

Pitkin hampain — vertailussa suomen ja ranskan hammas-idiomit

Henrietta Lastunen
Pro gradu -tutkielma
Turun yliopisto
Kieli- ja käännöstieteiden laitos
Ranska, käännösviestinnän opintopolku
Huhtikuu 2019

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

TURUN YLIOPISTO

Kieli- ja käännöstieteiden laitos / Humanistinen tiedekunta

LASTUNEN, HENRIETTA: Pitkin hampain — vertailussa suomen ja ranskan hammas-
idiomit

Tutkielma, 73 s., 9 liites.

Ranska, käännösviestinnän opintopolku

Huhtikuu 2019

Tämän pro gradu -tutkielman aiheena ovat suomen- ja ranskankieliset idiomit, joihin sisältyy sana *hammas* tai sen ranskankielinen vastine *dent*. Tutkimuksessa selvitetään, minkälaisia kyseisen sanan sisältäviä idiomeja suomessa ja ranskassa esiintyy sekä onko hammas-idiomeissa tai niiden esiintymismäärissä havaittavissa kieltenvälisiä eroja. Aineisto on muodostettu poimimalla internet-teksteistä koostuvasta käyttöliittymästä lähes 200 suomen- ja yhtä monta ranskankielistä virkettä, joissa esiintyy sana *hammas* tai *dent*. Aineiston idiomit on jaoteltu verbi- ja substantiivi-idiomeihin. Kategorisoinnissa on sovellettu Korhosen (2018) luokittelua.

Sekä suomessa että ranskassa noin viidesosassa aineiston virkkeistä on *hammas/dent*-sanan sisältävä idiomi. Tutkimushypoteesi on, että verbi- ja substantiivi-idiomeja on suomessa ja ranskassa suunnilleen saman verran ja että kielten välillä ei ole selvää eroa idiomien määrässä tai esiintyvyydessä. Tutkimuksen suomenkielisessä aineistossa erilaisia verbi- ja substantiivi-idiomeja onkin suunnilleen saman verran, kun taas ranskassa ensin mainittuja on lähes nelinkertainen määrä jälkimmäisiin verrattuna. Tilanne on molemmissa kielissä sama myös silloin, kun huomioidaan erilaisten idiomien sijaan kaikki aineiston idiomit yhteenlaskettuina eli myös useaan kertaan esiintyvät idiomit. Suppeahkon yksittäistutkimuksen perusteella, jossa aineistoon valikoidut idiomit ovat aiheensa puolesta tarkkaan rajattuja, ei voida tehdä yleistyksiä. Tutkimustulosten perusteella näyttää kuitenkin siltä, että ranskassa käytetään enemmän verbillisiä kuin verbitömiä hammas-idiomeja.

Asiasanat: idiomi, internet-aineisto, kielten vertailu, ranska, suomi

Sisällys

1. JOHDANTO	1
2. IDIOMI	6
2.1. KITEYTYNYT ILMAUS	6
2.2. IDIOMI JA MUITA KITEYTYNEITÄ ILMAUKSIA	8
2.3. IDIOMIN PIIRTEITÄ	11
2.3.1. POLYLEKSIKAALISUUS	11
2.3.2. EPÄKOMPOSITIONAALISUUS JA OPAAKKISUUS	13
2.3.3. KITEYTYNEISYYS JA MUUNNELTAVUUS	15
2.3.4. LEKSIKAALISTUMINEN	20
2.3.5. KUVAANNOLLISUUS.....	21
2.3.6. IDIOMIEN KÄÄNTÄMINEN	24
2.4. IDIOMIN TUNNISTAMINEN	26
2.5. IDIOMIN FUNKTIOITA	30
3. AINEISTO JA METODI	33
3.1. INTERNET AINEISTONA	33
3.2. TUTKIMUSAINEISTO	36
3.3. AINEISTON KERUU	38
3.4. LUOKITTELU	40
3.5. HAMMAS-IDIOMIT	42
3.6. KIELTEN VERTAILU	46
4. AINEISTON ANALYYSI	50
4.1. RAKENNEANALYYSI	50

4.1.1. SUOMENKIELISEN AINEISTON IDIOMIT	50
4.1.2. RANSKANKIELISEN AINEISTON IDIOMIT	53
4.1.3. HUOMIOT RAKENTEELLISESTA ANALYYSISTA	56
4.2. SEMANTTINEN ANALYYSI	57
4.3. VARIAATIO AINEISTON IDIOMEISSA	60
4.4. EI-KUVAANNOLLISET TAPAUKSET	61
5. TULOKSET JA POHDINTA	63
5.1. TUTKIMUSTULOKSET	63
5.2. KRIITTINEN NÄKÖKULMA	65
6. YHTEENVETO	67
LÄHTEET	70

LIITE 1

1. Johdanto

Idiomaattiset, kiteytyneet ilmaukset ovat erottamaton osa kielenkäyttöä. Idiomeja käytetään jatkuvasti puheessa ja kirjoituksessa niin tiedostamatta kuin tarkoituksellisestikin. Tarkoituksellisesta käytöstä esimerkkejä ovat tilanteet, joissa puhuja tai kirjoittaja pyrkii hauskuttamaan vastaanottajaa tai pehmentämään ilmaisuaan. Idiomit ja sananlaskut elävöittävät kieltä ja tuovat siihen keveyttä. Ne mahdollistavat sanoilla leikkittelyn ja edustavat kielen luovaa käyttöä.

Idiomit ja muut idiomaattiset ilmaukset ovat siinäkin mielessä erityislaatuinen osa kielenkäyttöä, että yksittäisen vastaanottajan näkemys korostuu niiden tulkinnassa enemmän kuin monien muiden kielen osa-alueiden tulkinnassa. Eri ihmiset saattavat löytää samasta idiomista toisistaan poikkeavia vivahde-eroja, ja jotkin ilmaukset ovat ymmärrettäviä vain rajatulle puhujaryhmälle. Kiteytyneet ilmaukset voivat siis luoda myös yhteisöllisyyttä ihmisten välille.

Idiomaattisille ilmauksille ominaista on niiden kieli- ja kulttuurispesifisyys. On olemassa runsaasti esimerkiksi raamatusta (*pestä kätensä jstak*) ja antiikin kirjallisuudesta (*leipää ja sirkushuveja*) peräisin olevia idiomeja, jotka esiintyvät useissa Euroopan kielissä. Toista ääripäätä puolestaan edustavat ilmaukset, joille ei muista kielistä löydy kirjaimellisesti kääntyvää vastinetta. Monesti muissa kielissä on kuitenkin samaa tarkoittava, vaikkakin muodoltaan poikkeava ilmaus. Yksi esimerkki on kandidaatintutkielmani (Lastunen 2016: 1) aineistossa esiintyvä ranskankielinen idiomi *avoir des mains de beurre* (’jklä on voiset kädet’), jonka samaa merkitsevä suomenkielinen vastine on *olla mämmikoura*. Esimerkki tuo ilmi niin idiomien kulttuurisidonnaisuuden kuin humoristisen luonteenkin.

Idiomien tutkimus kuuluu fraseologian alaan. Suomalainen fraseologiantutkija Korhonen (2018: 13) toteaa fraseologiassa olevan kaksi eri merkitystä. Termillä voidaan ensinnäkin viitata ”tietyn kielen kiteytyneiden ja vakiintuneiden sanayhtymien kokonaisuuteen”. Toisekseen se voidaan tulkita ”itsenäiseksi lingvistiikan osa-alueeksi, jonka tutkimuskohteena ovat kiteytyneet ja vakiintuneet sanayhtymät”. Jälkimmäiseen merkitykseen viitattaessa käytetään usein myös nimitystä fraseologiantutkimus, jolloin vältytään monitulkintaisuudelta.

Fraseologia voidaan jakaa edelleen alaluokkiin, joista yksi keskeisimmistä on idiomaatiikka. Se voidaan yläkäsitteensä fraseologian tavoin tulkita kahdella tapaa. Idiomaattisella viitataan siis yhtä lailla ”tietyn kielen idiomaattisten ilmausten kokonaisuuteen” kuin yhteen fraseologiantutkimuksen osa-alueeseen. (Korhonen 2018: 14.)

Idiomille on hankalaa, ellei jopa mahdotonta, antaa yksiselitteistä ja kaiken kattavaa määritelmää. Yleistäen voidaan kuitenkin todeta, että kyseessä on lauseenosana toimiva, useammasta kuin yhdestä sanasta koostuva ilmaus, jonka kokonaismerkitys ei ole pääteltävissä sen yksittäisten sanojen merkitysten perusteella. Korhonen (2018: 14) huomauttaa, että fraseologian ulkopuolella termillä idiomi saatetaan viitata myös yksittäisen puhujan tai rajatun puhujaryhmän erityiseen puhetapaan. Tässä tutkielmassa kyseinen määritelmä jätetään kuitenkin huomioimatta. Sen sijaan termillä idiomi viitataan edellä esiteltyyn tarkempaan ilmiön määritelmään.

Idiomia luonnehdittaessa on tapana esittää sille tyypillisiä piirteitä. Tällaisia ovat muun muassa epäkompositionaalisuus, opaakkisuus eli läpinäkymättömyys ja kiteytyneisyys. Piirteiden nimitykset alalla vaihtelevat jonkin verran tutkijasta riippuen. Myös kielikohdaiset erot tuovat omat haasteensa näkemysten vertailuun, sillä käsitteiden käyttö kielten välillä on osittain epäyhtenäistä. Tämän tutkimuksen lähteinä käytän niin suomen-, ranskan- kuin englanninkielistä tutkimuskirjallisuutta.

Yksi idiomien määrittelyn ja sen käsitteen rajauksen yhteydessä esiin nouseva seikka on ilmausten muunneltavuus. Vaikka idiomeja voidaan kärjistäen kutsua kiteytyneiksi, on selvää, että osa niistä sallii enemmän tai vähemmän muuntelua. Rajanvetoa sille, milloin kyseessä on saman idiomien vaihtoehtoinen ilmaisumuoto ja milloin puhutaan jo eri idiomista, ei suinkaan ole helppo tehdä. Lisäksi puhuja voi vahingossa tai vaikkapa hauskuuttamistarkoituksessa muokata vakiintunutta idiomia täysin uudenlaiseen muotoon siten, että kuulijan on kuitenkin yhä mahdollista aiempaan tietoonsa perustuen tunnistaa taustalla oleva alkuperäisversio (*jklla ei ole kaikki kotona > kaikki jakkarat keittiössä/ kaikki kortit pakassa*). Onko tällöin yhä kyseessä sama ilmaus, vaikka se olisikin saanut uusia vivahteita tai toisenlaisen merkityssisällön? Yleistävää vastausta on vaikea antaa, ja jopa yksittäistapauksen kohdalla määrittely saattaa olla vaikeaa.

Tutkimuksen aiheena ovat sellaiset suomen- ja ranskan-kieliset idiomit, joissa esiintyy sana *hammas* tai sen ranskan-kielinen vastine *dent*. Jäljempänä esitän perustelut kyseisten sanojen valinnalle. Tutkimuksen suomen- ja ranskan-kieliset aineistot ovat peräisin

Turku NLP Groupin (Tulevaisuuden teknologioiden ja Kieli- ja käännöstieteiden laitosten tutkijoista koostuva monitieteinen tutkimusryhmä) ylläpitämästä julkisesta käyttöliittymästä. Käyttöliittymään on kopioitu suuri määrä internetin sisältämää erikielistä tekstimassaa, josta on mahdollista tehdä hakuja yksittäisten sanojen perusteella. Tutkielmassani haen osumia, joissa esiintyy sana *hammas/dent*. Suoritan haun siis sekä ranskan että suomen kielellä, minkä jälkeen valitsen tarkasteluun kummallakin kielellä saaduista tuloksista 198 ensimmäistä. Syy sille, miksi virkkeitä on molemmissa kielissä 200 sijaan 198 käy ilmi myöhemmin aineiston keruuta käsittelevästä alaluvusta 3.3. Ylivoimaisesti suurin osa hakuosumista koskee kielenkäyttöä, jossa sanat *hammas* ja *dent* ovat kuvaannollisen sijaan kirjaimellisessa käytössään. Lähempään tarkasteluun poimin joukosta ne tapaukset, joissa kyseiset sanat ovat osana idiomia. Suuren osumäärän kerääminen on tarpeellista, jotta otannasta saadaan riittävän kattava ja tulosten pohjalta voidaan tehdä yleistyksiä.

Tutkijat ovat havainneet, että idiomeissa erilaiset ruumiinosat ovat tilastollisesti muita sanaryhmiä vahvemmin edustettuina (Nenonen 2002: 114). Valitsin idiomien keräämistä rajaavaksi perusteeksi sanan *hammas*, joka muiden ruumiinosien tavoin esiintyy lukuisissa idiomeissa. Verrattuna esimerkiksi yleisimmin esiintyviin ruumiinosannimityksiin *pää*, *silmä* ja *käsi*, on sanan *hammas* käyttö osana idiomia kuitenkin harvinaisempaa (Nenonen 2002: 115). Syy siihen, miksi kuitenkin päädyin juuri valitsemaani sanaan, on tulkinnan helppous. Kolme muuta edellä mainittua ruumiinosannimitystä esiintyvät myös monissa ilmauksissa, joissa rajanvetoa idiomin ja ei-idiomaattisen ilmauksen välillä on toisinaan hankala tehdä. Esimerkiksi sana *käsi* on varsin kieliopillistunut (Ojutkangas 2001). *Hampaalla* sen sijaan viitataan tavallisesti joko kirjaimellisessa merkityksessä ihmisen/eläimen hampaaseen tai vaihtoehtoisesti se esiintyy osana idiomia. Etenkin ranskankielisen, itselleni vieraskielisen, aineiston kohdalla monitulkin-taisuuden välttäminen helpottaa ilmausten tarkastelua.

Tarkoitukseni on selvittää, minkälaisia *hammas/dent*-sanan sisältäviä idiomeja suomes-sa ja ranskassa esiintyy. Vertailen aineistoja toisiinsa siitä näkökulmasta, onko hammas-idiomeissa tai niiden esiintymismäärissä havaittavissa kieltenvälisiä eroja. Tuloksista voidaan tehdä alustavia päätelmiä siitä, onko idiomien käyttö selvästi yleisempää jom-massakummassa tarkasteltavista kielistä. Saadut tulokset ovat suuntaa antavia, ja analy-soidun aineiston koon vuoksi tuloksiin on syytä suhtautua varauksella. Korpus edustaa

vain pientä osaa kielenkäytöstä. Ainoastaan noin 200 virkkeestä koostuva aineisto on pieni otos, ja tuloksia voi osin selittää myös puhdas sattuma.

Jaottelen idiomit Korhosen (2018: 50–53) esittämän luokittelun mukaan kahteen ryhmään sen perusteella, onko kyseessä verbi-idiomi (*purra hammasta*) vai substantiivi-idiomi (*ajan hammas*). On olemassa niin suomen- kuin ranskankielisiä idiomisanakirjoja sekä useita verkosta löytyviä idiomiluetteloita, mutta niiden sisältö on vaihtelevaa. Vertailukelpoista kielten välistä aineistoa idiomien osalta ei ole olemassa, eikä näin ollen tietoa idiomien kokonaismäärästä tai niiden yleisyydestä kielenkäytössä sen paremmin suomessa kuin ranskassakaan. Tällöin ei myöskään ole perusteita olettaa, että toisessa kielessä esiintyisi enemmän idiomeja. Tutkimushypoteesi siis on, että kielten välillä ei ole selvää eroa idiomien määrässä. Lisäksi oletan, että verbi- ja substantiivi-idiomeja käytetään niin suomessa kuin ranskassa suunnilleen saman verran.

Idiomeja käsittelevää kirjallisuutta on olemassa varsin paljon, mikä kertoo aiheen suosiosta tutkijoiden parissa. Käännöstieteellisissä tutkimuksissa aiheena on yleensä erikielisten idiomien vertailu. Kuitenkin tutkimuksia, joissa vertailun kohteena olisivat suomen- ja ranskankieliset ilmaukset, on vain vähän. Yksi esimerkki tällaisesta tutkimuksesta on Saarisen (2015) susi-aiheisia idiomeja ja muita sanontoja käsittelevä pro gradu -tutkielma. Ruumiinosisannimityksiä sisältävät idiomit ovat yleisyyteensä nähden yllättävän vähän tutkittu aihe, ja esimerkiksi hammas-idiomeihin liittyviä tutkimuksia ei ole tietääkseni julkaistu lainkaan. Suomen ja ranskan hammas-idiomien vertailu on siis aiheena uusi, samoin kuin valittu näkökulma eli tietyn sanan sisältävien idiomien yleisyys kielenkäytössä.

Idiomit ovat aihe, jota voi lähestyä hyvin monesta eri näkökulmasta. Lisäksi jo pelkääntään niiden määrittely ja luonteen kuvaus vaativat monipuolista tarkastelua. Tämä tekee tutkielmastani varsin teoriapainotteisen. Seuraavassa luvussa 2 lähden liikkeelle idiomien määrittelystä. Käsittelem idiomien yläkäsitettä, kiteytynyttä ilmausta, sekä joitain idiomien rinnakkaiskäsitteitä, kuten kollokaatiota. Lisäksi käyn läpi useita idiomien piirteitä sekä käsittelem idiomien tunnistamista ja funktioita. Kolmannessa luvussa esittelen tutkimusaineiston ja -metodin. Paneudun käytetyn aineiston lisäksi internet-aineistoihin ja hammas-idiomeihin yleisellä tasolla, minkä lisäksi esittelen tutkimuksessa soveltamani Korhosen (2018) luokittelun. Lopuksi käsittelem kielten vertailun teoriaa. Neljäs luku koostuu saatujen tulosten analysoinnista. Siinä esitellään kaikki aineistoissa esiintyvät

idiomit koottuina taulukoihin. Rakenteellisen luokittelun lisäksi tarkastelen analyysiluvussa aineiston idiomeja semantiikan ja variaation näkökulmasta. Idiomien lisäksi tarkastelen lyhyesti tutkimuksessa muutoin vaille huomiota jääviä ei-kuvaannollisia tapauksia, joita valtaosa niin ranskan- kuin suomenkielisen aineiston sisällöstä kuitenkin edustaa. Varsinaista pohdintaa saatujen tulosten pohjalta on luvussa viisi, ja tutkielman loppuun kuudennessa luvussa esitellään tutkimuksen myötä tehdyt johtopäätökset.

2. Idiomi

Kiinteitä sanontoja tutkivan fraseologian alan käsitteistö on osin kirjavaa ja epäyhtenäistä (Korhonen 2018: 14–15), mikä asettaa omat haasteensa idiomin ja sen lähikäsitteiden määrittelylle. Etenkin ylä- ja alakäsitteiden nimitykset vaihtelevat toisinaan niin määrittelijästä kuin määrittelykielestäkin riippuen. Tässä luvussa esittelen joitakin toisistaan poikkeavia näkemyksiä, minkä lisäksi pyrin avaamaan idiomin käsitettä mahdollisimman monipuolisesti suhteuttamalla sen muihin ilmauksiin ja esittelemällä idiomille tyyppillisiä piirteitä. Lopuksi pohdin idiomien tunnistamista ja niiden funktioita kielessä.

2.1. Kiteytynyt ilmaus

Kiteytynyt ilmaus on kielitieteessä yleinen käsite, jonka alle idiomi voidaan sijoittaa. Toisaalta samaan tarkoitteeseen viitataan toisinaan myös termillä kiteytynyt sanaliitto, joka esiintyy muun muassa Ison suomen kieliopin verkkoversion (VISK § 454) idiomimääritelmässä.

Sanaliitolla tarkoitetaan vähintään kahden sanan muodostamaa rakennetta (VISK MÄÄRITELMÄT, s.v. sanaliitto). Esimerkkinä vakiintuneesta sanaliitosta VISK mainitsee idiomin *kynsin hampain*. Vaatimus kahden tai useamman sanan rakenteesta toteutuu kaikissa kiteytyneiksi ilmauksiksi luokitelluissa tapauksissa. Suosin siitä huolimatta sanaliiton sijasta nimitystä ilmaus, joka on osuvampi termi kuvaamaan esimerkiksi kokonaisen virkkeen mittaista sananlaskua. Toinen peruste kiteytyneen ilmauksen suosimiselle on muiden kielten vaikutus. Esimerkiksi ranskan- ja englanninkielisissä lähteissä esiintyvät termit *expression figée* ja *fixed expression*, joissa kummassakin toinen osa on suora käännös suomen sanalle *ilmaus*.

Alan käsitteistön epäyhtenäisyys ja termien rinnastettavuuden vaikeus ilmenee tutustuttaessa tarkemmin vieraskielisiin määritelmiin. Siinä missä esimerkiksi Svensson (2004) ja Moon (1998) puhuvat idiomista yhtenä kiteytyneen ilmauksen alakäsitteenä, pitää Baker (1992) nämä kaksi käsitettä erillään niiden opaakkisuuteen eli läpinäkyvyyteen perustuen. Bakerin mukaan kiteytyneitä ilmauksia ovat esimerkiksi *all the best* ja *Ladies and Gentlemen*, joiden merkitys on helposti pääteltävissä, toisin kuin idiomien merkitys useimmissa tapauksissa.

Terminologisia eroja eri maiden välillä ovat havainneet myös Granger ja Paquot (2008: 27–28). He toteavat, että jopa naapurimaiden Ranskan ja Ison-Britannian käytännöt eroavat jonkin verran toisistaan. Siinä missä ranskalaisessa tutkimuksessa kiteytyneisyyttä (*figement*) pidetään merkittävässä roolissa, on brittiläisten tutkijoiden huomio ennen kaikkea vähemmän kiteytyneissä kollokaatioissa (emt.). Ero käykin selvästi ilmi esimerkiksi Bakerin (1992) teoksesta *In Other Words*, jossa kollokaation käsittely on saanut huomattavan paljon tilaa ja vähemmälle huomiolle puolestaan ovat jääneet idiomi sekä koko kiteytyneisyyden käsite.

