalta hyvin lyhyt. Sinänsä vähäinenkin tilastollisesti merkitsevä muutos – jollaisia tässä tutkimuksessa nähdään – kertoo, että muutosta tapahtuu kuten teoriaosuuden kielenmuutosta käsittelevät lähteet käytännössä yksimielisesti julistivatkin. Tässä tutkimuksessa tutkitun ilmiön eli olotilanilmausidiomien kongruoimattoman omistusliitteen syntyäikää ei valitettavasti pystytty selvittämään, sillä se oli osa kielenkäyttöä jo korpuksen ensimmäisten tekstien julkaisuaikaan.

Kielenhuolto ja normittaminen ovat välttämättömiä, jotta ihmiset voisivat jatkossakin ymmärtää toisiaan. Kieli muuttuu hitaasti, mutta riittävän pitkän ajan kuluessa heikosti ohjaillusta kielestä voisi periaatteessa muodostua variantteja, joiden puhujat eivät enää ymmärtäisi toisiaan kuten historiassa on käynytkin. Läheisinkään suomen sukulaiskielistä, viro, ei monista yhteisistä piirteistä huolimatta ole helposti ymmärrettävissä suomea äidinkielenään puhuville.

Toki tänä päivänä yhteydenpidon helppous pitkienkin matkojen päähän sekä painotuotteiden – myös digitaalisten tekstien – ja audiovisuaalisten viestinten tarjonnan saatavuus saattaa esimerkiksi suomen kielen kohdalla ennemminkin ohjata kieltä yhtenäisempään suuntaan kuin aikaisemmin on ollut mahdollista. Toisaalta juuri samat

tekijät helpottavat samanhenkisten ihmisten ryhmäytymistä ja kielenkäytön erityispiirteiden leviämistä pienissäkin alakulttuureissa.

Kaikkien tässä tutkimuksessa tutkittujen idiomien suosituksenvastainen käyttö on yhden idiomien yhtä yksittäistä vuotta lukuun ottamatta alle viisi prosenttia eli persoonakongruenssi tämän tyyppisten idiomien käytössä tuskin katoaa lähitulevaisuudessa. Toisaalta kuten aikaisemmin keskusteltiin, päätteiden katoamisen tai leksikaalistumisen vain yhteen vaihtelemattomaan muotoon voitaisiin nähdä vievän kieltä tehokkaampaan suuntaan, joka jättää vähemmän väärinymmärtämisen mahdollisuuksia.

Tämä tutkimus kertoo idiomien käytöstä ainoastaan Suomi24-keskustelupalstan käyttäjien keskuudessa ja heidänkin osaltaan vain palstalle kirjoitettaessa, mutta sen perusteella voidaan tehdä oletuksia ilmiön yleisyydestä puhekielessä. Yksittäisen kirjoittajan puhutun ja kirjoitetun kielen välillä voi olla suuriakin eroja. Puhuttu kieli myös eroaa eri puolilla Suomea ja näin ollen tämän tutkimuksen kartoittama ilmiö saattaa olla Kemissä aivan eri vaiheessa kuin Kaarinassa ja Kajaanissa taas eri vaiheessa kuin Kouvolassa. Tässä korpuksessa kaikki ovat kuitenkin samanlaisia palstan käyttäjiä asuinpaikasta riippumatta.

Vaikka suosituksenvastainen muoto särähtääkin pahasti osan kielikorvaan ja vaikka sitä Suomi24-keskustelupalstan kirjoitusten mukaan ovat käyttäneet niin kalliolaissyntyinen presidentti kuin huomattavasti pohjoisemmassa Suomessa pääosan elämästään vaikuttanut pääministerikin, ilmiö vaikuttaa kuitenkin edelleen melko harvinaiselta. Toki on huomioitava, ettei tämän tutkimuksen ole ollut mahdollista kartoittaa tämän tyyppisten ilmausten suosituksenvastaisen käytön osuutta puhekielessä vaan ainoastaan keskustelupalstalla käytetyssä kielessä. Se, ovatko Halonen ja Sipilä aktiivisia käyttäjiä Suomi24-palstalla ja miten heidän idiomimuotonsa siellä seuraavat suosituksia ja kongruoivat lauseen subjektin kanssa, jää myös tämän tutkimuksen rajauksen ulkopuolelle.

Tutkimuksen kolme idiomia vaikuttavat käyttäytyvän hieman eri lailla sen suhteen, miten tiukasti omistusliite niissä istuu. Otos on kuitenkin liian pieni siihen, että asia olisi yleistettävissä sisä- ja ulkopaikallissijoissa tai inessiivissä ja adessiivissa esiintyvien

idiomien suhteen. Ilmiö on sen vielä sen verran harvinainen, ettei edes Suomi24-korpuksen yli 4 miljardia sanetta riitä sen perusteellisempaan selvittämiseen ainakaan vain yksikön ensimmäisen persoonan osalta. Lisäksi on syytä muistaa se, että moni keskustelupalstojen käyttäjä pyrkii edelleen kirjoittaessaan korrektimpaan kieleen kuin puhuessaan eli suosituksenvastainen muoto oikeassa puhekielessä eikä vain sitä muistuttavassa keskustelupalstakielessä saattaa olla huomattavastikin yleisempi.

Kielen pyrkiessä parempaan tarkoituksenmukaisuuteen suomen omistusliite vaikuttaa ehkä hankalammalta tai tehottomammalta tavalta kertoa omistamisesta kuin viron tai meitä lähellä puhuttavien germaanisten kielten vastaavat omistusrakenteet. Possessiivipronomini ja omistusliite näyttävät nopeasti ajateltuna toimittavan samaa tehtävää ja usein niin onkin: asia voidaan helposti ilmaista vain toista käyttämällä esimerkiksi *haen palloni*.

Omistusliitettä vastaan puhuu ehkä sekin seikka, että sen käyttäminen on hankalampaa: ensimmäisessä ja toisessa persoonassa omistaminen käy ilmi omistusliitteestä, mutta kolmannen persoonan luvun suhteen neutraali muoto vaatii joka tapauksessa avukseen possessiivipronominin, jotta tiedetään, onko omistajia yksi vai useampia. Omistusliite tuntuu vain vaikeuttavan muutenkin hankalaa sanojen taivutusta. Kun possessiivipronomini joka tapauksessa tarvitaan, miksei se voi toimittaa omistuksenosoittajan virkaa kaikissa persoonissa? Myös kaikkialla läsnäolevien valtakielen paine saattaa olla osaltaan työntämässä suomea kohti omistusliitteettömiä muotoja.

Omistusta korostettaessa molempien muotojen käyttö on perusteltua esimerkiksi *Se on minun palloni* tai *Se oli sinun vikasi*. Tällaisessa tapauksessa vain lauseen merkitystä ajatellen possessiivipronominin muoto *Se on palloni* tai *Se on vikasi* vaikuttaa kömpelöltä, kun taas omistus käy helposti ilmi ilman possessiivikongruenssia *Se on minun pallo* tai *Se on sinun vika*. Ei ole vaikeaa ymmärtää, miksi näitä muotoja pääosan elämästään kuulleet ja käyttäneet nuoret eivät välttämättä kaipaa omistusliitteitä kieleensä. Toisaalta kuten Y. Lumio kirjoitti mielipideosastolla jo vuonna 1966 monen korvaan moiset muodot kuulostavat kammottavilta.

Tämän tutkimuksen käsittelemissä verbilausekeidiomeissa omistusliitteen totaalinen katoaminen vaikuttaa kuitenkin vielä kaukaiselta. Vaikka suosituksenvastainen kolmannen persoonan muoto on vallannut alaa ainakin yksikön ensimmäisen persoonan käytössä kuten esimerkiksi *olen huolissaan*, ilmaus sisältää omistusliitteen vaikkei se kongruoikaan subjektin kanssa. Muotoa *olen minun huolissa* tuskin tavataan ja vaikka tavattaisiinkin, kielenkäyttäjiltä menisi varmasti hyvä tovi yhdistää se *olla huolissaan* -konstruktion idiomaattiseen merkitykseen.