Korhosen (2018: 14) mukaan englannin kielessä termiä *idiom* käytetään selvästi laajemmassa merkityksessä kuin sanan vastinetta monissa muissa kielissä. Siinä missä anglistiikan piirissä yläkäsitteen asemassa on usein idiomi, suosii germaaninen tutkimus termiä fraseemi, jonka yhtenä alaluokkana ovat idiomit (emt.). Toisaalta myös englanninkielisessä kirjallisuudessa sana *idiom* luokitellaan toisinaan alakäsitteeksi (Korhonen 2018: 15). Ranskan kielessä taas suositaan perinteisesti termiä *locution* (emt. s.14), joka esiintyy esimerkiksi ranskankielisten idiomisanakirjojen nimissä.

Kuten Kortelainenkin (2011: 20) toteaa, kielentutkimuksessa ei ole päästy yhteisymmärrykseen siitä, mihin idiomit kielijärjestelmässä tulisi sijoittaa. Useiden tutkijoiden tavoin koen selkeimmäksi luokittelun, jonka mukaan idiomi on yksi kiteytyneen ilmauksen alaluokista. Svenssonin (2004) mukaan muita alakäsitteitä ovat kollokaatio, sananlasku, fraasi ja muut vakiintuneet sanonnat sekä gallisismi. Viimeksi mainitulla viitataan pelkästään ranskan kielessä esiintyvään ilmaukseen, ja sen esiintyminen selittyy Svenssonin teoksen ranskankielisyydellä. Englanniksi kirjoittavan Moonin (1998) luokittelussa gallisismi ei luonnollisesti esiinny, mutta siinä kiteytyneen ilmauksen alakäsitteinä mainitaan idiomien, kollokaation, sananlaskun ja fraasin lisäksi myös vertaus.

Edellä esitetyistä selvästi poikkeava on Korhosen (2018) melko yksityiskohtainen luokittelu. Korhonen käyttää yläkäsitteenä kiteytyneen ilmauksen sijaan germaanisessa tutkimuksessa esiintyvää fraseemia, jolle hän on määritellyt seuraavat alaluokat: sanaparit, vertausfraseemit, erisnimen sisältävät fraseemit, funktioverbirakenteet, kollokaatiot, rutiini-ilmaukset, lentävät lauseet, sananlaskut ja wellerismit. Idiomia ei siis tässä tapauksessa ole mainittu omana luokkana, mutta sen voidaan katsoa sisältyvän neljään edellä mainittuun kategoriaan. Havainto perustuu Korhosen kustakin erikoisluokasta antamiin lukuisiin esimerkkeihin, jotka voidaan luokitella idiomeiksi. Sanapareja edus-

taa esimerkiksi idiomi *pitkin ja poikin*, vertausfraseemeja *mennä kuin kuumille kiville* ja erisnimen sisältäviä fraseemeja *puhua kuin Runeberg*. Lentäviä lauseita taas ovat Korhosen mukaan muun muassa *otsa hiessä ja erottaa jyvät akanoista*. (Korhonen 2018: 54–90.)

Tutkielman edetessä nostan esiin useita Korhosen teoksessaan (2018) esittämiä näkemyksiä ja määritelmiä. Korhonen puhuu useimmiten fraseemeista, kun taas itse viittaa lähes kaikissa tapauksissa nimenomaan oman tutkimukseni kannalta oleelliseen idiomiin, joka kuitenkin kaikissa näissä tapauksissa sisältyy fraseemeihin. Tällöin mainitsen suluissa Korhosen alkuperäisviittauksessa käyttämän termin fraseemi. Suorissa lainauksissa luonnollisesti säilytän Korhosen käyttämän termin.

Seuraavassa alaluvussa pohjustetaan tutkielman pääaiheen eli idiomien määritelmää. Samalla käsitteeseen pääsevät monet muut jo mainitut kiteytyneen ilmauksen alalajit.

2.2. Idiomi ja muita kiteytyneitä ilmauksia

Idiomeja voidaan lähestyä määrittelemällä niiden suhde muihin kiteytyneisiin ilmauksiin tai tarkastelemalla niiden piirteitä. Lähdän liikkeelle idiomien lähikäsitteistä ja lähi-ilmiöistä. Ne luovat samalla kuvaa idiomien luonteesta. Vertaan idiomeja metaforiin, sananlaskuihin ja kollokaatioihin. Läheisimpään tarkasteluun pääsee viimeisenä käsiteltävä kollokaatio, jolla voidaan katsoa olevan paljon yhtäläisyyksiä, mutta toisaalta myös joitain perustavanlaatuisia eroja suhteessa idiomiin. Idiomien piirteitä tarkastellaan luvussa 2.3.

Cacciarin (2014: 269) mukaan idiomit ovat usein luonteeltaan vertauskuvallisia, sillä moni niistä on saanut alkunsa metaforasta. Erona idiomien ja metaforan välillä on kuitenkin idiomien konventionaalisuus. Siinä missä idiomilla on erityinen, kontekstista riippumaton merkityksensä, voi metaforan merkitys vaihdella tilanteesta riippuen. Esimerkkinä Cacciari mainitsee metaforan *John is an elephant*, joka on tulkittavissa eri konteksteissa eri tavoin. Metaforia voi myös kehittää spontaanisti, toisin kuin uusia idiomeja. (Cacciari 2014: 269.)

Idiomit ovat siinä mielessä epäitsenäisiä, että ne toimivat lauseenosina vaatiessa jonkin lauseyhteyden eli kontekstin toimiakseen. Muoto onkin merkittävin idiomien sananlaskusta erottava tekijä. Sananlasku on ilmaus, joka sisältää tyypillisesti kansanviisauden tai yleisesti tunnetun totuuden (Benson 1985: 66). Toisin kuin idiomi, sananlasku on jo

itsessään kokonainen, itsenäisesti toimiva lause (emt.). Sananlaskuja ovat esimerkiksi *ei kysyvä tieltä eksy* ja *aika parantaa haavat*.

Korhonen (2018: 82–83) huomauttaa idiomien ja sananlaskujen liittyvän sikäli hyvin läheisesti toisiinsa, että osaan sananlaskuista sisältyy idiomi. Tunnettuja sananlaskuja ovat muun muassa *ei pidä panna kaikkea yhden kortin varaan* sekä *ei pidä ostaa sikaa säkissä*, jotka sisältävät tunnetut idiomit *panna kaikki yhden kortin varaan* ja *ostaa sika säkissä*.

Korhonen (2018: 87) erottaa omaksi erityisluokakseen sananlaskua hyvin paljon muistuttavan wellerismin. Sananlaskun tapaan itsenäisenä kokonaisuutena toimiva wellerismi koostuu kolmesta osasta. Ilmaus alkaa sitaatilla, jota seuraavat maininnat sekä puhujasta että tilanteesta. Näin ollen wellerismiksi voidaan luokitella esimerkiksi seuraava: *Aika aikaa kutakin, sanoi pässi, kun päätä leikattiin*.

Mitä fraasiin tulee, antaa Tieteen termipankki siitä seuraavat kolme määritelmää: ”lauseen funktionaalinen rakenneosa, tiettyyn sananluokkaan kuuluvan sanan ympärille rakentunut syntaktinen konstruktio”, ”korulause, kulunut ilmaus; jollekin kielelle ominainen sanonta” sekä ”kliseemäinen, kulunut ilmaus”. Nykyajan kielenopas (luku 9 *Sanonnat*) puolestaan mainitsee esimerkkinä ilmauksen *tällä tietoa*, joka ilmaisee napakasti pidemmän version *tämänhetkisten tietojen mukaan*. Fraasin voitaneen siis todeta olevan kielenkäyttöön vakiintunut ilmaus, jonka merkitys on, tyypillisestä idiomista poiketen, helposti tai melko helposti sen osien perusteella pääteltävissä.

Kollokaatiot ovat Korhosen (2018: 15) määritelmän mukaan ”ennalta odotettavissa olevaan myötäesiintymiseen perustuvia rakenteita”. Kyse on siis yksinkertaisesti sanoista, joilla on taipumus esiintyä yhdessä. Useiden kielentutkijoiden, kuten Palmerin (1981: 79), näkemyksen mukaan idiomi on eräänlainen kollokaatio huolimatta monista eroavaisuuksistaan kollokaatioihin verrattuna.

Benson (1985: 61) jakaa kollokaatiot kahteen ryhmään: kieliopillisiin ja leksikaalisiin kollokaatioihin. Ensimmäisillä tarkoitetaan pääsanana vaatimaa tiettyä kieliopillista ainesta, josta esimerkkinä mainitaan tyypillisesti prepositiot (eng. *afraid of*; ra. *peur de*) (emt.). Suomessa prepositioita ei ole, mutta kieliopillista kollokaatiota kielessämme edustavat esimerkiksi verbien vaatimat rektiöt.

Kuten määritelmä usein yhdessä esiintyvistä sanoista antaa ymmärtää, viitataan kollokaatiolla kuitenkin tavallisesti nimenomaan leksikaaliseen kollokaatioon eli yhdessä esiintyviin kokonaisiin sanoihin. Esimerkiksi yksinään esiintyvä verbi *haukkua* tuo helposti mieleen sanan *koira*, ja sama pätee myös toisinpäin. Osa kollokaatioista on kielestä riippumatta samoja. Toisaalta on olemassa myös suuri määrä tyypillisesti yhdessä esiintyviä sanoja, joita vieraan kielen oppija ei pysty etukäteen päättelemään tai ainakaan tietämään varmaksi.

Bakerin (1992: 48–49) mukaan kollokaatiot ovat usein mielivaltaisia ja merkitykseltään riippumattomia. Vierasta kieltä puhuvalle tuttua on tilanne, jossa päänvaivaa aiheuttaa luontevimman verbin tai adjektiivin valinta tietyille substantiiville. Esimerkiksi adjektiiville *vahva* on lukuisia englanninkielisiä vastineita, mutta vahvasta teestä puhuttaessa käytetään nimenomaan sanaa *strong (tea)*.

Kielikohtaisuuteen liittyen Baker (1992: 49) nostaa esiin myös kollokaatioiden kulttuurisidonnaisuuden. Siinä missä englanniksi *buy a house* on varsin yleinen kollokaatio, Bakerin mukaan vuokra-asumista suosivassa Saksassa sanat esiintyvät selvästi harvemmin yhdessä. Vastaavanlainen kulttuuriero näkyy puhuttaessa sanoista *bread* ja *butter*, jotka etenkin useissa länsimaisissa kielissä kollokoivat todella vahvasti. Sen sijaan esimerkiksi arabian kielessä kyseisten sanojen yhteiskäyttö on huomattavasti harvinaisempaa erilaisesta ruokakulttuurista johtuen. (Baker emt.)

Kuten idiomien, myös kollokaatioiden avulla on mahdollista leikitellä kielellä. Baker (1992: 50) mainitsee poikkeavat, tunnusmerkilliset, kollokaatiot (*marked collocations*), jotka ovat yleisen kielikäsitelmän ja tavallisessa kielenkäytössä esiintyvien kollokaatioiden vastaisia. Niiden tarkoituksena on lähinnä epätavallisten mielikuvien luominen sekä lukijan huomion kiinnittäminen ja huvittaminen esimerkiksi mainonnassa.

Idiomien muunneltavuutta käsitellään myöhemmin omassa osiossaan (2.3.3.), mutta jo tässä vaiheessa voidaan muunneltavuus nostaa esiin yhtenä kollokaatiota ja idiomia erottavana tekijänä. Idiomeistakin osa on enemmän tai vähemmän muunneltavissa, kuten myöhemmin käy ilmi, mutta kollokaatioiden kohdalla rajoituksia ei käytännössä ole. Baker (1992: 63) mainitsee esimerkkinä englanninkielisen kollokaation *deliver a letter* ja sen lukuisat mahdolliset muunnelmat, kuten *delivery of a letter* ja *a letter has been delivered*. Bakerin mainitsemat esimerkit edustavat sanaluokan muunneltavuutta, mutta

yhtä lailla mahdollista tässä tapauksessa on myös leksikaalinen muuntelu. Sitä edustavat esimerkiksi ilmaukset *post a letter* ja *send a letter*.

Toisena eroavaisuutena voidaan pitää sitä, että kieleen syntyy useammin uusia kollokaatioita kuin uusia idiomeja. Toki myös idiomeja syntyy jatkuvasti lisää vanhojen hävittäessä kielestä (esim. Heinonen 2013a: 17). Tämä käy ilmi leksikaalistumista käsittelevästä osiosta 2.3.4. Kollokaatiot kuitenkin elävät vielä vahvemmin kielen mukana, sillä kieleen syntyy jatkuvasti uusia sanoja sekä uusia käyttötarkoituksia jo olemassa oleville sanoille (Baker 1992: 50). Bakerin (emt.) mielestä sellaista käsitettä kuin mahdoton kollokaatio ei lähtökohtaisesti ole olemassa. Hänen mukaansa uusia kollokaatioita tulee kieleen esimerkiksi analogian myötä ja puhujien muodostaessa tarkoituksellisesti epätavallisia sanayhdistelmiä.

2.3. Idiomien piirteitä

Tässä alaluvussa esittelen joukon piirteitä, jotka nousevat usein esiin kielitieteilijöiden luonnehtiessa idiomeja. Tällaisia ominaisuuksia ovat seuraavat: polyleksikaalisuus, epäkompositionaalisuus ja opaakkisuus, muunneltavuuden puute, leksikaalistuminen sekä kuvaannollisuus. Lisäksi otan kantaa idiomien kääntämiseen/kääntymättömyyteen, joka usein nostetaan esiin idiomien yhteydessä ja jota myös oma aiempi tutkielmani käsittelee. Alaluvun lopussa tarkastelen myös kiteytyneisyyden asteittaista luonnetta, josta monet kielentutkijat ovat yksimielisiä.

2.3.1. Polyleksikaalisuus

Määritelmä idiomista monisanaisena ilmauksena saa tukea useilta tutkijoilta. Esimerkiksi Heinonen (2013a: 8) luokittelee idiomien monisanaiseksi syntaktiseksi konstruktioksi. Hänen mukaansa *tohvelisankarin* kaltaiset yhdyssanat eivät siis ole idiomeja, vaan idiomaattisia yhdyssanoja. Myös Kortelainen (2011: 22) katsoo idiomien koostuvan vähintään kahdesta merkityksellisestä rakenneosasta viitaten rakenneosalla morfeemien sijaan kokonaisuun sanoihin. Kortelainen (emt.) perustelee näkemystään sillä, että muutoin idiomien yhteydessä ei voitaisi puhua sen osien merkitysten summasta.

Samoilla linjoilla idiomien polyleksikaalisuudesta on myös Korhonen (2018: 32), joka toteaa idiomien (fraseemien) koostuvan aina vähintään kahdesta erilleen kirjoitetusta sanasta. Korhonen kuitenkin tuo esiin myös kriittisen näkökulman aiheeseen. Hänen mukaansa se, että yksi idiomien (fraseemien) merkittävimmistä tunnusmerkeistä perus-

tuu puhtaasti oikeinkirjoitukseen, on saanut fraseologian alalla jonkin verran arvostelua osakseen (emt.).

Korhosen mainitsema kritiikki on helppo ymmärtää tarkasteltaessa kielten välisiä eroja. Useat suomen kielen yhdyssanat nimittäin vastaavat muissa kielissä monisanaisia ilmauksia. Yksi esimerkki tästä on *leipäläpi*, jonka ranskankielinen, näkemykseni mukaan idiomiksi luokiteltava vastine on *rue du pain* (Lastunen 2016). Se vastaa muodoltaan ja merkitykseltään suomenkielistä ilmausta, mutta ranskalle ominaiseen tapaan sanat on kirjoitettu erikseen. Pelkkä sanaväli siis ratkaisee sen, että toinen yllä mainituista on idiomaattinen yhdyssana ja toinen idiomi.

Tarkennuksena vaatimukseen vähintään kahdesta sanasta Korhonen (2018: 33) lisää, että useiden tutkijoiden mukaan idiomien (fraseemien) muodostavien sanojen tulee olla nimenomaan leksikaalisen merkityksen kantavia, niin kutsuttuja sisältösanoja. Tällöin ulkopuolelle jäävät funktiosanoiksi nimitetyt sanat, joilla on ainoastaan kieliopillinen tehtävä (emt.). Esimerkiksi ranskankielisessä ilmauksessa *rue du pain* sisältösanoja ovat *rue* ('katu'/'tie') sekä *pain* ('leipä'), kun taas funktiosanaksi voidaan luokitella artikkeliiin yhdistynyt prepositio *du*.

Lopuksi Korhonen (2018: 33) nostaa esiin fraseologian ja sananmuodostuksen välisen suhteen. Hän toteaa näillä kahdella kielitieteen osa-alueella olevan paljon yhtymäkohtia, mistä huolimatta ne on syytä pitää erillään. Esimerkkeinä Korhonen mainitsee idiomit *halkoa hiuksia* ja *jklla on kova sydän*, joista johdettuja ovat substantiivi *hiustenhalominen* sekä adjektiivi *kovasydäminen*. Hän toteaa jälkimmäisten olevan idiomaattisia yhdyssanoja, joita tulee tarkastella sananmuodostuksen, ei fraseologian, yksiköinä. Samaa idiomaattisten yhdyssanojen kategoriaan lukeutuvat myös muun muassa *onnenmyyrä*, *pahanilmanlintu* ja *itkupilli*.

Varsin ilmeiseltä tuntuva seikka, jonka Korhonen kuitenkin jättää mainitsematta, on se, että lisäämällä esimerkiksi edellä esitettyjen yhdyssanojen eteen *olla*-verbin on kyseessä jo idiomi. Se, milloin verbi määritellään osaksi idiomia, tuntuu melko oleelliselta kysymykseltä. Näin ollen onkin erikoista, etteivät kielentutkijat tunnu kiinnittävän huomiota asiaan. Kukin laskee tietyt osat tai tietyn verbin idiomiin kuuluvaksi oman näkemyksensä mukaan, mutta ilman perusteluita. Aiheen pohdinta voisikin saada aikaan uudenlaista keskustelua alalla. Toisinaan idiomien perusmuotoon kuuluvaksi määritelty verbi nimittäin vaihtelee lähteestä/kirjoittajasta riippuen. Esimerkiksi siinä missä Kielitoimiston

sanakirjassa mainitaan idiomi *pitää puoliaan kynsin hampain*, esiintyy Karin (1993: 29) idiomisanakirjassa versio *taistella kynsin hampain*. Kyse ei kuitenkaan ole siitä, että Kari olisi päättänyt systemaattisesti mainita verbin jokaisen idiomien kohdalla. Idiomi *hammasta purren* nimittäin esiintyy samassa listauksessa sellaisenaan, ja ainoastaan suluissa on mainittu esimerkkeinä verbit *totella* ja *ponnistella*.

Olisi tietysti loogista, mikäli idiomien osiksi ja tällöin idiomisanakirjoissa/listauksissa aiheellisesti mainittaviksi katsottaisiin vain ne verbit, jotka kuuluvat erottamattomasti ilmaukseen. Esimerkiksi *yhteen hiileen* aina puhalletaan ja henkilö nimenomaan *heittää veivinsä*, kun taas *kurkku suorana* voidaan niin huutaa, kiljua kuin kirkuakin ja *pitkin hampain* voidaan tehdä käytännössä mitä tahansa muutakin kuin vain suostua. Toisaalta tällöin listauksissa jäljelle jäisi enää lähinnä yksittäisiä sanoja ja yhdyssanoja.

Vastausta kysymykseen, mitkä osat lopulta kuuluvat idiomiin, on vaikea antaa. Kun idiomi esiintyy osana lausetta, on yleensä helppo määritellä, mitkä sanat muodostavat kyseisen idiomien. Sen sijaan sanakirjaan tai luetteloon idiomeita koottaessa on monessa tapauksessa kirjoittajan tai luettelon laatijan tehtävä päätös siitä, määritteleekö hän verbin kuuluvaksi idiomiin. Mikäli vaihtoehtoja on useita, on myös jollain perusteella tehtävä valinta niiden välillä.

Mitä idiomien sanastoon ja rakenteeseen tulee, nostan tässä kohtaa lyhyesti esiin idiomien leksikkoon perehtyneen Nenosen (2002) mainitsemia huomioita tyypillisistä suomenkielisistä idiomeista. Nenonen (2002: 65) toteaa idiomien olevan muodoltaan varsin yksinkertaisia. Verbillisissä idiomeissa esiintyy hänen mukaansa huomattavan paljon ”ruumiinosannimiä, kommunikaatioon liittyviä sanoja ja muihin yleisen tason käsitteisiin viittaavia sanoja” (emt., 114). Ruumiinosannimista useimmiten idiomeissa esiintyvät *pää, silmä, käsi, suu* ja *jalka*, mutta 20 yleisimmän joukkoon mahtuu myös *hammas* (emt., 115). Yleisimpiä verbejä puolestaan ovat perusverbit *olla, ottaa, saada, mennä* ja *pitää* (emt., 65).

2.3.2. Epäkompositionaalisuus ja opaakkisuus

Epäkompositionaalisuus ja opaakkisuus ovat erillisiä käsitteitä, mutta näen kuitenkin aiheelliseksi käsitellä niitä samanaikaisesti. Kuten tässä osiossa käy ilmi, ne liittyvät hyvin läheisesti toisiinsa ja tukevat toistensa määrittelyä. Esimerkiksi Svenssonin (2008: 81) näkemyksen mukaan epäkompositionaalisuus itse asiassa pitää sisällään vas-

takohtaisuutta ilmaisevan käsiteparin transparentti (läpinäkyvä) ja opaakki (läpinäkymättön).