Se, tavataanko kolmannen persoonan kivettynyttä omistusliitettä myös muualla kuin idiomeissa esimerkiksi *minun pallonsa* tai *teidän pallonsa* vaikuttaa pelkän intuition pohjalta epätodennäköiseltä, mutta sen selvittämiseen tarvitaan toki oma tutkimuksensa. Todennäköisempiä muotoja ovatkin tässäkin työssä monesti esille tulleet *minun pallo* tai *teidän pallo*.

Vaikka tämän tutkimuksen mukaan näiden kolmen idiomien leksikaalistuminen siten, että omistusliite kivettyy kolmanteen persoonaan muissakin persoonissa, on ainakin yksikön ensimmäisessä persoonassa vielä melko vähäistä, monessa lähteessä asiasta on keskusteltu ja todettu näin tapahtuvan puhekielessä. Tutkimustulokset tukevat aikaisempien tutkimusten havaintoja suosituksenvastaisten muotojen olemassaolosta ja jossain määrin myös niiden käytön lisääntymisestä, mutta liian pitkälle vietyjen johtopäätösten tekemistä kielenkäytöstä yleisellä tasolla on ilman lisätutkimusta viisainta välttää.

Tämä tutkimus antaa viitteitä ilmiön esiintymisestä ja kehityksestä suomen kielessä, mutta sitä olisi tarpeen kartoittaa isommassa mittakaavassa tulosten luotettavuuden parantamiseksi. Tutkimuksen laajentaminen vaatisi käyttöön useampia puhekielisiä tai puhekielenomaisia korpuksia tai vastaavasti internet-hakuja esimerkiksi Facebookista. Toinen mahdollisuus on kerätä lisää tietoa aiheesta olisi selvittää kivettyneiden omistusliitteen esiintymistä myös yksikön toisessa ja monikon ensimmäisessä ja toisessa persoonassa.

Myös idiomien määrää olisi laajemman korpuksen tai useampien korpusten tapauksessa mahdollista lisätä. Tärkein parannus tämän ilmiön tutkimiseen korpustutkimuksena olisi

kuitenkin se, jos siihen saataisiin dataa pitemmältä aikajaksolta. Isommat kieliresurssit mahdollistaisivat myös idiomien käyttöuseuden ja suosituksenmukaisen omistusliitteen välisen suhteen selvittämisen. Koska kielen innovaatiot tavataan pääsääntöisesti ensin informaalissa puhekielessä, josta osa niistä siirtyy aikanaan kirjoitettuun kieleen, ideaalitalanteessa tutkimusta tehtäisiin kuitenkin aitoa puhetta havainnoimalla ja siten, että saataisiin selvyyttä myös alueellisesta variaatiosta.

Lähteet

- Aitchison, Jean. 2001. *Language change: progress or decay?* Cambridge: Cambridge University Press.
- Chambers, J.K., Trudgill, Peter ja Schilling-Estes, Natalie (toim.). 2006. *The Handbook of Language Variation and Change*. Malden: Blackwell Publishing Ltd.
- Feagin, Crawford. 2006. "Entering the community: fieldwork". Teoksessa Chambers, Trudgill ja Schilling-Estes (toim.) *The Handbook of Language Variation and Change*. Malden: Blackwell Publishing Ltd. s. 20–39.
- Fillmore, Charles. 1985. "Syntactic Intrusions and the Notion of Grammatical Construction". *Berkeley Linguistic Society* 11. s. 73–86.
- Haddington, Pentti ja Sivonen, Jari (toim.). 2010. *Kielentutkimuksen modernit klassikot: kognitiivinen ja fuktiionaalinen kielitiede*. Helsinki: Gaudeamus Helsinki University Press.
- Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta ja Alho, Irja. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hoffmann, Thomas & Trousdale, Graeme. 2013. "Construction Grammar: Introduction". Teoksessa Hoffmann, Thomas & Trousdale, Graeme. (toim.) 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. s. 1–13.
- Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. Harlow: Addison Wesley Longman Limited.
- Kielitoimiston sanakirja (KTS). 2018. Saatavissa <https://www.kielitoimistonsanakirja.fi/>. Helsinki: Kotimaisten kielten keskus. (viitattu 25.5.2019)
- Korhonen, Riitta ja Lappalainen, Hanna. 2013. "Kieli muuttuu – rikkautta vai rappiota". *Kielikello*. 4/2016. Saatavissa: <https://www.kielikello.fi/-/kieli-muuttuu-rikkautta-vai-rappiota->. (viitattu 26.5.2019).
- Korhonen, Riitta. 2016. "Omistusliitteiden liikakäyttöä". *Kielikello*. 3/2016. Saatavissa: <https://www.kielikello.fi/-/omistusliitteiden-liikakayttoa>. (viitattu 24.5.2019).
- Kotilainen, Lari ja Pekkarinen, Heli. 2002. "Konstruktioita rakentamassa". *Virittäjä*. 4/2002. s.584–585.
- Kotimaisten kielten keskus. *Kielitoimiston ohjepankki*. "Omistusliitteet: millaisiin sanoihin omistusliite voi kiinnittyä?". Saatavissa: <http://www.kielitoimistonohjepankki.fi/ohje/589>. (viitattu 23.5.2019).
- Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: Chicago University Press.

Lars Borin, Markus Forsberg ja Johan Roxendal. 2012. *Korp – the corpus infrastructure of Språkbanken*.

Leino, Jaakko. 2003. *Antaa sen muuttua*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Leino, Jaakko. 2010. "Adele E. Goldberg - Väitöskirjalla klassikoksi". Teoksessa Haddington & Sivonen (toim.) *Kielentutkimuksen modernit klassikot: kognitiivinen ja fuktionaalinen kielitiede*. Helsinki: Gaudeamus Helsinki University Press. s. 71–97.

Lähdemäki, Eeva ja Bertills, Yvonne (toim.). 2000. *Fennistica fausta in honorem Mauno Koski septuagenarii. Juhlakirja Mauno Kosken 70-vuotispäivän kunniaksi 21.2.2000*, Turku: Åbo Akademis tryckeri.

Markkula, Karoliina. 2016. "Olin huolissaan sinusta – vai huolissani?". *Kielikello*. 3/2016. Saatavissa: <https://www.kielikello.fi/-/olin-huolissaan-sinusta-vai-huolissani> (viitattu 20.5.2019)

McEnery, Tony & Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Mielikäinen, Aila. 1982. "Nykypuhesuomen alueellista taustaa". *Virittäjä* Vol 86 Nro 3.

Mikone, Eve. 2000. "Idiomit ja kieliopillistuminen". Teoksessa Lähdemäki ja Bertills (toim.) *Fennistica fausta in honorem Mauno Koski septuagenarii. Juhlakirja Mauno Kosken 70-vuotispäivän kunniaksi 21.2.2000*. Turku: Åbo Akademis tryckeri.

Muikku-Werner, Pirkko. 2011. "Idiomeja ihan sikana". *Tiede*. 8.11.2011. Saatavissa: https://www.tiede.fi/artikkeli/jutut/artikkelit/idiomeja_ihan_sikana (viitattu 25.5.2019)

Mustanoja, Liisa. 2011. *Idiolekti ja sen muuttuminen*. Tampere: Tampereen yliopisto.

Nenonen, Marja. 2002. *Idiomit ja leksikko. Lauseidiomien syntaktisia, semanttisia ja morfologisia piirteitä suomen kielessä*. Joensuu: Joensuun yliopisto.

Ojutkangas, Krista, Larjavaara, Meri, Miestamo, Matti ja Ylikoski, Jussi. 2016. *Johdatus kielitieteeseen*. Helsinki: Sanoma Pro Oy.

Paunonen, Heikki. 1995. "Puhesuomen muuttuva omistusmuotojärjestelmä". *Virittäjä*. 4/1995. s. 501–531.

Remes, Hannu. 2009. *Muodot kontrastissa: Suomen ja viron vertailevaa taivutusmorfologiaa*. Oulu: Oulu University Press.