Epäkompositionaalisuus viittaa siihen, että idiomien merkitys ei ole yhtä kuin sen osien merkitysten summa (Lamiroy 2008: 6). Kyseinen ominaisuus liittyy siis idiomien semantiikkaan. Kuten jo edellisessä alaluvussa käy ilmi, idiomien merkitystä ei yleensä voi päätellä sen yksittäisten sanojen merkityksistä. Kyseessä onkin nimenomaan yleistys, sillä osa idiomeista on enemmän tai vähemmän kompositionaalisia ja tällöin yleensä myös läpinäkyviä, mikä havaitaan myös myöhemmin idiomien kiteytyneisyyttä ja muunneltavuutta käsittelevässä osiossa 2.3.3. Prototyypistä idiomia luonnehdittaessa epäkompositionaalisuus on kuitenkin syytä mainita, sillä se on nimenomaan idiomeille tyypillinen ominaisuus.

Epäkompositionaalisuuteen liittyy läheisesti jo useaan otteeseen mainittu termi opaakkiuus eli läpinäkymättömyys. Se on helppo mieltää vain idiomia koskevaksi piirteeksi, mutta tarkemmin ajatellen opaakkisuuden voidaan todeta koskevan itse asiassa lähes kaikkia tavallisia, yksittäisiä sanoja (Cacciari & Glucksberg 1991: 217). Kuten Cacciari ja Glucksberg toteavat, tavalliset sanat harvemmin kohtaavat äänteen ja merkityksen tasolla. Sen sijaan sanat ovat sopimuksenvaraisia ja ne täytyy erikseen oppia. Lauseet ja niitä laajemmat kokonaisuudet puolestaan ovat kompositionaalisia, sillä niiden merkitys on ymmärrettävissä yksittäisten sanojen merkitysten kautta (emt.). Tällöin oletuksena on luonnollisesti se, että sanat ovat jo entuudestaan tuttuja. Idiomien tallentumista kielenkäyttöön eli leksikaalistumista käsitellään tarkemmin osiossa 2.3.4.

Ilmauksen kompositionaalisuus ja sen merkityksen ymmärtäminen liittyvät siis läheisesti toisiinsa, ja niiden suhteesta voidaankin tehdä seuraavanlainen yleistys: useimmissa tapauksissa epäkompositionaalisuus ja opaakkisuus kulkevat käsi kädessä, samoin kuin kompositionaalisuus ja transparenttisuus (Svensson 2004: 85–86). Kyseessä on kuitenkin nimenomaan yleistys, ei ehdoton sääntö.

Osa ilmauksista voi olla tulkintatavasta riippuen transparentteja tai opaakkeja. Svensson (2004: 86) havainnollistaa tätä ranskankielisellä ilmauksella *les carottes sont cuites*. Kirjaimellisesti tulkittuna kyseessä on täysin arkipäiväinen toteamus 'porkkanat ovat kypsiä', joka siten tulkittuna on kompositionaalinen ja transparentti. Kuvaannollisessa merkityksessään ilmaus sen sijaan kuvastaa epätoivoa tilanteessa, jossa mitään ei ole enää tehtävissä. Tällöin ilmausta voidaan pitää epäkompositionaalisenä ja merkitystä

opaakkina. Samalla tavoin esimerkiksi suomenkielisen toteamuksen *sohaisin ampiais-pesää* voi tulkita kahdella eri tavalla. Transparenteissa tapauksissa tulkintaa ohjaa konteksti, josta lisää kuvaannollisuuden yhteydessä osiossa 2.3.5.

Palmerin (1981: 80) mukaan opaakki idiomi on usein semanttisesti rinnastettavissa yhteen sanaan. Taustalla on ajatus siitä, että läpinäkymättömän idiomien merkitys on sen yksittäisten osien merkitysten sijaan lähempänä jonkin aivan muun, idiomissa esiintymättömän sanan merkitystä. Esimerkiksi idiomi *heittää lusikka nurkkaan* on korvattavissa verbillä *kuolla*. Tyyllillisesti ilmausten välillä toki on eroa, vaikka merkitys olisikin tismalleen sama.

Monissa tapauksissa jokin idiomien osista esiintyy tavanomaisessa merkityksessään, mikä luonnollisesti helpottaa idiomien merkityksen päättelyä tai ainakin ohjaa sitä oikeaan suuntaan. Esimerkiksi idiomissa *it's raining cats and dogs* ('sataa kissoja ja koiria') verbi on tavanomaisessa merkityksessään ja sitä seuraavat substantiivit poikkeavassa. (Palmer 1981: 81.)

Opaakkisuus ja epäkompositionaalisuus ovat merkittävimpiä idiomia ja kollokaatiota erottavia tekijöitä. Toisin kuin idiomien, kollokaation merkitys on lähtökohtaisesti aina yhtä kuin siihen kuuluvien sanojen merkitysten summa, jolloin epäselvyyksiä merkityksen suhteen ei jää. Mikäli kollokaation merkitys ei sen yksittäisiä sanoja tulkitsemalla aukene, on kyse todennäköisesti idiomista.

2.3.3. Kiteytyneisyys ja muunneltavuus

Kiteytymisessä on nimensä mukaisesti kyse siitä, että ilmaus on jähmettynyt tiettyyn muotoon (Korhonen 2018: 34). Kiteytyneisyydestä puhuessaan Korhonen (2018: 34–35) erottaa toisistaan käsitteet leksikaalinen ja morfosyntaktinen kiteytyneisyys. Ensimmäinen viittaa sanastollisiin rajoituksiin eli siihen, että yksittäisiä sanoja ei voi vaihtaa ilmauksen kokonaismerkityksen muuttumatta. Jälkimmäisessä tapauksessa taas on kyse rakenteellisesta, eli esimerkiksi substantiivin numerusta koskevasta rajoittuneisuudesta (*pitää peukut/*peukku pystyssä*).

Kiteytyneisyyteen liittyy läheisesti myös Nenosen (2002: 127) isolaatiksi nimittämä idiomeissa esiintyvä osa. Se kuvaa ainoastaan idiomeissa tai muissa kiteytyneissä ilmauksissa esiintyvää kielenainesta. Korhonen (2018: 34) puolestaan viittaa samaan ilmiöön termillä unikaalinen komponentti. Esimerkkinä Korhonen (emt.) mainitsee sanan

härkänen, jota käytetään pelkästään osana idiomia *tehdä karpäsestä härkänen*. Sama koskee myös esimerkiksi *jälkijunaa*, josta puhutaan vain jonkun ollessa jälkijunassa (emt.).

Useat tutkijat ovat yhtä mieltä siitä, että on olemassa eriasteista kiteytyneisyyttä ja että idiomit sekä muut ilmaukset asettuvat sen suhteen ikään kuin jatkumolle. Mejri (2005: 193) määrittelee asteittaisen kiteytyneisyyden niin, että jokaiselle ilmaukselle on oma yksityiskohtainen kuvauksensa, josta ilmenee kyseisen idiomien suhteutuminen erilaisiin kiteytyneisyyden piirteisiin. Hän ei määrittele valmiita ryhmiä, joihin idiomit voitaisiin niiden kiteytyneisyyden perusteella jaotella, mutta neuvoo tarkastelemaan kunkin ilmauksen kiteytyneisyyttä useiden yksittäisten testien avulla (Mejri 2005: 191). Tällaisia testejä ovat esimerkiksi idiomien muuttaminen passiivi- tai kieltomuotoon.

Toisin kuin Mejri (2005), monet muut tutkijat luokittelevat idiomit valmiisiin ryhmiin niiden kiteytyneisyyden asteen mukaan. Kiteytyneisyyteen perustuvat luokittelutavat vaihtelevat yksityiskohtaisista, useita kategorioita käsittävistä aina luonteeltaan yleistävempiin jaotteluihin. Näistä enemmän jälkimmäistä ääripäätä edustaa esimerkiksi Lamiroy ja Kleinin (2005) luokittelu.

Lamiroy ja Klein (2005: 144–145) jaottelevat idiomit niiden kiteytyneisyyden perusteella kolmeen kategoriaan ottaen huomioon sekä idiomien merkityksen että rakenteen. Jaottelu perustuu kirjoittajien tutkimukseen, jonka aineisto koostuu ainoastaan verbillisistä ilmauksista. Näin ollen myös luokittelussa käytetyt esimerkit ovat kaikki verbin sisältäviä idiomeja. Kutakin kategoriaa havainnollistamaan olen seuraavassa valinnut suomenkielisiä idiomeja Lamiroy ja Kleinin esittämien ranskankielisten esimerkkien tilalle.

Ensimmäiseen ryhmään kuuluvat epäkompositionaaliset ja semanttisesti täysin opaakit idiomit. Suomenkielisistä idiomeista tähän ryhmään kuuluvaksi sopii esimerkiksi ilmaus *puhua läpiä päähänsä*, jonka merkitystä ei idiomia tuntematta voi helposti päätellä. Ilmauksella ei myöskään tunnu olevan erityisesti muunteluvaraa, sillä esimerkiksi **läpiä on puhuttu päähän(sä)* ja **puhuin läven päähäni* kuulostavat kaikkea muuta kuin oikeakielisiltä. Tämän esimerkin kohdalla muuntelu rajoittunee siihen, että verbi *puhua* on mahdollista, vaikkakin harvoin, korvata sanan synonyymilla (pälpättää läpiä päähänsä), verbin perään voidaan sijoittaa adverbiaali (puhua aina läpiä päähänsä) ja verbiä voidaan taivuttaa eri aikamuodoissa.

Toisen ryhmän Lamiroyn ja Kleinin (2005: 144–145) luokittelussa muodostavat idiomit, jotka ovat niin ikään syntaktisesti kiteytyneitä, mutta ensimmäisestä ryhmästä poiketen enemmän tai vähemmän läpinäkyviä. Tällaisina ovat esimerkiksi idiomit *seisoa omilla jaloillaan* ja *olla samalla aaltopituudella*. Niiden merkitys on pääteltävissä, mutta muunteluvaraa ei juuri ole.

Kolmas ja viimeinen ryhmä puolestaan kattaa idiomit, jotka vaikuttavat täysin läpinäkyviltä, mutta joissa ei kuitenkaan ole yhtä paljon muunteluvaraa kuin tavallisissa, kiteytymättömissä ilmauksissa (Lamiroy & Klein 2005: 144–145). Esimerkiksi se, että jokin *jättää arvet* ei vaikuta erityisen kiteytyneeltä ilmaukselta, sillä verbin ja substantiivin väliin voidaan lisätä monenlaista kielenainesta (jättää pysyvät arvet, jättää pysyvästi arvet). *Arpi*-sanankorvaaminen *ruvella* on kuitenkin mahdotonta ilman, että idiomien merkitys muuttuu ja idiomaattisuus kärsii. Lisäksi kyseistä ilmausta käytetään tavallisesti kuvaannollisessa merkityksessä.

Suomalaistutkijoista kiteytyneisyyden asteittaisuuteen kantaa ottavat Nenonen ja Korhonen. Nenonen (2002: 76) toteaa suomen idiomien ”asettuvan jatkumolle syntaktisesti täysin jähmeistä lausekkeista sellaisiin lausekkeisiin, jotka sallivat miltei minkä tahansa muunnelman menettämättä silti idiomaattista merkitystään”. Hän kuitenkin lisää, että minkäänlaista selkeää hierarkkista systemaattisuutta ei ainakaan suomen lausekeidiomeissa ole havaittavissa (emt.). Koska idiomeissa (fraseemeissa) esiintyy toisinaan vaihtelua merkityksen muuttumatta, voidaan kiteytyneisyydestä Korhosen (2018: 37) mukaan puhua absoluuttisen sijaan suhteellisena käsitteenä. Taustalla on siis tässäkin tapauksessa ajatus asteittaisesta tai osittaisesta kiteytyneisyydestä.

Muodon stabiilisuutta pidetään yhtenä idiomien tärkeimmistä kriteereistä. Yleistäen voidaan todeta tutkijoiden olevan sitä mieltä, että idiomit sallivat vain vähän tai ei lainkaan muuntelua. Osa on tiukasti sitä mieltä, etteivät ne ole lainkaan muunneltavissa. Valtaosa kielitieteilijöistä on kuitenkin lievemällä kannalla asian suhteen. Esimerkiksi Benson (1985: 66) kuvaa idiomia suhteellisen kiteytyneeksi (*relatively frozen*) ilmaukseksi. Hän toteaa joidenkin muutosten, olivatpa ne sitten leksikaalisia tai kieliopillisia, olevan mahdollisia.

On tärkeää erottaa toisistaan muuntelu ja variaatio, joten määrittelen tässä yhteydessä lyhyesti myös viimeksi mainitun. Korhonen (2018: 38–40) erottaa toisistaan käsitteet leksikaalinen variaatio (*lyödä rukkaset pöytään/tiski*) ja morfosyntaktinen variaatio

(*pistää jkta silmään/silmiin*). Toisinaan sekä sanasto että rakenne voivat vaihdella, jolloin mahdollisten varianttien määrä on vielä suurempi (esimerkiksi *panna/pistää isompi vaihde päälle/silmään*) (Korhonen 2018: 41). Kaikki mainitut esimerkit ovat siis samojen ilmausten eri variantteja, eivät muunneltuja versioita. Niissä merkitys ei juuri muutu esimerkiksi käytettäessä *panna*-verbin sijaan verbiä *pistää*, kun taas muuntelun taustalla lienee yleensä halu muuttaa idiomin tyyliä tai semantiikkaa.

Idiomien muuntelu ja variaatio herättävät väistämättä kysymyksen siitä, mikä lasketaan saman idiomin variantiksi ja mikä toiseksi idiomiksi. Useimmissa tapauksissa asia lienee selvä. Esimerkiksi idiomissa *viedä kieli mennessään/mukanaan* molemmat loppuosat ovat mahdollisia, vaikkakin ensimmäinen esiintyy useammin idiomilistauksissa ja tuottaa internethaussa lähes 80-kertaisen määrän osumia jälkimmäiseen verrattuna. On kuitenkin selvää, että merkitys on kummassakin versiossa sama.

Edellisessä esimerkissä idiomin rakenne muuttui ainoastaan siten, että yksi sana sai sitä korvaavan synonyymien. Sen sijaan ilmaukset *olla jonkun leivissä* ja *syödä jonkun leivää*, joista ensimmäinen on ainakin nykyään jälkimmäistä yleisempi, eroavat rakenteellisesti jo enemmän toisistaan sekä verbin että substantiivin sijamuodon vaihtuessa. Kuitenkin merkitys säilyy tässäkin tapauksessa samana. Erot ilmausten välillä ovat siis korkeintaan tyylillisiä, ja voidaan todeta synonymian olevan mahdollista idiomienkin keskuudessa.

Mitä taas idiomien muunteluun tulee, voidaan sitä Cacciarin ja Glucksbergin (1991: 224) mukaan tehdä seuraavin keinoin:

1. Vaihtamalla/lisäämällä adjektiivi
2. Vaihtamalla/lisäämällä adverbiaali
3. Muuttamalla substantiivin ilmaisemaa määrää
4. Muuttamalla aikamuotoa
5. Antonymialla eli vastakohtaisuudella (esimerkiksi negation kautta)

Listauksesta puuttuvat Bennettin (1997: 14–15) mainitsevat objektin korvaaminen pronominilla, objektin topikalisaatio eli siirtäminen lauseen alkuun sekä verbimuodon muuttaminen aktiivista passiiviksi. Niin Bennettin kuin Cacciarin ja Glucksbergin esittämät keinot ovat kaikki toisinaan mahdollisia ja toisinaan eivät. Palmer (1981: 80) to-

teaakin muuntelumahdollisuuksien olevan täysin idiomikohtaisia. Esimerkiksi rakenteeltaan samankaltaisista idiomeista *spill the beans* ja *kick the bucket* ensimmäinen voidaan Palmerin mukaan esittää passiivissa, kun taas jälkimmäinen olisi samassa muodossa kielenvastainen (*the beans have been spilled*; **the bucket has been kicked*).

Joitain yleistyksiä voidaan kuitenkin Palmerin mukaan tehdä. Hän toteaa esimerkiksi adjektiivien taivutuksen olevan käytännössä aina kielenvastaista (*red herring*; **redder herring*). Palmer myös nostaa esiin tyypillisen idiomirakenteen verbi + substantiivi, jota edustavat esimerkiksi edellä mainitut *spill the beans* ja *kick the bucket*. Niissä aikamuodon vaihtelu on lähes poikkeuksetta hyväksyttävää, kun taas substantiivin luvun muuttaminen lähestulkoon aina kielenvastaista (*he spilled the beans*; **spill the bean – he kicked the bucket*; **kick the buckets*). Tämä lienee hyvä, vaikkakaan ei kaikissa tapauksissa pätevä, yleistys.

Useimmiten idiomia muunneltaessa korvataan vain yksi sana toisella, toteaa Omazic (2008: 76). Esimerkkinä hän mainitsee idiomien *to be born with a wooden spoon* ('olla syntynyt puulusikka suussa'), joka on helppo yhdistää alkuperäisversioon hopealusikasta. Omazic (2008: 74) mukaan muuntelu ei ole täysin vapaata, vaan tunnistettavuuden vuoksi muunnellussa versiossa tulee olla jäljellä riittävästi alkuperäisidiomin aineksia. Muuntelua rajoittavina sääntöinä voidaan pitää myös sitä, että lopputulos ei ole kielenvastainen ja että muuntelulle on oltava jokin syy (Omazic 2008: 77). Esimerkiksi idiomien *olla syntynyt puulusikka suussa* on helposti tunnistettava ja kielellisesti oikein, sillä sen avulla voidaan ilmaista lyhyesti ja humoristisesti tilannetta, jossa joku on lähtöisin köyhistä oloista.

Kieli on aina produktiivista ja elää ihmisten mukana. Samoin kuin kaikessa kielenkäytössä, myös idiomeissa esiintyy vaihtelua. Ihmiset saattavat vahingossa tai tarkoituksellisesti käyttää idiomista alkuperäisestä poikkeavaa muotoa. Ensin mainittua tapausta Korhonen (2018: 189) nimittää tahattomaksi kontaminaatioksi. Hän on poiminut ilmiöstä lukuisia sanomalehti- ja internet-teksteistä peräisin olevia esimerkkejä. Yksi niistä on Professoriliiton jäsenkirjeeseen livahtanut ilmaus *sahata omaa jalkaansa*, joka lienee yhdistelmä idiomeista *sahata omaa oksaansa* ja *ampua itseään jalkaan* (Korhonen 2018: 190).

Tarkoituksenmukaisessa sekoittumisessa puolestaan kyse voi olla pyrkimyksestä omaperäisyyteen tai halusta hauskuuttaa. Muita mahdollisia motiiveja muunneltujen idiomi-

en tietoiselle käytölle etenkin uutisotsikoissa ovat esimerkiksi lukijan mielenkiinnon herättäminen tai yksinkertaisesti tilan säästäminen (Korhonen 2018: 45). Etenkin äidin-kielisen puhujan on kuitenkin yleensä helppo tunnistaa taustalla olevat alkuperäiset idiomit.

Myös uudelleenversioidut idiomit jäävät toisinaan kieleen elämään. Rajanvetoa sille, milloin kyseessä on muunneltu idiomiksi, jo olemassa olevan idiomien pohjalta luotu uusi idiomiksi tai jokin aivan muu kielellinen aines, on ymmärrettävästi hankala, ellei jopa mahdoton joissain tapauksissa tehdä.

2.3.4. Leksikaalistuminen

Idiomien vakiintumista kielenkäyttöön kutsutaan leksikaalistumiseksi. Heinosen (2013a: 16–17) mukaan leksikaalistumisella voidaan näkökulmasta riippuen viitata ilmauksen muistiin tallentumiseen, sen vakiintumiseen osaksi kielenkäyttöä tai useasta morfeemista koostuvan ilmauksen erikoistumiseen tiettyyn merkitykseen. Kaikki näistä kuitenkin liittyvät toisiinsa, joten leksikaalistumisessa on siten kyse ilmauksen varastoitumisesta kieleen.

Heinosen kanssa samoilla linjoilla on myös Korhonen (2018: 47), joka nimittää kieliyhteisön kollektiiviseen muistiin tallentuneita idiomeja ”kieliyhteisön käytettävissä oleviksi valmisosiksi”. Tällöin ne ovat leksikon muiden yksiköiden kanssa samanarvoisessa asemassa siinä mielessä, että niiden käyttö ja tunnistaminen tapahtuvat luonnostaan ilman erillistä vaivannäköä.

Idiomit voivat olla aluksi vain tilapäisiä, mutta tallentua myöhemmin pysyvästi leksikkoon (Korhonen 2018: 47). Suositut ilmaukset luonnollisesti vakiintuvat kielenkäyttöön, kun taas kertaluontoisiksi jäävät unohtuvat (Heinonen 2013a: 17). Esimerkkejä kadonneista tai vastaavasti pitkään kielessä säilyneistä idiomeista löytyy esimerkiksi tarkasteltaessa Elias Lönnrotin sanakirjoja (Korhonen 2018: 198). Lönnrotin 1800-luvulla mainitsemia ja yhä käytössä olevia ilmauksia ovat esimerkiksi *kynsin hampain* ja *pää kolmantena jalkana*, kun taas samaisessa sanakirjassa esiintyneitä mutta nyky-suomalaiselle tuntemattomia ovat muun muassa idiomit *keittää kieltä* ja *ottaa suu käteensä* (emt.).

Täydellisen säilymisen tai katoamisen väliin jäävät tapaukset, joissa idiomiksi on kokenut aikojen kuluessa ainoastaan sanastoon tai morfosyntaksiinsa kohdistuneen muutoksen

(Korhonen 2018: 198). Ensin mainitusta tapauksesta esimerkkinä mainittakoon idiomi *henki kurkussa*, jonka alkuperäisversio on *henki hampaissa* (emt.). Myös monet muut käyttämistämme idiomeista ovat muovautuneet nykyiseen muotoonsa vasta aikojen saatossa. Yllättäen esimerkiksi ilmaus *heittää kirves kaivoon* onkin risteymä idiomeista *heittää kirves järveen* ja *pudota kuin hohtimet kaivoon* (Anhava 2002). Anhava epäilee nykyisen version saaneen alkunsa tuoreuden tavoittelusta. Hän kuitenkin toteaa tehon vähentyneen selvästi, sillä kirveen noutaminen kaivosta on työläydestään huolimatta täysin toteutettavissa, kun taas ”järveen heittäminen ilmaisee toivottomuutta paljon vahvemmin”.

Idiomit siis elävät muun kielen tavoin, minkä vuoksi Lamiroy ja Klein (2005: 137) puhuvatkin niiden diakronisesta luonteesta. He muistuttavat, että kaikki nykyisin kiteytyneinä pitämämme ilmaukset ovat joskus syntyneet ja olleet tuolloin tavallisia, kiteyttömättömiä kielen yksiköitä.

2.3.5. Kuvaannollisuus

Cacciarin (2014: 267) mukaan idiomien merkittävin erikoispiirre on niiden kuvaannollinen merkitys. Ojutkangas (2001: 41) käyttää ilmiöstä puhuessaan termiä kuvallinen, jonka hän määrittelee seuraavasti: ”sanan ei-kirjaimellinen merkitys, jonka perustana voi olla metafora, metonymia tai mikä tahansa assosiatiivinen suhde”.

Idiomeille tyypillistä on, että ne voidaan tulkita sekä kuvaannollisesti että kirjaimellisesti. Ihminen todella voi repiä pelihousunsa tai hiukset päästään, mutta harvemmin näin kuitenkaan konkreettisesti tapahtuu. Toki kaikkien idiomien kohdalla tämä ei päde, sillä kukaan ei esimerkiksi voi todellisuudessa miettiä päätään puhki eikä ihmisellä ole siipeä, jolla joku voisi elää sen paremmin kuin eviäkään, joita olla liikuttamatta. Harvemmin myöskään syötetään toiselle pajunköyttä tai työnnetään lusikkaa toisen soppaan, vaikka se mahdollista olisikin. Idiomeja tarkasteltaessa voidaan kuitenkin tehdä yleistys, jonka mukaan niin kuvaannollinen kuin kirjaimellinenkin tulkinta ovat useimmissa tapauksissa mahdollisia. Monesti idiomien kirjaimellinen merkitys on kuitenkin niin absurdilta tai muutoin erittäin epätodennäköiseltä tuntuva, että se on helppo sulkea pois (Svensson 2004: 74). Tulkinta riippuu ennen kaikkea kontekstista, ja pragmaattisella tasolla vastaanottaja pystyykin yleensä saman tien päättämään sen, kumpaa tulkintatapaa kyseisessä tapauksessa haetaan (emt., 77).

Konteksti on tärkeässä asemassa etenkin sellaisten idiomien kohdalla, joiden kirjaimellinen merkitys on käytössä yhtä yleinen tai jopa yleisempi kuin kuvaannollinen. Näin voidaan todeta esimerkiksi idiomista *pestä kätensä*. Silti myös tällaisissa tapauksissa konteksti ohjaa vastaanottajan lähes poikkeuksetta oikeaan tulkintaan. Lindquist ja Levin (2008: 143) huomauttavat, että usein ihmisellä saattaa olla mielessään samanaikaisesti sekä idiomien kirjaimellinen että kuvaannollinen tulkinta. Kuitenkin yleensä jompikumpi on yksittäisellä hetkellä vallitsevampi kuin toinen (emt.). Sanojen metaforisen merkityksen tunteminen auttaa myös esimerkiksi idiomien *break the ice* (*rikkoo jää*) tulkitsemisessa. Sana *jää* on yleinen vertauskuva sosiaaliselle jännitteelle, ja etenkin *rikkoo*-verbin kanssa sen merkitys on varsin helposti pääteltävissä (Cacciari & Glucksberg 1991: 229).

Idiomien kuvaannollisuus selittyy tavallisesti niiden alkuperällä. Hablerin ja Hümmerin (2005: 104) mukaan monet kiteytyneet ilmaukset ovat alkujaan syntyneet kieleen abstraktion, piilomerkityksen tai tosielämän kokemuksen myötä. Abstraktio viittaa siihen, että jokin asia tai ilmiö on käsitteellistetty sen piirteisiin perustuen. Esimerkiksi ranskankielinen vastine perunalle, *pomme de terre*, on kirjaimellisesti suomennettuna 'maaomena'. Nimitys tuntuu loogiselta, sillä maassa kasvava peruna todella muistuttaa muodoltaan hieman omenaa.

Tässä kohtaa mainittakoon, että vaikka kiteytyneisyys mielletään helposti koskemaan lähinnä idiomeja ja muita selvästi luovaa kielenkäyttöä edustavia kiteytyneen ilmauksen alalajeja, voivat myös neutraalit perusnimitykset yhtä lailla olla kiteytyneitä (Habler ja Hümmer 2005: 103). Yksi esimerkki tällaisesta on nimenomaan yllä mainittu ranskankielinen nimitys *pomme de terre*.

Toisinaan idiomien taustalla on konkreettinen tapahtuma tai tilanne. Vastaanottajan on usein helppo ilman taustatietoakin päätellä tällaisen idiomien merkitys. Esimerkkinä Cacciari ja Glucksberg (1991: 230) mainitsevat idiomien *bury the hatchet* (*haudata sotakirves*). Kyseessä on nykyään enää kuvaannollinen ilmaus, joka on kuitenkin saanut alkunsa siitä, että joskus muinoin sotakirveet todella haudattiin kirjaimellisesti. Idiomien ymmärtämiseen ei vaadita tietämystä yksittäisten sanojen metaforisista merkityksistä. Pelkkä mielikuva siitä, kuinka sodassa käytettävä kirves haudataan maahan, riittää kertomaan, ettei sitä enää käytetä.

Osa idiomeista tuntuu hyvinkin abstrakteilta ja mahdottomilta nyky-yhteiskunnassa palttautaa konkreettisesti alkutilanteeseen (*päästä jyvälle asiasta*), kun taas osa niistä saataisi perustua Hablerin ja Hümmerin (2005: 104) mainitsemaan empiiriseen tietoon. Se, että joku vetää maton toisen jalkojen alta tai haistaa palaneen käryä ovat varsin konkreettisia ja yleisesti koettuja asioita. Jo yleistietämykseen ja loogiseen ajatteluun perustuen on helppo päätellä, mitä tapahtumista seuraisi ja mitä idiomeilla näin ollen tarkoitetaan. Tällöin voidaan puhua idiomien taustalla olevasta motivaatiosta. Heinonen (2013a: 48) mainitsee esimerkkinä idiomien *jklla menee jauhot suuhun*. Se kuvastaa tilannetta, jossa joku menee hiljaiseksi. Näin tapahtuisi myös siinä tapauksessa, että henkilöllä kirjaimellisesti olisi suu täynnä jauhoja. Näin ollen voidaan puhua motivoituneesta idiomista. Sen sijaan esimerkiksi idiomista, jonka taustalla ei ole motivaatiota, sopii Svenssonin (2004: 148) mukaan jo aiemmin mainittu *les carottes sont cuites* kuvaannollisesti käytettynä. Porkkanoiden kypsyyssasteen ja tilanteen toivottomuuden välillä ei ole siinä mielessä loogista yhteyttä, että idiomien merkitys aukeaisi automaattisesti sen avulla. Toki kyseisenkin ilmauksen taustalta on löydettävissä jonkinlainen logiikka.

Kuvaannollisilla ilmauksilla ja kulttuurilla on selkeä yhteys, toteaa Piirainen (2008: 207). Hänen (emt., 209) mukaansa kuvaannolliset ilmaukset, kuten idiomit, ilmentävät kulttuurissa vallitsevia normeja, asenteita ja arvoja. Hän jaottelee ilmaukset viiteen ryhmään (emt., 210–212) sen perusteella, miten kulttuuri niistä heijastuu. Piirainen (emt., 212) kuitenkin muistuttaa, etteivät rajat ole tarkoin vedettyjä tai yksiselitteisiä, vaan yksittäinen ilmaus voi kuulua useampaankin ryhmään.

Ensimmäisenä Piirainen (2008: 210–211) mainitsee tekstuaaliset viittaukset ja suorat lainaukset. Tunnettuja ovat esimerkiksi monet Raamatun lainaukset (*heittää ensimmäinen kivi*), jotka esiintyvät useissa eri kielissä. Myös klassikkoteokset, sadut ja elokuvat ovat tuoneet kieleen runsaasti ilmauksia. Toiseen kategoriaan kuuluvat tieteelliseen maailmankäsitykseen, kuten tähtitieteeseen tai ihmisbiologiaan, liittyvät ilmaukset. Vaikka esimerkiksi idiomeilla *olla syntynyt onnellisten tähtien alla* tai *jollakulla kiehuu sappi* ei varsinaista totuusperää olekaan, liittyvät ne kuitenkin tieteisiin.

Kolmanteen ryhmään Piirainen (2008: 211) sijoittaa kulttuuriset symbolit. Monissa kulttuureissa esimerkiksi valkoinen väri symboloi puhtautta ja rehellisyyttä, kun taas susi liitetään vaaraan ja muihin negatiivisiin asioihin (*susi lampaiden vaatteissa*). Seuraavaksi Piirainen (emt., 212) mainitsee aineellista kulttuuria heijastavat ilmaukset, jot-

ka ovatkin yleensä varsin kulttuurikohtaisia. Siinä missä suomalaisille on luonnollista käyttää sauna-aiheisia idiomeja (*heittää lisää löylyä kiukaalle* kuvaannollisessa merkityksessä), suosivat Piiraisen mukaan luxemburgilaiset viininviljelyyn (emt., 212) ja espanjalaiset härkiin (emt., 216) liittyviä ilmauksia. Piirainen (emt., 223) huomauttaakin, että kuvaannollisilla ilmauksilla on taipumus paljastaa kullekin kulttuurille tärkeitä käsitteitä.

Luokittelun viimeinen ryhmä liittyy sosiaaliseen kanssakäymiseen (Piirainen 2008: 212). Siihen sisältyvät esimerkiksi kunnioitusta ilmaiseva *nostaa hattua* sekä ilmaukset, joiden avulla vältytään mainitsemasta suoraan tabuna pidetty aihe, kuten humala tai kuolema. Kiertoilmauksia tarkastellaan lisää idiomien funktioita käsittelevässä alaluvussa 2.5.

2.3.6. Idiomien kääntäminen

Vaikka tämän tutkielman aineisto on kaksikielinen, ei sen aiheena ole kääntäminen vaan kielten vertailu. Siitä huolimatta käsittelem idiomien kääntämistä lyhyesti omassa alaluvussa, sillä idiomien kääntäminen/kääntymättömyys jakaa mielipiteitä tutkijoiden keskuudessa.

Ajatus idiomien kääntymättömyydestä muille kielille toistuu varsin usein kielitieteilijöiden määritelmässä. Sen mainitsevat muun muassa Lamiroy ja Klein (2005: 136) sekä ranskankielisiä idiomeja ja muita kiteytyneitä ilmauksia sanakirjamuotoon keränneet Rey ja Chantreau (2003: 9). Viimeksi mainittu määrittelee teoksessaan idiomien sanayhdistelmäksi, jota ei voida kääntää sanasta sanaan muille kielille. Määritelmä toki kuvailee idiomien luonnetta, mutta samalla se on turhan kapea ja yleistävä.

Pohdittaessa sitä, mitkä ilmiöt ovat kääntämisen kannalta erityisen haasteellisia, nousevat idiomit ja muut kiteytyneet ilmaukset helposti esiin. Idiomeja voidaanakin kuvata useissa tapauksissa vaikeasti käännettäviksi. Sen sijaan tutkijoiden suosimaa, ehdottomutta kuvastavaa sanavalintaa mahdoton voitaneen kutsua jopa hieman harhaanjohtavaksi. Aikaisemman tutkielmani (Lastunen 2016) tulokset osoittavat, että yllättävän usein idiomi on mahdollista kääntää käyttämällä kohdekielestä löytyvää, merkitykseltään vastaavaa idiomiä. Usein tällaisissa tapauksissa erikieliset vastineet toki eroavat muodon ja sisällön puolesta toisistaan ainakin jossain määrin. On kuitenkin olemassa myös idiomeja, jotka ovat sanasta sanaan samoja kahdessa eri kielessä ja myös niiden

merkitys on täysin sama. Seuraavaksi esittelen lyhyesti aiemman tutkielmani (emt.) tulokset.

Kandidaatintutkielmani aineisto koostuu 60 ranskankielisestä idiomista, joissa kussakin esiintyy sana *pain* ('leipä') tai *beurre* ('voi') tai molemmat. Ranskankielessä on runsaasti idiomeja ja muita ilmauksia, joissa kyseiset sanat esiintyvät. Useimmiten tällaisten ilmausten aihepiirinä on työ/raha (*gagner son pain* = (kirj.) *ansaita leipänsä*). Tutkimuksen idiomit kerättiin sanakirjoista sekä sähköisistä aineistoista. Kaikille 60 idiomille etsittiin suomenkieliset vastineet, minkä perusteella idiomit jaettiin kolmeen eri ryhmään.

Aineiston jaottelu perustuu Bakerin (1992) käänösstrategialuokitteluun. Idiomit luokiteltiin vastaavuuden perusteella kolmeen käänösstrategiaryhmään, jotka ovat seuraavat:

- 1) semanttisesti ja muodollisesti vastaavan idiomien käyttö
- 2) semanttisesti vastaavan, mutta muodoltaan poikkeavan idiomien käyttö
- 3) parafrasien käyttö

Bakerin alkuperäiseen luokitteluun kuuluu myös kaksi muuta ryhmää, jotka ovat pois jättäminen sekä kirjaimellinen, epäidiomaattinen käänös. Ne kuitenkin jätettiin tutkimuksen ulkopuolelle. Tutkielmassa idiomeja tarkastellaan aidon käyttökontekstin sijaan niiden sanakirjamuodossa, joten luokittelu neljänteen ja viidenteen ryhmään olisi mahdollonta.

Ranskankielisille idiomeille pyrittiin siis ensisijaisesti löytämään vastineeksi niin merkitykseltään kuin muodoltaankin vastaava suomenkielinen idiomiksi. Tämä toteutui kuudesosassa tapauksista (10/60). Ensimmäiseen ryhmään luokiteltavia idiomeja edustaa jolla esiintyvä *gagner son pain* (= *ansaita leipänsä*), samoin kuin esimerkiksi *ôter le pain de la bouche de qqn* (= *viedä leipä jonkun suusta*).

Valtaosalle idiomeista ei siis ollut löydettävissä ensimmäisen kategorian mukaista käänösvastinetta, joten niille pyrittiin löytämään semanttisesti mahdollisimman läheinen vastine suomenkielisistä idiomeista. Suurin osa aineiston idiomeista (36/60) kuuluu tähän toiseen kategoriaan. Esimerkkeinä mainittakoon ranskankieliset ilmaukset *être comme pain et beurre* ('olla kuin leipä ja voi') sekä *baigner dans le beurre* ('kylpeä

voissa’), joiden merkitykseltään vastaavat suomenkieliset idiomit ovat *olla kuin paita ja peppu* sekä *kylpeä rahassa*.

Tapauksissa, joissa minkäänlaista idiomivastinetta ei löytynyt (14/60), muotoiltiin idiomien sisältö parafrasiksi. Näin toimittiin esimerkiksi idiomien *être né avant son pain* (’olla syntynyt ennen leipäänsä’) kohdalla. Ilmaus viittaa orpona ja varattomana olemiseen.

Tutkimus perustuu luonnollisesti osin omaan intuitiooni ja suomen kielen tuntemukseeni. Apuna ja varmistuskeinoina käytin ranskan- ja suomenkielisiä idiomisanakirjoja sekä internetistä löytyvää aiheeseen liittyvää tietoa. Tutkimuksen suppeudesta johtuen aineisto koostui vain 60 idiomista, mikä ei vielä riitä varsinaisten yleistysten tekemiseen. Tulokset ovat kuitenkin suuntaa antavia ja osoittavat vähintäänkin sen, ettei idiomien kääntäminen useimmissa tapauksissa suinkaan ole mahdotonta. Määrittelytavasta toki riippuu se, lasketaanko semanttisesti toisiaan vastaavat mutta muodoltaan erilaiset idiomit toistensa käännösvastineiksi. Oman näkemykseni mukaan kyllä, mutta asiasta voi olla montaa mieltä.

2.4. Idiomien tunnistaminen

Koska yhtenä idiomien ominaispiirteistä pidetään sen opaakkisuutta, on aiheellista pohdita, mistä idiomien tunnistaa. Etenkin sellaiselta ei-äidinkieliseltä henkilöltä, jonka kohdekielisten idiomien tuntemus on heikkoa, iso osa idiomeista jäänee helposti havaitsematta. Ihminen osaa tiedostamattaankin odottaa tiettyjä sanoja sekä päätellä idiomien kokonaisuudessaan pelkän alkuosan perusteella. Jonkun aloittaessa puheenvuoronsa esimerkiksi sanoilla *minua vietiin kuin* ja pitäessä sen jälkeen lyhyen tauon on todennäköistä, että kuulija ehtii mielessään lisätä lauseelle lopun *pässiä narussa*. Sen sijaan suomea vieraana kielenä puhuva henkilö, jolle idiomit on entuudestaan tuntematon, ei voi arvata sen loppuosaa. Häneltä saattaa myös jäädä idiomit kokonaan huomaamatta, koska se ei herätä minkäänlaista muistikuvaa.

Baker (1992: 65) toteaa, ettei idiomien tunnistaminen suinkaan ole itsestään selvää edes kääntäjille, jotka kuitenkin ovat kielen ammattilaisia. Lähes kaikki idiomit on nimittäin mahdollista tulkita myös kirjaimellisesti. Yleiseksi nyrkkisäännöksi tunnistamiseen Baker mainitsee idiomien outouden. Mitä vaikeammin ymmärrettävä ja mitä oudomman

kuuloinen idiomi tilanteeseen nähden on, sitä helpompi se hänen mukaansa on tunnistaa idiomiksi.

Idiomeja paljon tutkinut Cacciari on osassa tutkimuksistaan perehtynyt nimenomaan idiomien tunnistamiseen. Näiden tutkimusten perusteella hän on määritellyt joukon keinoja ja piirteitä, jotka auttavat idiomien tunnistamisessa. Seuraavaksi esitän Cacciarin kahdessa artikkelissaan esiin nostamia tunnistustapoja. Suomenkieliset esimerkit ovat minun lisämiäni, eivätkä ne liity sisältönsä tai merkitysvastaavuutensa puolesta Cacciarin artikkeleissaan esimerkkeinä käyttämiin englanninkielisiin idiomeihin. Kaikki suomenkieliset esimerkit kuitenkin kuvaavat nähdäkseni sitä, mitä Cacciari kunkin keino kohdalla tarkoittaa.

Cacciari (1993: 40–41) mainitsee useita keinoja, joiden avulla kielenpuhujan on mahdollista päätellä entuudestaan tuntemattoman idiomien merkitys. Ensimmäiseksi Cacciari kehottaa etsimään muistista samantapaista idiomia ja vertaamaan sitä uuteen idiomiin. Esimerkiksi samankaltainen rakenne voi auttaa hahmottamaan myös tuntemattoman idiomien merkityksen. Jos tietää, mitä tarkoitetaan sanottaessa *ettei jollakulla ole kaikki muumit laaksossa*, on se helppo yhdistää muodoltaan vastaavaan ilmaukseen *kaikki inkkarit kanootissa*. Toinen keino on tulkita jokin osa idiomia kirjaimellisesti ja pyrkiä sitä kautta päättämään idiomien kokonaismerkitys. Esimerkiksi idiomien *sataa kissoja ja koiria* ymmärtämisessä verbillä *sataa* on suuri merkitys.

Kolmas tapa liittyy yksittäisten osien sijaan idiomien tarkasteluun laajemmin. Cacciari (1993: 41) kehottaa pohtimaan idiomien kuvastamaa tilannetta ja päättämään, onko kyseessä jonkin asian syy vai seuraus. Esimerkiksi mielikuva siitä, mitä tapahtuu kun kirjaimellisesti *sohaisee ampiaispesää* auttaa päättämään myös idiomien kuvaannollisen merkityksen. Neljäs keino on tarkastella yksittäisiä sanoja semantiikan näkökulmasta. Toisena mainitusta keinosta poiketen tässä kohtaa Cacciari (emt.) neuvoo keskittymään sanojen kirjaimellisen tulkinnan sijaan niiden mahdollisiin vertauskuvallisiin merkityksiin. Sana *napa* mielletään yleensä jonkin keskipisteeksi, joten *oman navan tuijottelun* voi sitä kautta tulkita itsekkääksi käytökseksi.

Viides keino liittyy jälleen idiomien tarkasteluun laajemmasta näkökulmasta. Sen yhteydessä Cacciari (1993: 41) kehottaa idiomien tulkitsijaa kuvittelemaan mielessään idiomien kuvaaman tilanteen. Konkreettinen mielikuva esimerkiksi *kalasta kuivalla maalla* paljastaa myös kyseisen idiomien merkityksen.