Schilling-Estes, Natalie. 2006. "Field methods - Introduction". Teoksessa Chambers, Trudgill ja Schilling-Estes (toim.) *The Handbook of Language Variation and Change*. Malden: Blackwell Publishing Ltd. s. 17–19

Schneider, Edgar W. 2006. "Investigating Variation and Change in Written Documents". Teoksessa Chambers, Trudgill ja Schilling-Estes (toim.) *The Handbook of Language Variation and Change*. Malden: Blackwell Publishing Ltd. s.67–96.

Tähtinen, Juhani, Laakkonen, Eero & Broberg, Mari. 2011. *Tilastollisen aineiston käsittelyn ja tulkinnan perusteita*. Turku: Turun yliopiston kasvatustieteen tiedekunnan julkaisuja C:20.

Weisser, Martin. 2016. *Practical Corpus Linguistics: An introduction to corpus-based language analysis*. Chichester: John Wiley & Sons, Inc.

Viertö, Annastiina. 2009. "Ritolat ja vanhaset". *Kotimaisten kielten keskus: Nyt*. Saatavissa: [https://www.kotus.fi/nyt/kolumnit_artikkelit_ja_esitelmat/kuukauden_kielijuttu_\(2003_2011\)/ritolat_ja_vanhaset](https://www.kotus.fi/nyt/kolumnit_artikkelit_ja_esitelmat/kuukauden_kielijuttu_(2003_2011)/ritolat_ja_vanhaset) (viitattu 30.5.2019)

Zetterberg, Seppo. 2017. *Uusi Viron historia*. Helsinki: Suomalaisen kirjallisuuden seura.

Aineistolähteet:

Aller Suomi. *Suomi24-keskustelupalsta*. Saatavissa: <https://www.aller.fi/suomi24/> (viitattu 20.5.2019)

Aller Suomi. *Suomi24-tilasto*. Saatavissa: https://www.aller.fi/wp-content/uploads/2019/01/S24_Fiam_2018.pdf (viitattu 30.5.2019)

Kielipankin Suomi24 virkkeet -korpus (2017H2) (beta)

Tieteen termipankki 28.05.2019: *Nimitys:annotointi*. Saatavissa: <https://www.tieteentermipankki.fi/> (viitattu 16.5.2019)

Svensk sammanfattning – Ruotsinkielinen tiivistelmä

En studie om personkongruens vid idiomatiska uttryck som uttrycker tillstånd med hjälp av Kielipankkis Suomi24-korpus

Inledning

Precis som gällande allt annat i livet sker det hela tiden kontinuerlig och oavbruten förändring i människospråk. De ända språken som inte förändras – blir bättre eller sämre beroende på synvinkeln – är de döda språken det vill säga språken som ingen pratar som sitt modersmål. Man kan försöka styra språkförändringen men det är omöjligt att stoppa eller ens kontrollera den fullt ut. Det sker ändringar hela tiden, vissa snabbare och andra långsammare.

Det är en smaksak om språkförändringen är en sak som man bör vara entusiastisk för eller orolig över liksom det att i vilken mån man bör vårda språk och normera språkbruket. Mycket sträng normering och reglering kan leda till en situation där skriftspråk och talspråk liknar varandra relativt lite och det blir svårt att lära sig språket även för dem som lär sig det som sitt första språk. Å andra sidan studeras de stora världsspråken världen runt med lärare som inte har nämnda språken som sina modersmål så det är viktigt att det finns klara regler.

I denna pro gradu-avhandling undersöker jag språkförändring i det finska språket med närmare fokus på idiom som beskriver tillstånd och där det enligt rekommendationer bör finnas personkongruens mellan satssubjektet och det obligatoriska possessivsuffixet som hör till den andra idiomkomponenten. till exempel i uttrycket eller konstruktionen *olla huolissaan* (vara bekymrad). Ett exempel på en sådan konstruktion är idiomet *olla huolissaan* där i första person singularis krävs possessivsuffixet *-ni* d.v.s. *olen huolissani*.