Toisessa aihetta koskevassa artikkelissaan Cacciari (2014: 270) toteaa kielellisten erikoisuuksien auttavan usein idiomien tunnistamisessa. Siihen liittyen hän myös luettelee lisää keinoja idiomien tunnistamiseen. Ensimmäiseksi idiomien erikoispiirteeksi Cacciari (emt.) mainitsee kieliopillisen ja käsitteellisen luvun ristiriitaisuuden. Esimerkiksi ilmausten *olla päissään/naamat* substantiivi on monikkomuotoinen myös silloin, kun kyse on vain yhdestä henkilöstä. Toki kyseisissä esimerkeissä idiomaattisesta kielenkäytöstä kielii jo kirjaimellisen tulkinnan mahdottomuus.

Muiksi tunnistamista auttaviksi tekijöiksi Cacciari (2014: 270) mainitsee määriteosan eli tarkenteen puuttumisen, syntaktisesti poikkeuksellisen rakenteen sekä sellaiset sanat tai muut kielelliset ainekset, jotka esiintyvät ainoastaan idiomeissa. Myös se, että ilmausta ei pysty muuntelemaan samalla tavoin kuin muita ilmauksia, paljastaa usein kyseessä olevan idiomi.

Yleensä lähdetään liikkeelle oletuksesta, että idiomit ovat opaakkeja ja niiden merkitys vaikeasti pääteltävissä. Heinonen (2013a: 10) kuitenkin katsoo, että merkityksen ennakkoimattomuutta monesti liioitellaan. Hänen mukaansa ”tutun kulttuurisen taustatiedon avulla on mahdollista päätellä yhtä ja toista”.

Svensson (2004: 33) toteaa kielitieteilijöiden pitävän usein omaa intuitiotaan merkittävänä keinona ilmausten kiteytyneisyyden tunnistamisessa, olipa se sitten tietoista tai ei. Fraseologia ja etenkin idiomitutkimus ovatkin aloja, joilla subjektiivisuus lienee tavallista suuremmassa asemassa. Idiomeja on tutkittu runsaasti jo vuosikymmenten ajan eri puolilla maailmaa, mutta siitä huolimatta idiomeille ei ole onnistuttu antamaan yksimielistä ja kaiken kattavaa määritelmää. Yleistysten tekeminen ja idiomeja yhdistävien piirteiden löytäminen on helppoa, kun taas tarkemmat käsitykset siitä, miten idiomit tulisi määritellä ja mitkä ilmaukset voidaan luokitella kyseiseen kategoriaan kuuluviksi vaihtelevat enemmän tai vähemmän määrittelijästä ja myös kielestä riippuen.

Edellä on tarkasteltu idiomien tunnistamista teoreettisesta näkökulmasta. Näin ollen on luonteva siirtyä käsittelemään sitä, miten oman tutkimukseni aineistossa esiintyvät idiomit on käytännössä tunnistettu ja todettu idiomeiksi. Idiomien piirteistä eniten aineiston analyysissä korostuvat kuvaannollisuus ja leksikaalistuminen. Toki yhtä lailla myös muut edellä mainitut ominaisuudet ovat jokainen tärkeitä ja osallisia siihen, että yksittäisiä ilmauksia pystytään luokittelemaan idiomeiksi.

Käydessäni aineistoa läpi pysähdyin jokaisen *hammas/dent*-sanan ympärille rakentuvan ilmauksen kohdalla pohtimaan ennen kaikkea sitä, esiintyykö *hammas/dent* siinä kirjaimellisessa vai kuvaannollisessa merkityksessään. Erityisen tärkeää ja monesti ratkaisevaa tämä oli ranskankielisen aineiston kohdalla. Osa ilmauksista tuotti vaikeuksia, mutta aineiston tapauksissa lauseyhteys tai viimeistään alkuperäiskontekstin tarkastelu kokonaisuudessaan paljasti aina sen, onko kyseessä idiomi vai kirjaimellisesti tulkittava ilmaus. Esimerkiksi *montrer les dents* (‘näyttää hampaansa’) on kontekstista riippuen mahdollista tulkita joko kuvaannollisesti tai kirjaimellisesti, ja tässä tapauksessa se olisi-kin ollut toisinaan hankala todeta ilman mahdollisuutta tarkastella alkuperäistekstiä, josta kyseisen idiomien sisältämä virke kussakin tapauksessa on peräisin, kokonaisuudessaan. Sen sijaan esimerkiksi ilmaus *avoir les dents longues* (‘olla kunnianhimoinen’; kirj. *olla pitkät hampaat*) on kirjaimellisesti tulkittuna niin erikoinen, että se on helppo tunnistaa idiomiksi.

Suomenkielisen aineiston läpikäymisessä myös ilmauksen leksikaalistuminen eli se, että se on varastoitunut osaksi kielenkäyttöä, teki idiomien tunnistamisen helpoksi. Kuten leksikaalistumista käsittelevästä osiosta 2.3.4. käy ilmi, äidinkielen kielenpuhujan kohdalla idiomien tunnistaminen tapahtuu luonnostaan, ilman erillistä vaivannäköä. Esimerkiksi se, onko *x hammasta kolottaa* tai *kynsin hampain* idiomi vai ei, on suomenpuhujalle itsestäänselvyys.

Kuvaannollisuuden ja leksikaalistuneisuuden ohella yhtenä kriteerinä aineiston idiomien tunnistamisessa oli polyleksikaalisuus. Kaikki niin suomen- kuin ranskankieliset ilmaukset nimittäin koostuvat vähintään kahdesta sisältösanasta. Jotkin ranskankieliset substantiivi-idiomit ovat niin kirjaimellisesti kuin merkityssisällöltään suomennettuina yhdyssanoja (*en dent de scie* = ‘epäsäännöllisesti’; kirj. *sahanhampain*). Siinä ei kuitenkaan ole ristiriitaa, sillä kumpakaan suluissa mainittua suomennosta ei muutoinkaan luokitella idiomiksi. Mikäli kuitenkin olisi epäselvää, ovatko adverbit *epäsäännöllisesti* ja *sahanhampain* idiomeja, olisi polyleksikaalisuus, tai tässä tapauksessa sen puuttuminen, ratkaiseva kriteeri.

2.5. Idiomien funktioita

Yksi idiomien tehtävistä on hauskuuttaa ja viihdyttää. Kortelainen (2011: 50) toteaaakin idiomien tekevän keskustelusta kevyempää, hauskempaa ja rennompaa. Idiomit, aivan kuten sanonnatkin, elävöittävät puhetta ja tekstiä. Ne myös muodostavat kielen osa-alueen, joka mahdollistaa monia muita kielenaineiksia paremmin luovuuden ja sanoilla leikkittelyn. Idiomeja voi myös muunnella tilanteen mukaan, jolloin syntyy omaperäisiäkin ilmauksia. Yksi kielellä leikkittelyn keino on käyttää idiomeja siten, että se voidaan tulkita sekä kuvaannollisesti että kirjaimellisesti (Muikku-Werner 2009: 86). Esimerkkinä tapauksesta, jossa ”kielen eri tasot sekoittuvat hupaisasti” Muikku-Werner (emt.) mainitsee ilmauksen *kissa nukkuu koiranunta*.

Muikku-Wernerin (2009: 84) mukaan kirjoittaja pystyy idiomien avulla säätelemään tekstin tyyliä ja näin ollen myös virallisuutta. Esimerkiksi ilmausta *heittää veivinsä* ei ole sopivaa käyttää kaikissa tilanteissa kuolemasta puhuttaessa. Äidinkieliselle puhujalle idiomien tyyllitason tunnistaminen on yleensä helppoa, kun taas vieraskieliselle se on huomattavasti haastavampaa. Tyylin määrittelyssä käytetään tavallisesti neljää eri tasoa, jotka ovat seuraavat: ylätyylinen (*olla vuoteen omana*), neutraali (*olla kieli keskellä suuta*), arkityylinen (*lähteä lätkimään*) sekä alatyylinen (*olla kuin paita ja perse*) (Korhonen 2018: 156–160).

Toinen Muikku-Wernerin mainitsemista idiomien funktioista on kielenkäyttäjän suhtautumisen ilmaiseminen (2009: 85). Yllättävän usein puhujan tai kirjoittajan valitsema idiom kertookin jotain tämän omasta asennoitumisesta. Muikku-Werner (emt.) käyttää esimerkkinä lauseita *Ville on hyvä suustaan* ja *Pekka ei sylje lasiin*. Ensimmäinen on yleensä positiiviseksi mielletty tokaisu, kun taas jälkimmäisessä selvästi kritisoidaan henkilön juomatottumuksia, vaikka se tehdäänkin epäsuorasti.

Arkaluontoiset asiat ja hallitsemattomat tunteet verhotaan usein eufemismeihin eli kiertoilmauksiin (Muikku-Werner 2009: 86–87). Esimerkiksi alkoholin vaikutuksen alaisena olemista kuvaavia idiomeja on tunnetusti paljon eri kielissä. Idiomien kiertämisfunktion liittyen Korhonen (2018: 162) mainitsee myös aineenvaihduntaan (*tehdä tarpeensa*) sekä sukupuolisuuteen (*olla siunatussa tilassa*) viittaavat ilmaukset. Habler ja Hümmer (2005: 107) toteavat negatiivista toimintaa ja tunteita kuvastavia ilmauksia olevan yleisesti ottaen enemmän kuin ilmauksia, joilla kuvataan positiivisia tunteita

tai yhteiskunnan normien mukaista toimintaa. Voidaan siis todeta, että ihmisellä on luonnostaan tarve kehittää ikäviä ja noloja asioita kiertäviä tai lieventäviä ilmauksia.

Edellä esitetylle havainnolle on helppo saada vahvistusta tarkastelemalla suomalaisia idiomeja. Esimerkiksi valehtelua tai muutoin epärehellistä toimintaa kuvaavia ilmauksia (*viilata linssiin; puhua itsensä pussiin*) löytyy runsaasti, kun taas rehellisyyteen viittaavia (*olla sanansa mittainen*) on huomattavasti vaikeampi keksiä. Sama ilmiö esiintyy myös monilla muilla aihealueilla. Siinä missä tyhmyyttä ja ärtymystä kuvaavia idiomeja on kielessämme paljon, löytyy älykkyyteen ja hyväntuulisuuteen viittaavia ilmauksia selvästi vähemmän. Kortelainen (2011: 6) huomauttaa tyhmyyden olevan ilmeisen yleinen puheenaihe suomen kielessä, sillä tapoja sen ilmaisemiseksi on erityisen paljon. Kortelainen on perehtynyt aiheeseen, sillä tutkielmansa nimen (*Ei ole kaikki muumit laaksossa*) mukaisesti hän tarkastelee nimenomaan tyhmyyttä kuvaavia idiomeja, ja tarkemmin erilaisia idiomien *ei ole kaikki kotona* variaatioita.

Yllä esitetty päätelmä negatiivissävytteisten idiomien positiivisia suuremmasta määrästä saa tukea myös tässä tutkimuksessa tehtävistä havainnoista. Lähempi *hammas*-aiheisten idiomien tarkastelu osoittaa, että sanan sisältävien ilmausten joukossa yleisimpiä aihepiirejä ovat muun muassa pelko, vastentahtoisuus, katkeruus ja vihamielisyys. Aihetta käsitellään tarkemmin alaluvussa 3.5.

Habler ja Hümmel (2005: 107) mainitsevat yhdeksi idiomien funktioksi kielenkäytön taloudellisuuden. Heidän mukaansa on olemassa käsitteitä, joita ei voida ilmaista muulla tavoin. Kyse on siis samasta asiasta, joka helposti aiheuttaa ongelmia myös idiomien kääntämisessä. Moni idiomi nimittäin sisältää erilaisia sävyjä, sivumerkityksiä ja jopa viitteitä puhujan asennoitumisesta. Jotta tämä kaikki saataisiin muulla tavoin ilmaistua, vaatisi se huomattavasti monisanaisempaa kerrontaa.

Edellä mainittujen lukuisten funktioiden lisäksi on syytä mainita myös idiomien tehtävä yhteisöllisyyden luojina. Jo yhteinen kieli itsessään saa ihmiset kokemaan yhteenkuuluvuuden tunnetta, samoin kuin pienemmässä mittakaavassa tietyn alueen sisällä puhuttavat murteet tai tietyn ikäryhmän suosima puhetapa/sanasto. Kaikki tietynkieliset ilmaukset eivät kuitenkaan automaattisesti kuulu kaikkien kyseistä kieltä käyttävien puhevarastoon. Toisin sanoen saman kielen puhujat eivät käytä tai välttämättä edes ymmärrä samoja idiomeja, sillä myös niiden käytössä esiintyy alueellisia eroja (Lamiroy & Klein 2005).

Usein yhteisöllisyyden kokemukseen liitetään myös huumori. Idiomit mielletäänkin monesti hauskoiksi tai muutoin viihdyttäviksi, ja näin ollen niiden viljely tuo keskusteluun huumoria. Kortelainen (2011: 59) tiivistää idiomien hauskuuttamisfunktion seuraavasti: ”Idiomeja ei (kuitenkaan) tyypillisesti käytetä vitseinä sellaisenaan ja kontekstista irrallisina, vaan ne ovat osana vapaan kerronnan sanailua. Nokkelat sanavalinnat oikeanlaisessa tilanteessa tekevät idiomista koomisen.” Leikilliseen asennoitumiseen liittyen idiomilla voidaan osoittaa myös ironiaa, kuten käytettäessä ilmausta *elää uutta nuoruutta* (Korhonen 2018: 162).

Tässä luvussa käsiteltiin idiomeja useasta näkökulmasta. Kuten tarkastelusta käy ilmi, aiheeseen liittyvää kirjallisuutta ja tutkimuksia on runsaasti erikielistä. Seuraavaksi esittelen oman tutkimukseni aineiston ja metodin.

3. Aineisto ja metodi

Tässä luvussa käsittelen ensin internet-aineistoa yleisesti tarkastelemalla sen hyviä ja huonoja puolia. Sen jälkeen esittelen tutkimuksen korpuksen ja sen taustalla olevan projektin. Lisäksi kerron aineiston keruumenetelmästä ja luokittelusta, jota sovellan aineistoon. Luvun loppupuolella käsittelen omissa alaluvuissaan myös tutkielman aiheena olevia *hammas*-idiomeja sekä kielten vertailun teoriaa. Tässä luvussa esiteltävät metodit luovat pohjaa seuraavassa luvussa tehtävälle analyysille.

3.1. Internet aineistona

Kilgarriff ja Grefenstette (2003: 333) kuvaavat internetiä valtavaksi ja vapaasti käytettävissä olevaksi aineistoksi, jossa on monipuolisesti edustettuna eri kieliä ja murteita. He kutsuvatkin sitä humoristisesti kielitieteilijöiden ”unelmien leikkikentäksi” (2003: 333). Mutta voidaanko internetiä sen valtavan koon vuoksi pitää korpuksena? Kilgarriffin ja Grefenstetten mielestä voidaan, sillä yhden määritelmän mukaan korpus on kokonaisuus tekstejä silloin, kun tekstit ovat kielen- tai kirjallisuudentutkimuksen kohteena (2003: 334).

Internet-aineiston etuna on luonnollisesti internetin laajuus (Heinonen 2013a: 27). Mitä enemmän dataa, sen parempi tutkimuksen kannalta, toteavat Kilgarriff ja Grefenstette (2003: 336). He myös huomauttavat, että datan määrän lisääminen on yleensä tehokkaampaa kuin algoritmien hienosäätö.

Heinonen (2013a: 26) kuvaa internet-aineistoa sekakoosteiseksi. Aineiston ehdottomana vahvuutena hän pitää arkikielisten, eli ei kielen standardeja tarkkaan noudattavien tekstien suurta osuutta. Tämän tyyppisten tekstien määrä selittyy esimerkiksi keskustelupalstojen olemassaololla. Ei-standardinmukaisuus on hyödyllinen piirre tämän tutkimuksen kannalta, sillä sitä todennäköisesti viljellään enemmän epävirallisessa kuin virallisessa kielessä. Näin ollen se, että internet-teksteistä koostettu aineisto sisältää niin akateemista tekstiä kuin myös ihmisten keskustelupalstoille ja uutisten kommentointiosioon tuottamaa arkipäiväistä ja toisinaan puhekielistäkin sisältöä, lisää aineiston esiintyvien idiomien määrää. Aineiston sekakoosteisuus myös parantaa korpuksen edustavuutta.

Suuria tekstimassoja sisältävä internet-korpus on tutkimuksessa erityisen hyödyllinen harvinaisia rakenteita tarkasteltaessa. On esimerkiksi paljon kielellisiä ilmiöitä ja rakenteita, jotka esiintyvät teksteissä vain harvoin tai eivät lainkaan (Luotolahti, Kanerva, Laippala, Pyysalo & Ginter 2015: 212). Tällöin ainoa mahdollinen keino niiden tutkimiseen saattaa olla internet-aineiston hyödyntäminen. Kuten jo edellisestä kappaleesta käy ilmi, voitaneen myös idiomit yleisesti määritellä sellaiseksi kielelliseksi ilmiöksi, jonka esiintyminen ja yleisyys riippuvat genrestä ja ennen kaikkea rekisteristä. Siirryttäessä yleiseltä tasolta harvinaisempiin tapauksiin, kuten *hammas*-sanana sisältäviin idiomeihin, on mahdollisimman suuresta ja sekakoosteisesta tekstimassasta etua.

Genren eli tekstilajin voi määritellä usein eri tavoin. Asheghi, Sharoff ja Market (2016: 605) toteavat useiden tutkijoiden korostavan määrittelyissään tekstin muotoa ja käyttötarkoitusta. Erityisen pätevänä he pitävät alun perin Orlikowskin ja Yatesin (1994: 543) laatimaa määritelmää, jonka mukaan genre ei ole ainoastaan tiettyjä ominaispiirteitä sisältävä, viestinnälliseen tarkoitukseen soveltuva kategoria, vaan sen on oltava myös yhteisön yleisesti tunnistama (Asheghi, Sharoff & Market 2016: 606).

Internet on tuonut mukanaan myös painetuista teksteistä poikkeavia tekstilajeja, minkä vuoksi Asheghi, Sharoff ja Market (2016: 606) nostavatkin esiin jaottelun web- ja perinteisiin genreihin. Web-genreä edustavat esimerkiksi ainoastaan internetissä julkaistavat blogit. Näiden kahden ääripään väliin sijoittuu erilaisia välimuotoja. Esimerkiksi verkko-utinen voi sisällöllisesti vastata täysin sanomalehdessä julkaistavaa uutista, mutta mahdollisine hyperlinkkeineen ja kommentointimahdollisuuksineen se kuitenkin poikkeaa painetusta tekstistä.

Rekisterillä viitataan tilannekohtaiseen kielen vaihteluun. Rekisteri voi olla tarkoin rajattu, kuten yksittäisen kirjailijan teoksissaan käyttämä kieli tai jollain tieteenalalla käytettävä tarkoin määritelty tyyli. Toisinaan taas rekisteri voi olla niinkin yleistettävä kuin esimerkiksi pelkkä keskustelu. Rekisteri on syytä erottaa murteesta. Siinä missä viimeksi mainitulla viitataan määrättyllä alueella asuvan tai määrättyä sosiaaliryhmää edustavan puhujajoukon käyttämään kieleen, on rekisterissä kyse nimenomaan tilannekohtaisesta vaihtelusta. Siihen vaikuttavat muun muassa keskustelun tarkoitus, aihe ja ympäristö sekä keskustelukumppanit. (Biber, Conrad & Reppen 1998: 135.)

Korpustutkijan on kiinnitettävä huomiota ja otettava kantaa korpuksensa edustavuuteen. Biberin, Conradin ja Reppenin (1998: 246) mukaan korpus ei ole ainoastaan kokoelma

tekstejä, vaan se myös pyrkii edustamaan kieltä yleensä tai jotakin kielen osa-aluetta. Korpuksen edustavuus puolestaan määrittää heidän mukaansa sen, millaisia tutkimuskysymyksiä voidaan esittää ja kuinka yleistettävänä saatuja tuloksia voidaan pitää.

Kun tutkitaan kielellistä variaatiota, korpuksen tulisi olla mahdollisimman edustava eikä vain edustaa rajallista otosta mahdollisesta vaihtelusta. Otosmenetelmä toimii esimerkiksi laadittaessa poliittisten vaalien gallupeja, joissa suhteellisen pieni otanta voidaan yleistää kuvastamaan koko väestöä. Sen sijaan kielellistä vaihtelua tarkasteltaessa laaja kirjo on välttämätön, jotta vaihtelu olisi aineistossa mahdollisimman kattavasti edustettuna. (Biber, Conrad & Reppen 1998: 247.)

Edustavan korpuksen tärkeimmät ominaisuudet ovat monipuolisuus sekä riittävä koko. Monipuolinen aineisto sisältää laajasti eri rekistereitä. Korpuksen koossa oleellista ei ole ainoastaan suuri sanamäärä vaan myös se, että kutakin eri tekstikategoriaa edustaa riittävän monta tekstiä. Riittävä määrä on kuitenkin suhteellinen käsite, sillä siihen vaikuttaa luonnollisesti tutkittava ilmiö. Mitä harvinaisemmasta ilmiöstä on kyse, sitä suurempi aineisto tarvitaan. Kokoavasti voidaan todeta, että vaikka jokaisella korpuksella on omat rajoituksensa, on asianmukaisesti laadittu korpus siitä huolimatta hyödyllinen väline kielellistä vaihtelua tutkittaessa. (Biber, Conrad & Reppen 1998: 248–250.)

Internetin edustavuus on asia, joka jakaa mielipiteitä. Vastauksena kritiikkiin, jonka mukaan internet korpuksena ei edusta kielellistä variaatiota riittävän monipuolisesti, Kilgarriff ja Grefenstette (2003: 340) esittävät vastakysymyksen: mikä sitten edustaa? Jos määritelmänä pidetään kielen neljää osa-aluetta (puhuttu, kirjoitettu, luettu ja kuultu) ja sitä, että ne olisivat kaikki yhtä lailla edustettuina, on kielellistä variaatiota mahdollisimman kattavasti edustavan korpuksen laatiminen lähes mahdotonta. Kilgarriff ja Grefenstette (2003: 340–341) nostavat esiin yksittäisiä edustavuuteen liittyviä ongelmia, kuten sen, että iso osa käyttämästämme kielestä on lainaamista eli aiemmin sanotun tai kirjoitetun kopiointia. Lainaukseksi luokitellaan myös kääntäminen ja tulkkaukset, sillä vaikka kieli tällöin vaihtuu, säilyy viestin sisältö lähtökohtaisesti samana. Kilgarriff ja Grefenstette (emt.) muistuttavat myös esimerkiksi unissa puhutun tai tiedostamatta kuullun kielen määrittelyn ongelmallisuudesta. Samoilla linjoilla internet-aineiston edustavuudesta on myös Heinonen (2013a: 26). Hänen mukaansa on hankala sanoa, kuinka tasapuolista ja näin ollen edustavaa aineisto on. Samalla Heinonen huomauttaa, että on vaikea edes määritellä sitä, mitä todellinen edustavuus loppujen lopuksi on.

Painettuihin teksteihin verrattuna internet-teksteissä esiintyy selvästi enemmän kielivirheitä, sillä niitä tuottavien henkilöiden kirjo on huomattavasti laajempi (Kilgarriff & Grefenstette 2003: 342). Ihmisten kirjoittaessaan tekemät virheet eivät luonnollisesti vähennä aineiston edustavuutta, mutta asia on silti hyvä pitää mielessä. Yhtä lailla virheiden määrässä on eroa myös eri tekstilajien välillä. Idiomeja tarkasteltaessa tuleekin muistaa, että esimerkiksi väärässä merkityksessä käytetty idiomi esiintyy todennäköisemmin verkkokeskustelussa tai blogissa kuin ammattikirjoittajan tuottamassa ja oikolukijan tarkastamassa uutistekstissä.

Yhtenä internet-aineiston heikkoutena Kilgarriff ja Grefenstette (2003: 341) nostavat esiin kielen mallintamiseen liittyvät haasteet. Heidän mukaansa mallintamista vaikeuttaa se, ettei verkkotekstejä pystytä kunnolla luokittelemaan tekstilajeittain. Kilgarriff ja Grefenstette (2003: 345) toteavatkin lopuksi, että tulevaisuudessa koko internetin sisältö voitaisiin jaotella eri tekstilajeihin. Lisäksi he ehdottavat, että verkkotekstit voitaisiin jäsentää kieliopillisesti. Nyt 15 vuotta Kilgarriffin ja Grefenstetten artikkelin julkaisun jälkeen voidaan kuitenkin todeta, että heidän ehdotuksensa ovat jo osin toteutuneet ja jatkavat yhä toteutumistaan seuraavassa alaluvussa esiteltävän projektin ansiosta.

3.2. Tutkimusaineisto

Tutkimukseni suomen- ja ranskankieliset aineistot ovat peräisin Turku NLP Groupin ylläpitämästä julkisesta käyttöliittymästä, jossa valittavana on useiden eri kielten korpuksia. Turku NLP Group on Tulevaisuuden teknologioiden ja Kieli- ja käännöstieteiden laitosten tutkijoista koostuva monitieteinen tutkimusryhmä, jonka tutkimusalana on kieliteknologia (<https://turkunlp.github.io/> 8.1.2018). Lyhenne NLP viittaa luonnollisen kielen käsittelyyn (Natural Language Processing). Ryhmä vastaa Finnish Internet Parsebank -projektista, jonka johtajina toimivat apulaisprofessorit Veronika Laippala ja Filip Ginter. Projekti on osa kansainvälistä Universal Dependencies (UD) -järjestelmää, johon kuuluu tällä hetkellä yli 100 puupankiksi kutsuttua tietokantaa yli 60:lla eri kielellä (Universal Dependencies 8.1.2018). UD-järjestelmän tarkoituksena on jäsentää kielet samaa mallia noudattaen, jolloin eri kielten rakenteet ovat keskenään vertailukelpoisia (emt.). Projektin taustalla on siis kielitypologinen näkökulma (emt.). Käytännössä kaikissa käyttöjärjestelmään syötetyissä erikielisissä virkkeissä on määritelty jokaisen yksittäisen sanan sanaluokka sekä sanojen välisiä suhteita (esim. apu- ja pääverbin erottaminen).

UD-järjestelmän avulla koottu datakokoelma, johon erikieliset tietokannat kuuluvat, on nimeltään Universal Parsebanks. Sen koko kymmeniä miljardeja sanoja käsittävä sisältö koostuu yksinomaan internet-teksteistä, jotka on koneellisesti kopioitu. Aineisto on koottu alun perin kansainvälistä tieteellistä kilpailua varten. Siinä eri maiden osallistujat kilpailivat syntaksijäsentämisessä. Kaikki erikieliset tietokannat on syntaksijäsennetty automaattisesti saman annotaatiojärjestelmän avulla, minkä ansiosta haku voidaan kielestä riippumatta suorittaa samoilla kriteereillä. (Suullinen tiedonanto, Veronika Laippala 18.1.2018.)

Suomenkielisen internetin sisällöstä koostuva korpus, Finnish Internet Parsebank, sisältää toistaiseksi lähes neljä miljardia sanaa. Turku NLP -tutkimusryhmällä on käynnissä projekti, jonka tavoitteena on luoda morfologian ja syntaksin osalta automaattisesti analysoitu kieliaineisto. Sen avulla voidaan tarkastella esimerkiksi lauserakenteita. Lisäksi tavoitteena on luokitella koko kieliaineisto järjestelmällisesti tekstilajeittain. Käyttöliittymä on julkinen ja näin ollen vapaasti hyödynnettävissä osoitteessa www.utu.fi/dep_search/. (Turku NLP Group 8.1.2018.)

Suomenkielisen korpuksen aineisto on kerätty kahdesta eri lähteestä. Osa on peräisin yhdysvaltalaisen Common Crawl -järjestön keräämästä aineistosta, osa taas on suomalaishankkeen oman robotin internetistä kopioimaa sisältöä (Luotolahti ym. 2015: 212). Aineistolle tehdään automaattinen syntaksianalyysi käyttäen UD-järjestelmää, jonka avulla jäsenitys voidaan tehdä kielestä riippumatta (Luotolahti ym. 2015: 218). Analyysi tapahtuu Finnish Dep Parserilla ([turkunlp.github.io/Finnish-dep-parser/](https://github.com/turkunlp/Finnish-dep-parser/)) (Luotolahti ym. 2015: 214).

Aineiston saattaminen siihen muotoon, että virketason jäsenitys on mahdollista, vaatii jonkin verran suodatamista. Iso osa internetin sisällöstä on esimerkiksi luetteloita ja valikoita, joille perinteistä syntaksianalyysia ei voida suorittaa. Näin ollen tällainen sisältö on suodatettu pois aineistosta. Toinen haaste projektissa on useaan otteeseen toistuva sisältö, kuten eri sivustoilla esiintyvät samat mainokset tai Wikipediasta muille sivustoille kopioidut tekstit. Myös tällaiset toisinnokset on kuitenkin mahdollista tunnistaa automaattisesti. (Luotolahti ym. 2015: 212.)

Kieliopillisen jäsenityksen lisäksi Finnish Internet Parsebank -projektissa huomioidaan myös tekstilajit, joihin internet-tekstit luokitellaan niiden URL-osoitteen perusteella. Toistaiseksi kategorioita on neljä: blogit, verkkokeskustelut, tietosanakirja-sivut ja uutii-

set. Tulokset osoittavat luokittelun toimivan hyvin sekä riittävän luotettavasti, ja tutkimusryhmän onkin tulevaisuudessa tarkoitus määritellä lisää kategorioita neljän edellä mainitun rinnalle. Jaottelun kannalta ongelmallisia ovat tapaukset, joissa yksittäisessä tekstissä on piirteitä useammasta kuin vain yhdestä tekstilajista. (Laippala, Luotolahti, Kyröläinen, Salakoski & Ginter 2017: 152.)

Aiemmin kaikki nettikorpusten sisältämä teksti on ollut tasalaatuista, joten Turku NLP Groupin kehittämä tekstilajiluokittelu edesauttaa jatkossa aineiston hyödyntämistä tutkimuskäytössä ja parantaa sen käytettävyyttä (Laippala ym. 2017: 152–153). Tekstilajiluokat eivät kuitenkaan vielä ole mukana käyttöliittymässä, joten niitä ei myöskään omassa tutkimuksessani varsinaisesti oteta huomioon. Teen kuitenkin omiin havaintoihini perustuen joitain havaintoja ja yleistyksiä myös tekstilajeihin liittyen, sillä alkuperäinen internet-sivu on käytännössä aina mahdollista jäljittää tavallisella Googlehaulla.

Käyttöjärjestelmä on poiminut osumat satunnaisesti, joten aineistossa ovat monipuolisesti edustettuina erilaiset tekstilajit ja -tyypit. Toisaalta se, että virkkeessä edellytetään esiintyvän sana *hammas/dent* väistämättä rajaa saatuja osumia. Näin ollen aineistosta ei löydy esimerkiksi lakitekstejä eikä muutamaa poikkeusta lukuun ottamatta uutisia, kun taas mainokset ja lääketieteelliset julkaisut ovat erityisen hyvin edustettuina.

Ranskankielinen aineisto French Parsebank toimii käyttöliittymässä samalla tavoin kuin suomenkielinen. Sekä ranska että suomi kuuluvat ensimmäisiin kieliin, joilla Universal Parsebanks -aineistot julkaistiin vuonna 2015. Myös ranskankielistä aineistoa on sen jälkeen päivitetty useampaan kertaan. Viimeisimmän version takana ovat Marie-Catherine de Marneffe, Bruno Guillaume, Matias Gironi, Carly Dickerson ja Guy Perrier. (GitHub – Universal Dependencies/UD_French-GSD 10.9.2018.)

3.3. Aineiston keruu

Aineiston keruussa käytetty metodi on yksinkertainen. Käyttöliittymä löytyy jo edellä mainitusta osoitteesta bionlp-www.utu.fi/dep_search/. Sivun yläalaidassa ensimmäisenä on pudotusvalikko, josta valitaan kieli. Suomenkielisen aineiston kohdalla valitaan "Finnish Internet Parsebank" ja ranskankielisen "French Parsebank max 2M". Valikkoa seuraa hakuikkuna, johon sana(t) kirjoitetaan. Hakuja on mahdollista tehdä niin sanojen kuin syntaksirakenteidenkin mukaan, mutta tässä tapauksessa kyse on ensin mainitusta.

Kun haettavan sanan eteen lisätään kirjain *L* sekä yhtäsuuruusmerkki =, kohdistuu haku kaikkiin kyseisen sanan taivutusmuotoihin. Suomenkielisen aineiston keräämisessä käytetty hakusana on näin ollen ”L=hammas” ja ranskankielisen ”L=dent”. Viimeisenä valitaan omasta pudotusvalikostaan haluttujen tulosten määrä. Vaihtoehtoja siinä ovat 10, 50, 100, 250 ja 1000.

Alun perin tarkoituksena oli valita molemmista kielistä 100 ensimmäistä osumaa tarkasteltaviksi. Määrä tuntui kuitenkin liian pieneltä kattavan kuvan saamiseen ja yleistysten tekemiseen, joten päädyin valitsemaan 200 ensimmäistä. Valitsin molemmissa hauissa siis pudotusvalikosta vaihtoehdon ”250”. Suomenkielisiä tuloksia löytyi huomattavasti enemmänkin (lähes 500), mutta yllättäen ranskankielisiä osumia ainoastaan 198. Määrä on kuitenkin tarpeeksi lähellä 200:a, joten otin tarkasteluun ne kaikki. Tasapuolisuuden vuoksi valitsin myös suomenkielisistä osumista 200:n sijasta 198 ensimmäistä.

Suoritin aineiston keruun 10. – 11.4.2018, ja kaikki saadut tulokset perustuvat kyseisenä ajankohtana saatuihin osumiin. Jos/kun aineistoa muutetaan tulevaisuudessa, on mahdollista, että käyttöliittymästä saatavat tulokset ovat myöhemmin toistettaessa poikkeavia. Osumien esiintymisjärjestys ja näin ollen niiden päätyminen osaksi tutkimusaineistoa on sattumanvaraista (suullinen tiedonanto Veronika Laippala 18.1.2018).

Käyttöliittymässä pystyy suodattamaan tuloksia usein eri tavoin, ja niitä on mahdollista tarkastella hyvinkin yksityiskohtaisesti syntaktisella tasolla. Tässä tutkimuksessa tarkoituksena on selvittää ainoastaan se, kuinka usein sanat *hammas* ja *dent* esiintyvät osana idiomeja sekä se, millaisissa suomen- ja ranskankielisissä idiomeissa ne esiintyvät. Näin ollen tulosten tarkastelua varten valittu menetelmä on hyvin yksinkertainen. Kävin molemmilla kielillä saadut 198 ensimmäistä osumaa läpi yksitellen niiden esiintymisjärjestyksessä. Totesin jokaisen kohdalla, onko kyseessä idiomi vai ei (idiomien tunnistamisesta tarkemmin s. 32–33). Näin muodostuivat tulokset, joista on helppo todeta molemmissa aineistoissa esiintyvien idiomien määrä sekä se, mitkä idiomit niissä esiintyvät. Myös ei-kuvaannollisessa merkityksessä esiintyvistä tapauksista, joita ylivoimaisesti suurin osa osumista edustaa, on helppo tehdä tulkintoja sen suhteen, millaisissa yhteyksissä tutkimuksessa käytetyt hakusanat tyypillisesti esiintyvät suomessa ja ranskassa.

Jokainen käyttöliittymän antama tulos on kielestä riippumatta yhdestä kokonaisesta virkkeestä koostuva. Useimmissa tapauksissa se riittää antamaan tarpeeksi laajan kuvan kontekstista oikean tulkinnan tekemiseen. Näin ei kuitenkaan aina ole. Alkuperäinen

tekstiyhteys, jossa virke esiintyy, on kuitenkin kaikissa tapauksissa mahdollista jäljittää kirjoittamalla se Googlen hakukenttään kokonaislainausmerkkien (” ”) väliin. Tästä huolimatta niin suomen- kuin ranskankielisissä tuloksissa esiintyy muutama yksittäinen epäselvä tai muutoin tulkinnanvarainen kohta. Se ei kuitenkaan ole yllättävää kun otetaan huomioon se, että kuka tahansa on saattanut tuottaa kyseisen virkkeen internetiin. Epäselville tapauksille yhteistä onkin se, että ne ovat esimerkiksi sekoitus useammasta eri idiomista tai muutoin antavat vaikutelman idiomaattisesta käytöstä olematta silti varsinaisia, tunnistettavia idiomeja. Ongelmallisia tapauksia käsitellään tarkemmin aineiston analyysin yhteydessä luvussa 4.

Tutkimuksessa käytettävän aineiston etuja ovat jo edellä mainitut virkkeiden sattumanvarainen valikoituminen sekä melko laaja otanta. Näiden lisäksi vertailun kannalta ehdottoman tärkeää on kahden aineiston vertailukelpoisuus. Edellytyksinä onnistuneelle vertailulle ovat yhtenäiset lähtökohdat ja aineistojen riippumattomuus toisistaan (Järventausta 2013: 103), mitkä tässä tapauksessa toteutuvat. Mitä aineiston heikkouksiin tulee, on syytä huomioida sen antama suppea kuva kielenkäytöstä. Vaikka internettekstit edustavat varsin laajaa kirjoa aina mainoksista ja puhekielisistä keskusteluista uutisiin ja tieteellisiin tutkielmiin, jäävät tarkastelun ulkopuolelle luonnollisesti kaikki puhuttu kieli sekä vain painetuissa lähteissä esiintyvä teksti. Lisäksi vajaan 200 virkkeen otanta on toki rajallisesti edustava.

3.4. Luokittelu

Idiomeja on mahdollista luokitella monin eri tavoin. Kategorisointiin vaikuttaa luonnollisesti se, mistä näkökulmasta idiomeja lähestyy ja mitä ominaisuutta niistä halutaan tarkastella. Monissa luokitteluissa mukana on idiomien lisäksi myös muunlaisia kiteytyneitä ilmauksia.

Tässä tutkimuksessa aineiston luokittelu perustuu Korhosen (2018: 50–53) esittämään nelikategoriseen jaotteluun. Mitä tulee syihin, joiden vuoksi luokittelua ylipäätään tehdään, toteaa Miestamo (2013: 49) seuraavasti:

Luokittelu ei ole itsetarkoitus, vaan se tehdään aina jotakin tarkoitusta varten. Luokittelun on siis oltava hyödyllinen esimerkiksi teorianmuodostuksen ja selitysten etsimisen kannalta.

Luokittelua voidaan siis pitää ennen kaikkea apuvälineenä varsinaisen tavoitteen saavuttamisessa. Tässä tapauksessa koen valitsemani kategorisoinnin luontevaksi ja selkeäksi keinoksi ryhmitellä ja jäsentää korpuksessa esiintyvät idiomit. Lisäksi se auttaa vertailemaan suomen- ja ranskankielisiä idiomeja keskenään. Luokittelu perustuu idiomien morfosyntaktiseen luonteeseen, eli se nostaa esiin ilmausten rakenteellisiin ominaisuuksiin liittyviä eroja ja yhtäläisyyksiä.

Korhosen (2018: 50–53) käyttämät neljä kategoriaa ovat: verbifraseemit, substantiivifraseemit, adjektiivifraseemit ja adverbifraseemit. Aineistoni koostuessa yksinomaan idiomeista tulen kuitenkin jatkossa käyttämään edellä luetelluista nimityksiä verbi-idiomit, substantiivi-idiomit, adjektiivi-idiomit ja adverbi-idiomit. Jokaisella kategorialla on enemmän tai vähemmän alaluokkia, joita esitellään seuraavaksi. Suluissa annetut esimerkit ovat kaikki Korhosen käyttämiä. Joissain tapauksissa suomenkielisen esimerkin keksiminen olisi tuottanut vaikeuksia, joten yhteneväisyyden vuoksi pitäydyin annetuissa englanninkielisissä. Hankaluuksia suomenkielisten esimerkkien keksimisessä aiheuttavat ennen kaikkea moniin alaluokkiin sisältyvät adpositiot, joita suomessa on useisiin muihin kieliin verrattuna hyvin vähän.

Eniten alaluokkia löytyy verbi-idiomeista, jotka ovatkin selvästi laajin luokka. Ne voidaan yksinkertaisimmillaan jaotella verbin luonteen perusteella infinitiiviin (*to spill the beans*), indikatiivin preesenssiin (*sb.s hackles rise*) ja sanapariin (*to wait and see*). Lisäksi verbi-idiomeja voidaan ryhmitellä verbistä riippuvien jäsenten lukumäärään perustuen. Jos ilmauksen verbillä on yksi alisteinen jäsen, voi tämä jäsen olla esimerkiksi infinitiivin/partisiipin muotoinen verbi (*to be taken with sb*), (prepositionaalinen) substantiivi (*to be in touch with sb; to beggar description*), substantiiviryhmä (*to press every button*) tai (prepositionaalinen) adjektiivi (*to go short*). Mahdollisia alisteisia jäseniä ovat lisäksi muun muassa lukusana (*to take five*) ja interjektio (*to say ugh to sb*). Toisinaan alisteisia jäseniä saattaa olla yhden sijasta kaksi tai peräti kolme, jolloin mahdollisia yhdistelmiä ovat esimerkiksi substantiivi + prepositionaalinen substantiivi (*not to see the wood for the trees*) tai kaksi prepositionaalista substantiivia + adverbi (*to stand with both feet firmly on the ground*). (Korhonen 2018: 50–52.)

Substantiivi-idiomit puolestaan jaetaan prepositionaaliin ja ei-prepositionaaliin, joista molemmat jakautuvat edelleen idiomin jäsenten sanaluokkien perusteella useampaan alaluokkaan. Prepositiota saattaa seurata esimerkiksi sanapari (*by leaps and bounds*) tai

substantiivi, jolla on attribuutti (*with a grain of salt*). Yksi mahdollinen yhdistelmä on myös prepositio + substantiivi + prepositio + substantiivi (*from cradle to grave*). Ei-prepositionaaliset substantiivi-idiomit taas voivat koostua esimerkiksi sanaparista (*bag and baggage*) tai kieltosanasta ja substantiivista (*not a scrap*). (Korhonen 2018: 52–53.)

Alaluokkien määrä vähenee huomattavasti siirryttäessä adjektiivij- ja adverbi-idioimeihin. Näistä molemmat jakautuvat vain prepositionaalisiin (*in brief; by now*) ja ei-prepositionaalisiin (*dead and buried; now and again*) (Korhonen 2018: 53). Mahdollinen yhdistelmä on myös adjektiivista ja adverbista koostuva sanapari (emt.).