Konstruktionsgrammatik

Konstruktionsgrammatikerna anser att grammatiken består av olika slags konstruktioner som är etablerade kombinationer av form och betydelse. Det finns till exempel ordkonstruktioner, idiomkonstruktioner som denna avhandling fokuserar sig på, komparativkonstruktioner, morfem, resultativkonstruktioner och så vidare.

Konstruktionsgrammatikens popularitet beror på möjligheten att beskriva all grammatik enligt ett och samma mönster vilket i sin tur gör det möjligt att utnyttja den i forskningen av mycket flera olika företeelser än de mer traditionella betraktelsesätten.

Idiom

Den konstruktionstyp som är den viktigaste för denna avhandling är idiomkonstruktionen. Vi möter idiom dagligen såväl i olika sorters diskussioner, nyheter, vetenskapliga studier som skönlitteratur och filmkonsten. Man kan inte veta eller resonera sig fram till ett idioms betydelse med kunskap om vad dess olika delkomponenter betyder. Med andra ord är ett idiom inte genomskeinligt till sin betydelse. Verbfrasidiom som består av ett verb och ett komplement är den klart vanligaste idiomtypen i finskan. För att förmedla sin betydelse måste idiom följa bestämda regler som inte beror på grammatik t.ex. i finskans idiom *olla huolissaan* (vara bekymrad eller orolig) måste komplementet förekomma i pluralis och inessiv och det måste innehålla ett possessivsuffix som enligt rekommendationerna bör kongruera med satssubjektet.

Kongruens och possessivsuffix

Kongruens är ett begrepp som betyder ett samband mellan ord och fraser i en konstruktion. De måste anpassa sig till varandra så att de bilda en grammatiskt enhetlig helhet. Man böjer olika komponenter i en konstruktion i kasus, numerus och/eller person och på det sättet ökar deras kohesion och visar att de hör ihop. Kongruens kan förekomma inom en sats eller sammansatta verbformer men också i nominalfraser eller substantivfraser. Personkongruensen i idiom som denna avhandling fokuserar sig på förekommer på det sättet att verbdelen böjs enligt subjektet och nomendelen får ett possessivsuffix som också väljs utgående från subjektet.

Possessivsuffixet är ett suffix som placeras efter ett substantiv för att ange vem dess ägare är. Possessivsuffixen i finskan är *-ni* i första person singularis och *-mme* i första person pluralis, *-si* i andra person singularis och *-nne* i andra person pluralis samt *-nsa* eller *-Vn* i tredje person som har samma suffix i både singularis och pluralis. Possessivsuffixet placeras efter kasusändelsen men före en enklitisk partikel t.ex. *pallonikin*. Possessivsuffixet i idiomtypen som denna avhandling behandlar skall enligt kongruera med subjektet t.ex. *olen pahoillani, olet huolissasi, olemme innoissamme*.

Språkvariation och förändring

Ett levande språk förändras hela tiden. Det är omöjligt att undvika det. Förändringen kan vara snabbt och synligt eller mycket långsamt och svår att upptäcka men den påverkar alla språkets delområden. Under de senaste åren har sortimentet texttyper blivit mycket varierande och mångskiftande. Med nya kommunikationsmedier har det uppstått en hel del nya behov att kommunicera och det har i sin tur lett till det att det har uppkommit en hel del nya textarter.

Det finns olika sätt att vy språkförändring. Man kan anse att språket degenererar och blir sämre eller att det går framåt och blir mer effektivt genom förändring. Ett tredje sätt som man kan närma sig språkförändring på är att tycka att det med tanke på degenerering och utveckling hålls på ungefär samma ställe eftersom det sker både positiva och negativa förändringar. Fast det finns människor som har mycket starka åsikter om saken är det rent lingvistiskt långt en smaksak vilka förändringar som är positiva och vilka negativa.

Korpuslingvistik

Korpuslingvistik är en metod där man använder korpusar för att analysera språk. En korpus är en stor och systematiskt samlad kollektion texter eller tal som man har samlat för att studera olika fenomen inom språket och att göra lingvistisk analys. I praktiken är alla korpusstudier i dag datorbaserade. Korpuslingvistikens tillämpningsområden inkluderar bland andra att studera språkets karaktär, beskriva dess strukturer, utreda dess användningssätt och användningssituationer.