Koska peruste aineiston idiomien valikoitumiselle on *hammas/dent*-sanon esiintyminen, sulkee se tässä tapauksessa kokonaan pois kaksi viimeistä kategoriaa. Mikäli idiomiin sisältyy verbi (*purra hammasta*), on kyseessä verbi-idiomi. Jos taas verbi ei ole osallimausta, on pääkomponenttina kaikissa aineiston tapauksissa sana *hammas/dent* (*pitkin hampain*), eli substantiivi. Toisin sanoen idiomien pääsanana ei koskaan ole adjektiivi tai adverbi. Analyysivaiheen luokittelu supistuu siis neljän sijaan kahteen pääluokkaan.

On mielenkiintoista nähdä, onko suomen ja ranskan välillä havaittavissa eroja verbillisten ja verbittömien idiomien jakaumassa ja jos näin on, mikä voisi olla siihen syynä. Analyysissa on siis mukana sekä määrällinen että laadullinen näkökulma. Tarkastelen suomen- ja ranskankielisiä aineistoja yhdessä ja erikseen tehden havaintoja niiden eroista, yhtäläisyyksistä ja muista esiin nousevista piirteistä. Lisäksi tarkastelun kohteena on eri idiomien määrä ja yleisyys kummassakin kielessä.

3.5. Hammas-idiomit

Hammas on yksi 20 yleisimmistä suomenkielisissä idiomeissa esiintyvistä ruumiinosanimistä, toteaa Nenonen (2002: 115). Korhonen (2018) puolestaan ei ota erikseen kantaa kyseisen sanon esiintyvyyteen, mutta huomauttaa kielessä olevan sanoja, joilla on ”huomattava taipumus muodostaa fraseemien osia” (2018: 130). Tällaisia ovat tyypillisesti esimerkiksi eläin-, väri- ja ruumiinosaviitteiset sanat (emt.). Idiomeja, jotka sisältävät yhden tai useamman ihmisen/eläimen ruumiinosan tai elimen nimityksen kutsutaan somatismeiksi (Korhonen 2018: 131). Korhosen mukaan somatimit ovat keskeisessä asemassa fraseologiassa sen vuoksi, että niiden avulla pystytään ilmaisemaan ihmisen olennaisimpia tuntemuksia, ominaisuuksia ja toimintoja (emt.).

Hammas-aiheisten idiomien yleisyys huomioiden voidaan olettaa, että monipuolisesta kielenkäytöstä koostuvassa internet-aineistossa esiintyy sanan sisältäviä idiomeja. *Hammas* ei kuitenkaan ole maailman kielissä varsinaisesti kieliopillistunut, kuten monet muut ruumiinosannimet. Tämä käy ilmi esimerkiksi siitä, ettei sana esiinny lainkaan Heinen ja Kutevan (2002) kieliopillistuneita sanoja, useat ruumiinosannimet mukaan lukien, yhteen koonneessa teoksessa *World Lexicon of Grammaticalization*. Useammin kieliopillistuneita ovat esimerkiksi *käsi* ja *korva* (Ojutkangas 2001). Näin ollen *hammas*-sanan sisältävien lauseiden luokittelu sen perusteella, esiintyykö sana osana idiomia vai ei, on suhteellisen helppoa. Yksittäisiä epäselviä tai monitulkintaisia tapauksia saattaa esiintyä, mutta suuressa otannassa sillä ei liene merkittävää vaikutusta lopullisiin tutkimustuloksiin. Yleisesti ottaen voidaan todeta, että useimmissa tapauksissa *hammas*-sana esiintyy selvästi joko osana idiomia tai vaihtoehtoisesti varsinaisessa merkityksessään viitaten konkreettisesti ihmisen tai eläimen hampaisiin.

Heinosen (2013a: 94) mukaan osviittaa sanan käytöstä ja sen yleisyydestä osana idiomeja antaa mahdollisesti se, paljonko tilaa ja montako eri merkitysryhmää sanalla Kielitoimiston sanakirjassa on. Sanalle *hammas* annetaan kolme merkitysryhmää, joista ensimmäisenä on luonnollisesti sanan yleisin merkitys: ”useiden selkärankaisten suussa olevia luumuodostumia, joiden päätarkoituksena on ruoan hienontaminen”. Toinen määritelmä on ”terävähkö ulkonema; sakara, väkä, pii”, jota käytetään esimerkiksi sahan tai postimerkin hampaista puhuttaessa. Kolmas ja viimeinen määritelmä on seuraava: ”kuv. ajan hammas: ajan kuluttava vaikutus”. Tapauksista, joissa sanaa *hammas* käytetään kuvaannollisessa merkityksessä eli osana idiomia ainoastaan *ajan hammas* siis muodostaa oman kategoriansa. Kaikki muut ilmaukset puolestaan sijoittuvat Kielitoimiston sanakirjan määrittelyssä ensimmäisen merkitysryhmän alle.

Ensimmäisen merkitysryhmän yhteydessä Kielitoimiston sanakirja tarjoaa useita esimerkkejä niin sanan kirjaimellisesta (*tekohampaat; hammas puhkeaa; saada hampaita*) kuin kuvaannollisestakin (*olla ihmisten hampaissa; kahvihammasta kolottaa*) käytöstä. Kuvaannollisiksi määritellyt ilmaukset ovat kaikki idiomeja, Sekä kirjaimellista että kuvaannollista käyttöä kuvaavaksi on luokiteltu esimerkiksi *kiristellä hampaitaan*, jonka merkitys voidaan tulkita eri tavoin kontekstista riippuen.

Yksikielinen ranskan sanakirja *Le Robert de poche 2009* (2008) tarjoaa kaksi määritelmää ranskankieliselle *hammas*-sanan vastineelle *dentille*. Ensimmäisen määritelmän

mukaan kyse on ihmisen tai eläimen suussa esiintyvistä hampaista, kun taas jälkimmäinen määritelmä on kaikessa yksinkertaisuudessaan 'kuvaannolliset ilmaukset' (*locutions figurées*) ja se sisältää siten kaikki idiomit, joissa sana esiintyy.

Idiomi- ja muita sanakirjoja (esim. Tervomaa 2002: *Päin mäntyä*; Kari 1993: *Naulan kantaan* ja Rey & Chantreau 2003: *Dictionnaire des expressions et locutions*) tarkasteltaessa havaitaan useita erilaisia aihepiirejä, joiden alle *hammas*-aiheisia idiomeja on mahdollista jaotella. Selvästi suurin osa teemoista niin suomessa kuin ranskassa ilmentää negatiiviseksi miellettyjä tunteita. Havaintoa tukee myös Korhosen (2018: 125) näkemys, jonka mukaan useille idiomeille (fraseemeille) tyypillistä on nimenomaan negatiivinen arvottaminen.

Rey ja Chantreau (2003: 297–299) nostavat idiomisanakirjassaan esiin asioita, joita sana *dent* symboloi. Useimmin toistuvia teemoja ovat vastentahtoisuus, vihamielisyys, uhkaaminen sekä kauna, jotka ovat pitkälti samoja kuin suomalaisissa *hammas*-idiomeissa. Suuri joukko idiomeja onkin niin muodon kuin merkityksen puolesta tismalleen samoja molemmissa kielissä. Seuraavaksi mainituista esimerkeistä osa on poimittu tutkimuksen aineistosta. Aineistossa esiintymättömät ranskankieliset ilmaukset ovat peräisin Reyn ja Chantreaun (2003) idiomisanakirjasta. Suomenkieliset, aineistossa esiintymättömät idiomit puolestaan ovat sellaisia, jotka äidinkielenä puhujana olen oman kielentutkimukseni perusteella katsonut ranskankielisiä vastaaviksi.

Vastentahtoinen henkilö *puree hammasta* (*serrer les dents*), mistä seuraa *itkua ja hammasten kiristelyä* (*pleurs et grincements de dents*). Vihaisemmaksi sävy muuttuu, kun ihminen alkaa *kiristellä hampaitaan* (*grincer des dents*). *Hampaiden näyttämiseen* (*montrer les dents*) liittyy pelottelu ja halu saada toinen kokemaan olonsa uhatuksi. Siinä missä ihmisestä puhuttaessa ilmaus on tulkittavissa lähinnä kuvaannollisesti, on eläinten kohdalla kyse yleensä kirjaimellisestikin hampaiden näyttämisestä. Uhattu osapuoli puolestaan *joutuu toisen hampaisiin*. Hieman yllättäen vastaava ranskankielinen ilmaus *être sur les dents* ('olla hampaissa') viittaakin kiireisenä olemiseen.

Edellä mainitun lisäksi on myös muita tapauksia, joissa toisiaan muistuttavat idiomit saattavatkin merkitä osin tai täysin eri asioita. Suomessa pitkähampaisuus (*tehdä jtkä pitkin hampain*) viittaa vastentahtoisuuteen, kun taas ranskankielinen ilmaus *avoir les dents longues* ('jklä on pitkät hampaat') tarkoittaa, että henkilö on erittäin kunnianhimoinen. Idiomille *pitkin hampain* kuitenkin löytyy myös ranskankielinen *dent*-sanan

sisältävä vastine *du bout des dents*. Siinä tekeminen tapahtuu hampaiden kärjillä, mikä tuntuu varsin loogiselta.

Toisinaan kahdella erikielisellä idiomilla on sama merkitys, mutta niiden muodossa on pientä eroavaisuutta. Esimerkiksi kaunaa kantavalla suomalaisella *on jotakin hampaan-kolossa* toista kohtaan, kun taas saman tunteen jakavalla ranskalaisella on 'hammas jotakuta vastaan' (*avoir une dent contre qqn*).

Osa *hammas/dent*-idiomeista edustaa yksittäistä merkityskategoriaa, kuten molemmissa kielissä samanmuotoinen ja -merkityksinen *olla hampaisiin asti/hampaita myöten aseistautunut* (*être armé jusqu'aux dents*). Ranskalainen ilmaus *quand les poules auront les dents* ('kun kanoilla on hampaat') taas viittaa erittäin epätodennäköisesti tapahtuvaan asiaan. Luonteva suomennos sille on *kun lehmät lentää*. Muita oman merkitysryhmänsä muodostavia *hammas*-idiomeja ovat esimerkiksi *ilman sarvia ja hampaita* (puhua vilpittömästi/kiertelemättä) sekä *panna hampaat naulaan* (olla köyhä; elää puutteessa).

Kenties produktiivisin ja siten mielenkiintoisin on suomalainen idiomiksi *kahvihammasta kolottaa*. Kuten Heinonen (2013b) toteaa, on tullut tavaksi hyödyntää samaa pohjakavaa myös muiden mielitekojen ilmaisemiseen. Esimerkkeinä hän mainitsee kolottavan/pakottavan *viina-*, *karkki-*, *sokeri-*, *nami-* ja *herkkuhampaan*. Näiden lisäksi Heinonen (2013b) mainitsee törmänneenä internet-teksteissä myös *hääkaku-*, *makkara-*, *pavlova-*, *sahti-*, *sipsi-* ja *viskihampaiden* kolotukseen. Erikoisimmista tapauksista idiomin käyttö on laajentunut koskemaan myös ei-syötäviä asioita, kun jollakulla on alkanut pakottaa *auto-* tai *mökkihammasta* (emt.). Mahdollisuuksia erilaisille ja vieläpä toimiville sanavaihdoksille on siis paljon. Harvaa idiomia voitaneen pitää yhtä produktiivisena.

Niin ranskaa kuin suomeakin tarkasteltaessa voidaan todeta, että sana *kynsi* (ra *ongle*) on jossain määrin verrattavissa *hampaaseen*. Sana *kynsi/ongle* vaikuttaa molemmissa kielissä esiintyvän selvästi harvemmin osana idiomia kuin *hammas/dent*, mutta joitakin yhtymäkohtia on silti havaittavissa etenkin suomen osalta. Esimerkiksi idiomeissa *näyttää hampaansa/kyntensä* sekä *joutua jonkun hampaisiin/kynsiin* merkitys säilyy samana huolimatta siitä, kumpaa sanaa käytetään. Ilmauksessa *kynsin hampain* puolestaan esiintyvät molemmat. Se, että sanat ovat joissain tapauksissa korvattavissa toisillaan, tuntuu loogiselta. Ovathan hampaat ja kynnet ainoita teräviä ruumiinosia, ja näin ollen ne on luonnollista kokea uhkaavina. Mitä ranskaan tulee, on idiomeissa *être armé jusqu'aux*

dents (olla hampaita myöten aseistautunut) muinoin käytetty *dentin* tilalla sanaa *ongles* (Rey & Chantreau 2003: 297).

3.6. Kielten vertailu

Kontrastiivinen lingvistiikka on yksi kielentutkimuksen osa-alueista (Järventausta 2013: 96). Sen tavoitteena on kuvata eri kielten välisiä eroja ja yhtäläisyyksiä (emt.). Pääasiallisena lähteenä menetelmän esittelyssä käytän Järventaustan (2013) artikkelia, johon on koottu vertailun keskeisiä periaatteita ja käsitteitä.

Tutkimuksessani tarkastelen suomen- ja ranskankielisiä internet-korpuksia erillisinä kokonaisuuksina, mutta ennen kaikkea vertaillen niitä toisiinsa. Koska kyseessä on kaksi erikielistä keskenään vertailukelpoisia alkuperäistekstejä sisältävää korpusta, voidaan aineistosta puhua verrannollisena korpuksena (Järventausta 2013: 111). Sitä, mikä on käännettyjen ja mikä alkuperäistekstien osuus korpuksista, on vaikea arvioida. Toisaalta se ei myöskään ole tutkimuksen kannalta oleellista.

Vertaileva kielentutkimus on joko diakronista tai synkronista (Järventausta 2013: 96). Tämä tutkimus edustaa jälkimmäistä, sillä historiallisen muutoksen sijaan tarkastelun kohteena ovat tällä hetkellä internetissä olevat tekstit. Niiden julkaisuajankohdalla ei tutkimuksen kannalta ole merkitystä. Valtaosa teksteistä on kuitenkin oletetusti internetin aikakaudella tuotettuja, eli historiallisesta näkökulmasta tuoreita.

Synkroninen vertaileva kielentutkimus voidaan jakaa edelleen kielitypologiseen ja kontrastiiviseen tutkimukseen. Molemmissa suuntauksissa pääpaino on kielen käytön tarkastelun sijaan kielijärjestelmien tutkimisessa. Kielitypologisessa tutkimuksessa näkökulma on kuitenkin tyypillisesti funktionaalinen, kun taas kontrastiivisessa kielentutkimuksessa perinteisesti rakenteellinen. Lisäksi ensin mainitussa tutkimussuuntauksessa tarkastelun kohteena on useita kieliä, ja näin ollen valittujen ilmiöiden tarkastelu jää usein melko yleiselle tasolle. Siinä kiinnostus kohdistuu ennen kaikkea kieltenvälisiin yhtäläisyyksiin. Kieltenvälisiin eroihin keskittyvässä kontrastiivisessa kielentutkimuksessa puolestaan tarkastellaan tyypillisesti vain kahta kieltä, jolloin eroavaisuuksia voidaan tutkia hyvinkin perusteellisesti. (Järventausta 2013: 96–97.)

On kuitenkin syytä huomioida, että yllä esitetyt piirteet ovat nimenomaan kahdelle mainitulle suuntaukselle tyypillisiä ominaisuuksia, eivät ehdottomia rajanvetoja. Järventausta (2013: 98) huomauttaakin, että kontrastiivisen kielentutkimuksen kentällä on siir-

rytty vähitellen pelkistä rakenteista kohti tekstien ja ylipäätään kielen käytön tarkastelua. Myös tässä tutkimuksessa huomio on ennen kaikkea kielen todellisessa käytössä, vaikka tutkimus onkin luokiteltavissa enemmän kontrastiiviseksi kuin kielitypologiseksi. Luokittelu kontrastiiviseksi tutkimukseksi perustuu lähinnä tarkasteltavien kielten (suomi ja ranska) määrään, sillä kiinnostuksen kohteena tutkimuksessa ovat niin erot kuin yhtäläisyydetkin idiomien käytössä kahden eri kielen välillä.

Kontrastiivisen kielentutkimuksen analyysissä on kolme vaihetta: kuvaus, verrattavuuden toteaminen ja varsinainen kontrastointi (Järventausta 2013: 102). Kuten jo edellä käy ilmi, kontrastiivinen kielentutkimus on perinteisesti keskittynyt lähinnä kielen rakenteisiin, ja kielioppia käsiteltäessä juuri *kuvaus* onkin osuva nimitys. Kielen käyttöä tarkasteltaessa kuvauksen voidaan kuitenkin tulkita käytännössä viittaavan aineiston hankkimiseen molemmista kielistä, tässä tapauksessa siis korpuksen hankkimiseen ja rajaamiseen. Näiltä kielikohtaisilta kuvauksilta, eli keskenään vertailtavilta aineistoilta, edellytetään yhtenäisiä lähtökohtia ja riippumattomuutta toisistaan (Järventausta 2013: 103).

Toinen vaihe kontrastiivisessa analyysissä on verrattavuuden toteaminen. Siinä määritellään tutkimuksen kannalta oleellinen vertailuperusta eli *tertium comparationis*. Kyse on niin sanotusti muuttumattomasta suureesta, jota kutsutaan kielitieteessä myös nimellä 'vertailun kolmas'. *Tertium comparationis* on siis perusta, johon kielikohtaiset erot ja yhtäläisyydet suhteutetaan. Järventausta luonnehtii sitä ”kontrastiivisen analyysin välttämättömäksi edellytykseksi ja verrattavuuden takeeksi”. (Järventausta 2013: 104.)

Järventaustan (2013: 107) mukaan ei ole olemassa yhtä yhtenäistä vertailuperustaa, joka soveltuisi käytettäväksi kaikissa kontrastoinneissa. Sen sijaan jokaista erilaista tutkimusta varten on luotava oma vertailuperustansa, tai tarvittaessa useampikin. Tekstilähteisessä kontrastoinnissa, mitä tämäkin tutkimus edustaa, *tertium comparationiksena* toimii parhaiten käänös- tai tilastollinen ekvivalenssi (Järventausta 2013: 106). Koska tässä tutkimuksessa tarkastellaan yksittäisen ilmiön esiintymistä kahdessa erikielisessä aineistossa ilman käänösnäkökulmaa, soveltuu lähtökohdaksi tässä tapauksessa tilastollinen ekvivalenssi. Tilastollinen ekvivalenssi toteutuu, mikäli kontrastoitavilla ilmiöillä on ”tilastollisesti samankaltainen jakauma verrattavien kielten verrannollisissa teksteissä” (emt.).

Jo ennen aineistojen tarkastelua ja vertailua voidaan muodostaa erilaisia olettamuksia. Sekä suomessa että ranskassa on lukuisia *hammas*-sanan sisältäviä idiomeja, kuten osiosta 3.5 käy ilmi. Voidaan siis olettaa, että niitä esiintyy vähintään jonkin verran molemmista aineistoissa. Joissain tapauksissa molemmista kielistä löytyy kirjaimellisesti ja merkityssisällöltään täysin vastaava idiomi (esimerkiksi *kiristellä hampaitaan* = *grincer des dents*). Ja vaikka kaikille idiomeille ei löydy muodon puolesta vastaavaa ilmausta toisesta kielestä, ovat aihepiirit usein yleisesti ottaen samoja (esimerkiksi pelko ja vastentahtoisuus). Tarkemmista yhtäläisyyksistä tai eroista aineistojen välillä on kuitenkin hankala sanoa etukäteen mitään, samoin kuin esiintymien määrästä.

Vertailuperustan löytäminen ennen analyysin kolmatta vaihetta eli varsinaista kontrastointia, jossa aineistoja verrataan kohta kohdalta toisiinsa (Järventausta 2013: 102), ei ole aivan yksinkertaista. Se ei toisaalta ole myöskään tarpeen, toteaa Chesterman (1998: 59). Siinä missä *tertium comparationis* määrittellään perinteisesti tutkimuksen lähtökohdaksi, pitää Chesterman (emt.) sitä pikemminkin tutkimuksen päämääränä, joka syntyy kontrastiivista analyysia tehtäessä. Hän kuvaa *tertium comparationista* kontrastoitavien kielten yhteiseksi perustaksi, joka tutkijan tulee löytää analyysin aikana (emt., 30). Sen sijaan lähtökohtana tutkimuksessa voidaan pitää kahden verrattavan kielen välillä havaittua samankaltaisuutta ilman yksityiskohtaisempaa kuvausta (emt., 59).

Myös Järventausta (2013: 113) huomauttaa, että kontrastiivinen analyysi on toisinaan mahdollista tehdä ilman *tertium comparationiksen* määrittelyä missään vaiheessa. Tällöin kyse on kuitenkin aina yksisuuntaisesta kontrastoinnista. Siinä vertailuperustana ovat lähtökieleksi valitun kielen ilmaukset tai muut ominaisuudet, joihin toista kieltä verrataan. Yksisuuntaisen kontrastoinnin etuna on, ettei *tertium comparationista* tarvitse lainkaan määrittellä (emt., 115). Toisaalta tällöin tarkastelu on helposti suppeaa, jolloin kattavan kokonaiskuvan saaminen kontrastoitavasta ilmiöstä hankaloituu (emt.).

Sen sijaan kaksisuuntaisessa kontrastoinnissa, jossa kieliä verrataan toisiinsa yhteisen *tertium comparationiksen* kautta, on helpompi saada laaja kokonaiskuva verrattavasta ilmiöstä. Toisaalta siinä ongelmia saattaa tuottaa nimenomaan sopivan vertailuperustan löytäminen ja muotoileminen. *Tertium comparationikselta* edellytetään aina riippumattomuutta verrattavista kielistä, mikä antaa tutkittavasta ilmiöstä usein kattavamman ja todenmukaisemman kuvan kuin mitä yksisuuntaisella kontrastoinnilla pystyttäisiin saavuttamaan. (Järventausta 2013: 120.)