I korpuslingvistiska studier försöker man besvara forskningsfrågor genom att göra sökningar i korpusar och inte genom att observera eller intervjua språkanvändare. Korpusstudier skiljer sig från majoriteten av andra lingvistiska forskningsområden på det sättet att de är inte koncentrerade på någon enstaka språklig aspekt utan det viktiga är olika metoder och procedurer som man kan använda till att tackla nästan vilket som helst lingvistiskt forskningstema.

Analys och resultat

Denna studie är närmast en korpusstudie och som material har jag valt korpusen Suomi24 virkkeet -korpus (2017H2) (beta) som är en av Kielipankkis (Språkbanken) korpusar.

Suomi24-korpusen innehåller alla diskussioner som har publicerats på diskussionsforumet Suomi24 mellan 2001 och 2017 och den är morfosyntaktiskt strukturerad. Allt som allt innehåller Suomi24-korpusen mer än fyra miljarder textord.

Jag gjorde mina korpussökningar med Kielipankkis webbaserade konkordansverktyg KORP och bearbetade sökresultaten och gjorde alla räkneoperationer med Microsoft Excel. Alla datatabeller och grafer är också gjorda i Excel. Den statistiska analysen gjorde jag med SPSS för att utreda om förändringar och skillnader mellan olika idiom var statistiskt signifikanta.

Till denna studie hör en liten förstudie som gjordes för att hitta de mest användbara finska idiomerna för huvudstudien. Till huvudstudien behövde jag idiom som skulle resultera i ett antal sökträffar som är tillräckligt stor för statistisk analys. Förstudien innehöll 25 vanliga finska idiom av rätt form som jag plockade ur Nenonens (2002) lista på mer än 1600 olika idiom.

Som resultat av förstudien valde jag idiom *olla huolissaan*, *olla pahoillaan* (vara ledsen) och *olla innoissaan* (vara ivrig) till huvudstudien. Alla dessa idiom består av verbet *olla* (vara) och andra idiomkomponenter (nomendelar) som böjs i pluralis och lokalkasus närmare bestämt inessiv (*huolissaan* och *innoissaan*) eller adessiv (*pahoillaan*) och som innehåller också possessivsuffixet som bör enligt rekommendationen variera beroende på satsens subjekt. Forskningen avgränsade jag att omfatta endast första person singularis.

Avhandlingens syfte var att studera huruvida de ovannämnda idiomkonstruktionernas icke-rekommendationsenliga användning med tanke på possessivsuffix har ökat under de sju åren som Suomi24-korpusen omfattar. Med andra ord har formerna av typen *olen huolissaan*, *olen pahoillaan*, *olen innoissaan* (i stället för rekommendationsenliga *olen huolissani*, *olen pahoillani*, *olen innoissani*) blivit vanligare under de senaste åren. Andra forskningsfrågan var huruvida det finns skillnader på icke-rekommendationsenlig användning mellan olika idiom.

För att kunna besvara den första frågan delade jag korpusen i tre jämförelseperioder:

2001–2006, 2007–2011 och 2012–2017. De siffror som jag ville jämföra var andelen icke-rekommendationsenlig användning av totalanvändningen av idiomerna. Dessa resultat finns samlade i tabell 1.

Tabell 1. *Idiomens 1:a person singularis personkongruens i Suomi24-korpusen*

Olla huolissaan				
Tidsperiod	kongruerar ej	kongruerar	total användning	kongr. ej %
2001–2006	289	9 710	9 999	2,89
2007–2011	615	15 788	16 403	3,75
2012–2017	542	12 532	13 074	4,15
Totalt	1 446	38 030	39 476	3,66
Olla pahoillaan				
Tidsperiod	kongruerar ej	kongruerar	total användning	kongr. ej %
2001–2006	52	17 003	17 055	0,30
2007–2011	95	19 598	19 693	0,48
2012–2017	74	14 278	14 352	0,52
Totalt	221	50 879	51 100	0,43
Olla innoissaan				
Tidsperiod	kongruerar ej	kongruerar	total användning	kongr. ej %
2001–2006	71	3 113	3 184	2,23
2007–2011	102	4 014	4 116	2,48
2012–2017	77	2 610	2 687	2,87
Totalt	250	9 737	9 987	2,50

Största delen av alla sökträffarna var rekommendationsenliga. Det fanns endast ett år av sjutton möjliga där ett av tre idiom producerade resultat där andelen icke-rekommendationsenlig användning var mer än fem procent. Med andra ord endast en av 51 möjligheter visade den rekommendationsenliga användningen vara mindre än 95 procent. Därför verkar det att personkongruensen vid idiom av denna typ inte håller på att försvinna i en nära framtid.

Olla huolissaan var det idiomet där det fanns mest icke-rekommendationsenligt bruk med högsta andel 4,15 procent under den sista tidsperioden och 3,66 procent i hela korpusen. *Olla pahoillaan* däremot används rekommendationsenligt oftare än 99 procent av all användning. Resultaten för *olla innoissaan* ligger mellan de andra två men även det idiomet används korrekt i över 97 procent av fallen. Tabellen visar dessa andelar för alla idiom under alla tidsperioder samt i hela korpusen.

Jag använde statistikprogrammet SPSS för att kontrollera huruvida ökningen av icke-rekommendationsenlig användning var statistiskt signifikant. Till det använde jag korstabellering och Pearsons χ^2 -test. Resultaten visar att det finns statistiskt signifikanta förändringar mellan första och andra tidsperioden för *olla huolissaan* och *olla pahoillaan* men förändringen mellan andra och tredje perioden är inte statistiskt signifikant för någondera. Det vill säga att icke-rekommendationsenlig användning har ökat mellan första och andra perioden men andelar har stabiliserats efter det. Det finns inga statistiskt signifikanta förändringar för idiomet *olla innoissaan*. Statistiska värden för denna analys finns i tabell 2.

Tabell 2. Statistiska värden från SPSS-analysen för idiom

Idiom	$\chi^2(2)$	p-värde	Cramérs V
Olla huolissaan	25,895	p<0,001	0,026
Olla pahoillaan	9,890	0,007	0,014
Olla innoissaan	2,431	0,297	0,016

Jag jämförde också andelen icke-rekommendationsenlig användning mellan de tre studerade idiomerna. Under den första tidsperioden finns det inte en statistiskt signifikant skillnad mellan *olla huolissaan* och *olla innoissaan*. Alla andra tidsperioder och hela Suomi24-korpusen visade statistiskt signifikant skillnad mellan idiomerna det vill säga *olla huolissaan* visar mest icke-rekommendationsenligt användning och *olla pahoillaan* visar minst. *Olla pahoillaan* skiljer sig från de andra även under den första tidsperioden.

Tabell 3. Statistiska värden från SPSS-analysen för tidsperioderna

Tidsperiod	$\chi^2(2)$	p-värde	Cramérs V
2001–2006	333,426	p<0,001	0,105
2007–2011	487,615	p<0,001	0,110
2012–2017	405,147	p<0,001	0,112
Hela korpus	1264,088	p<0,001	0,112

Sammanfattning

Denna studie visar att icke-rekommendationsenlig användning av idiom som uttrycker tillstånd är på en relativt låg nivå. Det är rätt sällsynt att man använder åtminstone dessa tre idiom utan korrekt böjt possessivsuffix. Fast korpusens 17 år är en mycket kort tid med tanke på språkförändring hittade jag statistiskt signifikanta förändringar i användningen av två **av** idiomerna så man kan konstatera att det verkar ske förändringar

beträffande detta. Dessutom kan man konstatera att det finns skillnader mellan idiom med tanke på hur ofta det korrekta possessivsuffixet blir ersatt av en rekommendationsstridig form. Det som man måste beakta är det att dessa resultat gäller endast Suomi24-korpusen och språkanvändningen där. De indicerar dock att sådana former används också i talspråket men för att få reda på hur vanliga de är skulle man behöva forska vidare med språkresurser som på riktigt baserar sig på talspråk.