Tulosten todenmukaisuuden ja yleistettävyyden takaamiseksi sovellan tutkimuksessa kaksisuuntaista kontrastointia. Äidinkielenä suomenpuhujana ranskankielisen aineiston vertaaminen suomenkieliseen olisi toki luontevaa, mutta koen aineistojen erillisen ja tasapuolisen käsittelyn selkeämmäksi. Vertailen siis yhtä lailla suomenkielistä aineistoa ranskankieliseen kuin myös päinvastoin.

Kuten edellä käy ilmi, on yksi tämän tutkimuksen *tertium comparationis* tilastollinen ekvivalenssi, jossa huomion kohteena on tarkasteltavien idiomien yleisyys. Toisin sanoen se, kuinka usein hammas-idiomit esiintyvät internet-teksteissä. Vaikka idiomien laadun lisäksi tarkastelukohteena tutkimuksessa on myös niiden esiintyvyys, on tilastolliseen ekvivalenssiin vertailuperustana syytä suhtautua tässä kriittisesti. Jotta todella voitaisiin verrata suomea ja ranskaa tilastollisesta näkökulmasta, vaatisi se tarkkojen todennäköisyyksien laskemista. Tällöin varsinaisten yleistystenkin tekeminen olisi mahdollista ja perusteltua. Tällä kertaa pääpaino on kuitenkin laadullisessa tarkastelussa, ja idiomien esiintyvyyteen liittyvät havainnot ovat lähinnä suuntaa antavia, eivät tarkkoja tai yleistettävissä olevia huomioita.

Toinen vertailuperusta, joka on samalla edellytyksenä idiomien yleisyyden tarkastelulle, on idiomien määritelmä itsessään. Kun idiomit poimitaan aineistojen virkkeistä, on jokin niistä jollain perusteella luokiteltava idiomiksi. Kriteerit aineiston idiomien tunnistamiselle on esitetty alaluvun 2.4. lopussa. Niitä ovat kuvaannollisuus, leksikaalisuus sekä polyleksikaalisuus. Etenkin suomenkielisen aineiston kohdalla tunnistaminen on äidinkieliselle puhujalle lähes kaikissa tapauksissa helppoa, mutta ranskankielisten virkkeiden tapauksessa vaikeampaa vaati enemmän taustatyötä ja apuvälineitä kuten idiomisanakirjoja. Kolmantena vertailukohtana voidaan pitää alaluvussa 3.4. esitettyä Korhosen luokittelua. Sovellan samaa rakenteellista luokittelutapaa, eli jakoa verbi- ja substantiivi-idiomeihin, molemmissa kielissä.

4. Aineiston analyysi

Ensimmäisessä alaluvussa 4.1. esittelen omilla osioissaan kaikki suomen- ja ranskan-kielisten aineistojen *hammas*-idiomit (4.1.1. ja 4.1.2.). Suomenkielisessä aineistossa esiintyi yhteensä 12 erilaista hammas-idiomia. Kaikkiaan hammas-idiomeja oli 37. Ranskankielisessä aineistossa erilaisia idiomeja oli 20, jotka esiintyivät aineistossa 44 kertaa. Kaikki erilaiset idiomit on koottu ensimmäisessä alaluvussa yhteensä neljään taulukkoon: verbi-idiomit (su), substantiivi-idiomit (su), verbi-idiomit (ra) ja substantiivi-idiomit (ra). Taulukoista ilmenee kunkin idiomien rakenne sekä tarkka esiintymismäärä aineistossa.

Kussakin rakenneanalyysiosion taulukossa idiomit on listattu niiden yleisyysjärjestyksessä alkaen yleisimmästä ja päättyen idiomeihin, jotka esiintyvät aineistossa ainoastaan kerran. Osassa tapauksista yhdeksi idiomiksi on luokiteltu saman idiomien useita erilaisia variantteja. Tällöin yksittäiset variantit ovat omilla riveillään, mutta samaa idiomiä edustavat eri variantit ovat koottuina yhteen ja samaan ruudukkoon. Esimerkiksi heti ensimmäisen taulukon alussa kaikkien neljän variantin (*joutua jkn hampaisiin*, *jättää jkn hampaisiin*, *olla jkn hampaissa* sekä *päästä jkn hampaista*) katsotaan edustavan yhtä idiomiä. Jokaisen variantin perässä on ilmoitettu sulkeissa kyseisen variantin esiintymismäärä aineistossa, ja taulukon kolmannessa sarakkeessa puolestaan kaikkien varianttien yhteenlaskettu esiintymismäärä.

Rakenneanalyysia seuraa semanttisen analyysin osio (4.2.), jossa tarkastellaan molempien kielten aineistossa esiintyviä idiomeja siitä näkökulmasta, ovatko ne sävyiltään negatiivisia, neutraaleja vai positiivisia. Alaluvussa 4.3. tarkastellaan idiomeissa esiintyvää variaatiota. Ei-idiomaattisia tapauksia käsitellään lyhyesti viimeisessä alaluvussa 4.4.

4.1. Rakenneanalyysi

4.1.1. Suomenkielisen aineiston idiomit

Taulukkoon 1 ovat koottuina kaikki suomenkielisessä aineistossa esiintyvät verbi-idiomit mahdollisine variantteineen. Käytetyn luokittelutavan mukaan suomenkielisessä aineistossa esiintyy kahdeksan eri verbi-idiomiä. Ensimmäisessä taulukossa tapauksia, joissa *hammas*-sanaa ylipäätään käytetään osana idiomiä, on kaikkiaan 20.

VERBI-IDIOMIT (SU)		
Idiomi	Rakenne	Yhteismäärä aineistossa
joutua jkn hampaisiin (4) jättää jkn hampaisiin (2) olla jkn hampaissa (1) päästä jkn hampaista (1)	verbi + pron. + subst.	8
purra hammasta (3) purra hampaat yhteen (1) hammasta purren (1)	verbi + subst. verbi + subst. + adverbi subst. + verbi	5
hampaat kalisevat (1) hampaat kalisten (1)	subst. + verbi subst. + verbi	2
sihahtaa hampaidensa lomas- ta(1) /välistä(1)	verbi + subst. + postp.	2
(makean) hammasta kolottaa	subst. + verbi	1
kiristellä hampaita	verbi + subst.	1
saada hampaat kirskumaan	verbi + subst. + verbi	1
Yht.		20

Taulukko 1. Aineiston suomenkieliset verbi-idiomit

Hammasta purren ja *hampaat kalisten* ovat verbi-idiomeja ja näin ollen mukana luettelossa, vaikkakin ne E-infinitiiviin taivutettuina poikkeavat listauksen muista verbeistä. Muut verbit on taulukossa esitetty A-infinitiivimuodossa, mutta alkuperäiskonteksteissaan useimmiten taivutettuina eli finiittimuotoisina.. Lauseenjäsenenä *hammasta purren* ja *hampaat kalisten* ovat adverbiaaleja, toisin kuin muut luokituksen verbi-idiomit, jotka ovat predikaatteja. Korhonen (2018: 52) toteaaakin verbi-idiomien (verbifraseemien) olevan useimmiten predikaatteja. Samalla hän kuitenkin mainitsee myös subjektin, objektin, adverbiaalin sekä predikatiivin olevan mahdollisia lauseenjäsenrooleja. Idiomin lauseeseen tuoman merkityksen kannalta on sama asia, ilmaistaanko lause esimerkiksi muodossa *hän teki sen hammasta purren* vai *hän puri hammasta ja teki sen*.

Idiomien rakenne käy ilmi taulukon keskimmäisestä sarakkeesta, jossa on määritelty idiomien osien sanaluokka. Rakenteet niin suomen- kuin ranskankielisissä taulukoissa ovat itse määriteltyjä. Taulukosta käy ilmi, että yleisin verbi-idiomien yhdistelmä on verbi + substantiivi. Tulos ei yllätä, sillä onhan kyse nimenomaan verbillisistä idiomeista, joiden yhtenä osana on substantiivi *hammas*. Monien muiden kielten vastaavissa taulukoissa mukana olisi myös prepositioita, kuten ranskankielisen aineiston tarkastelu myöhemmin osoittaa. Suomessa samaa asiaa ajavat sijapäätteet, joita rakenneanalyysissa ei kuitenkaan ole erikseen huomioitu. Taulukossa esiintyvät kerran adverbi *yhteen* sekä postpositiot *lomasta* ja *välistä*.

Kuten taulukosta 2 käy ilmi, erilaisia substantiivi-idiomeja esiintyy suomenkielisessä aineistossa viisi ja tapauksia, joissa *hammas* on osana substantiivi-idiomia, on yhteensä kaikkiaan 17. Määrä on vain hieman pienempi verrattuna verbi-idiomeihin, joiden osalta vastaavat luvut ovat edellä mainitut 8 ja 21.

SUBSTANTIIVI-IDIOMIT (SU)		
Idiomi	Rakenne	Yhteismäärä aineistossa
kynsin hampain	sanapari	5
pitkin hampain	adj.-attrib. + subst.	5
hampaat irvessä (3) irvi hampaassa (1)	subst. + subst.	4
ajan hammas	gen.-attrib. + subst.	2
ilman sarvia ja hampaita	prep. + sanapari	1
Yht.		17

Taulukko 2. Aineiston suomenkieliset substantiivi-idiomit

Toisin kuin verbi-idiomeissa, substantiivi-idiomien kohdalla kaikki rakenteet poikkeavat jonkin verran toisistaan. Idiomien *kynsin hampain* sekä *ilman sarvia ja hampaita* rakenne on Korhosen (2018: 51) luokittelua mukaillen nimetty sanapariksi. Viimeksi mainitussa esiintyy myös suomen kielelle melko harvinaisesti prepositio. Kertaalleen taulukossa esiintyvät myös niin adjektiivi- kuin genetiiviattribuutti. *Ajan hammasta* luokun ottamatta kaikki taulukon idiomit ovat lauseenjäsenrooliltaan adverbiaaleja.

Ajan hammas esiintyy aineistossa kerran siten, että sitä edeltää verbi *kestää* ja kerran ilman verbiä. Jälkimmäisessä tapauksessa kyse on auton ilmastointihuoltoa koskevasta tekstistä, ja sanoilla leikittelevä virke menee kokonaisuudessaan seuraavasti: *Ajan hammas ja maantiesuola syövät tekniikan kuntoa. Kestää ajan hammasta* on tunnettu yhdistelmä, ja se olisi sellaisenaan ollut mahdollista sijoittaa verbi-idiomien joukkoon. *Ajan hampaalla* on kuitenkin yksi ja sama merkitys molemmissa tapauksissa, joten olisi turhaa mainita se erikseen molemmissa taulukoissa.

4.1.2. Ranskankielisen aineiston idiomit

Taulukkoon 3 ovat koottuina ranskankielisessä aineistossa esiintyvät verbi-idiomit suomennoksineen.

VERBI-IDIOMIT (RA)		
Idiomi	Rakenne	Yhteismäärä aineistossa
grincer les dents (1) grincer des dents (5) (‘kiristellä hampaita’; kirj. mm. <i>narskutella hampaita</i>)	verbi + subst.	6
croquer à pleine dent (‘nauttia täysin rinnoin’; kirj. <i>haukata täysin hampain</i>)	verbi + prep. + adj.-attrib. + subst.	5
se faire les dents sur qc (‘harjoitella’; kirj. <i>laittaa hampaat jhk</i>)	verbi + subst. + prep.	5
avoir la dent dure (‘olla hyökkäävä/armoton’; kirj. <i>olla kova hammas</i>)	verbi + subst. + adj.	2
avoir les dents longues (‘olla kunnianhimoinen’; kirj. <i>olla pitkät hampaat</i>)	verbi + subst. + adj.	2
casser les dents (‘epäonnistua’; kirj. <i>rikkoo hampaat</i>)	verbi + subst.	2
être armé jusqu’aux dents	verbi + adj. + prep. + subst.	2

(’olla hampaita myöten aseistautunut’)		
prendre le mors aux dents (’ryhtyä suurella palolla jhk’; kirj. <i>ottaa kuolain hampaista</i>)	verbi + subst. + prep. + subst.	2
(rien/qc) à se mettre sous la dent (’suuhun pantavaa’; kirj. <i>ei mitään/jtak hampaan alle laitettavaa</i>)	pron. + verbi + prep. + subst.	2
avoir le couteau entre les dents (’olla kostonhimoinen’; kirj. <i>olla veitsi hampaiden välissä</i>)	verbi + subst. + prep. + subst.	1
avoir une dent contre qqn (’olla pahantahtoinen’; kirj. <i>olla hammas jkta vastaan</i>)	verbi + subst. + prep. + pron.	1
claquer des dents (’hampaat kalisevat’; kirj. <i>kalistaa hampaita</i>)	verbi + subst.	1
montrer les dents (’näyttää hampaansa’(su. yl. ’kyntensä’)	verbi + subst.	1
murmurer entre ses dents (’mumista hampaiden lomasta’)	verbi + posses.pron. + subst.	1
péter les dents (’lyödä hampaat kurkkuun’; kirj. <i>rikkoo hampaat</i>)	verbi + subst.	1
serrer les dents (’purra hampaat yhteen’; kirj. <i>painaa/puristaa tms. hampaita</i>)	verbi + subst.	1
Yht.		35

Taulukko 3. Aineiston ranskankieliset verbi-idiomit

Siinä missä eri suomenkielisiä verbi-idiomeja oli seitsemän, on niitä samankokoisessa ranskankielisessä aineistossa 16 eli yli puolet enemmän. Eroa siis on, vaikkakaan ei huomattavan suurta. Ranskassa vaikuttaa tutkimuksen perusteella olevan enemmän erilaisia idiomeja kuin suomessa. Samoin verbi-idiomien kokonaismäärissä on aineistoissa eroa kielten välillä (20 vs. 35). Tässäkin tapauksessa on kuitenkin otettava huomioon

sattuman mahdollisuus ja aineistojen suppeus. Rakenteen osalta yllä olevan taulukon idiomit eivät erityisesti poikkea suomenkielisistä. Suurin ero on preposition esiintymisen puolessa tapauksista, mikä selittyy puhtaasti kielten välisellä erolla.

Kunkin ranskankielisen idiomien perään on suluissa lisätty ilmauksen merkitys suomeksi. Yksinkertaisiin lainausmerkkeihin olen lisännyt luontevan tavan ilmaista sama asia suomeksi. Mikäli idiomille on suomessa idiomivastine, kuten puolessa tapauksista, olen maininnut sen. Toisinaan suomessa ja ranskassa on tismalleen sama idiomien (esim. *être armé jusqu'aux dents* ~ olla hampaita myöten aseistautunut), kun taas toisinaan molemmissa kielissä on täysin toisistaan poikkeavat mutta samaa merkitsevät idiomit (*prendre le mors aux dents* ~ ryhtyä suurella palolla jhk). Osassa tapauksista olen käyttänyt parafrasaa (*se faire les dents sur qc* ~ harjoitella). Ainakin joissain tällaisissa tapauksissa myös suomesta olisi ollut mahdollista löytää jokin merkitykseltään ranskankielistä vastaava idiomien. Nimenomaan idiomien löytäminen suomennokseksi ei kuitenkaan ole tarkoitus itsessään, vaan pääasia on, että idiomien merkitys käy taulukosta ilmi. Mikäli ranskankielisen ilmauksen kirjaimellinen käännös ei ole sama kuin lainausmerkeissä esitetty luonteva/idiomaattinen suomennos, on sulkeissa annettu myös kirjaimellinen käännös kursiivein kirjoitettuna.

Suomennokset perustuvat omiin suomen taitoihini äidinkielenä puhujana. Joidenkin ranskankielisten idiomien merkitys oli minulle pitkään ranskaa opiskelleena entuudestaan tuttu. Tuntemattomien ranskankielisten idiomien merkitykset puolestaan selvitin *Dictionnaire des expressions et locutions* -sanakirjasta (Rey & Chantreau 2003) sekä verkkosanakirjasta *Linternaute*. Sama koskee myös seuraavaksi esiteltävien substantiivi-idiomien suomennoksia.

Neljännessä taulukossa esitetään ranskankieliset substantiivi-idiomit. Niitä esiintyy aineistossa ainoastaan neljä, eli selvästi vähemmän kuin verbi-idioimeja (16). Toisaalta niitä on lähes saman verran kuin suomenkielisen aineiston substantiivi-idioimeja (5). Erona suomen vastaavaan aineistoon on kuitenkin se, että siinä missä suomenkielisistä substantiivi-idioimeista yhtä lukuun ottamatta kaikki toimivat lauseessa adverbialina, on alla olevassa taulukossa ainoastaan yksi selvä adverbialia (*en dent de scie*).

SUBSTANTIIVI-IDIOMIT (RA)

Idiomi	Rakenne	Yhteismäärä aineistossa
en dent de scie (’epäsäännöllisesti’; kirj. <i>sahanhampain</i>)	prep. + subst.	4
grincements de dents (’hampaiden kiristely’; kirj. mm. <i>hampaiden narina</i>)	subst. + subst.	3
dent contre dent (kirj. <i>hammas vasten hammasta</i> . Huom.: ei vars. idiomi; taustalla <i>dent pour dent</i> =’hammas hampaasta’)	subst. + prep. + subst.	1
un coup de dent (’iva’/’mustamaalaus’; kirj. <i>hammasisku</i>)	subst. + subst.	1
Yht.		9

Taulukko 4. Aineiston ranskankieliset substantiivi-idiomit

Dent contre dent on aineistossa epäselvä tapaus. Sen suora käänös on ’hammas vasten/kohti hammasta’, mikä on itsessään jokseenkin järkevä ilmaus. Alkuperäiskontekstin tarkastelu kuitenkin osoittaa, että sillä on tavoiteltu merkitystä ’joku joutuu maksamaan teoistaan’. Tällöin onkin oletettavasti tarkoitettu idiomiä *dent pour dent*, joka vastaa suomen idiomiä *hammas hampaasta*. Idiomaattista merkitystä lienee joka tapauksessa haettu, joten luokittelen sen tässä tapauksessa idiomiksi.

4.1.3. Huomiot rakenteellisesta analyysistä

Aineistossa esiintyvien idiomien rakenteiden tarkastelu ei nostanut esiin suuria eroja tai muita yllättäviä seikkoja. Substantiivi on hyvin yleinen ja todennäköinen sanaluokka idiomeissa, ja tässä tapauksessa vaatimus *hammas/dent*-sanon esiintymisestä teki siitä välttämättömän osan jokaista ilmausta. Verbillisissä idiomeissa puolestaan esiintyy luonnollisesti verbi. Toisin kuin ranskassa, suomessa on olemassa vain vähän prepositioita, sillä niiden tilalla käytetään useimmiten erilaisia sijapäätteitä. Näin ollen se, että aineiston suomenkielisissä idiomeissa esiintyy vain yksi prepositio ja ranskankielisissä kymmenen oli täysin odotettavissa.

Adjektiivien määrä on molempien kielten aineistoissa vähäinen. Sekä suomessa että ranskassa esiintyy kerran adjektiiviattribuutti (*pitkin hampain; croquer à pleine dent*), kun taas varsinaisia adjektiiveja on ainoastaan ranskankielisissä idiomeissa. Adjektiiveja on kolme, ja ne esiintyvät idiomeissa *avoir la dent dure, avoir les dents longues* sekä *être armé jusqu'aux dents*. Näistä kolmesta tapauksesta kaksi ensimmäistä adjektiivia, 'kova' ja 'pitkä', viittaavat hampaisiin, kun taas viimeisen ilmauksen 'aseistautunut' ihmiseen. Näiden havaintojen perusteella herääkin kysymys siitä, ovatko adjektiiviattribuutit ylipäättään yleisiä hammas-idiomien yhteydessä ja millaisten idiomien yhteydessä ne yleensä esiintyvät.

Korhonen (2018: 54–55) toteaa fraseologisten sanaparien olevan yleisiä monissa kielissä. Aineiston idiomeissa sanapareja esiintyy kaksi, ja ne ovat molemmat suomenkielisiä: *pitkin hampain* sekä (*ilman*) *sarvia ja hampaita*. Näistä ensimmäinen on asyndeettinen, sillä siinä ei ole yhdistävää konjunktiota sanojen välillä (emt.). Vastaavasti jälkimmäinen esimerkki edustaa syndeettistä sanaparia (emt.). Sanaparin molemmat osat kuuluvat tyypillisesti samaan sanaluokkaan (emt.), kuten aineiston tapauksissa.

4.2. Semanttinen analyysi

Rakenteellisten piirteiden lisäksi idiomeja voidaan tarkastella semantiikan näkökulmasta. Sekä suomen- että ranskankielisten idiomien tarkastelu osoittaa, että aineistossa esiintyy vain vähän tai ei lainkaan positiivissävytteisiä hammas-idiomeja. Kummassakin kielessä idiomit ilmaisevat lähinnä vihan, katkeruuden ja pelon kaltaisia negatiivisiksi mielletäviä tunteita.

Aineiston kaikki idiomit on sijoitettu kielikohtaisiin taulukoihin 5 ja 6 sen perusteella, onko ilmaus sävyltään negatiivinen, neutraali vai positiivinen. Mikäli idiomilla on useita variantteja, on taulukossa mainittu niistä vain yksi. Osassa tapauksista ei ollut lainkaan itsestään selvää, mihin kolmesta sävykategoriasta idiomi tulisi sijoittaa. Molempia taulukkoja seuraakin lyhyt kielikohtainen pohdinta, jossa mainitaan tällaiset tapaukset.

negatiivinen	neutraali	positiivinen
hampaat irvessä	ajan hammas	
hampaat kalisevat	ilman sarvia ja hampaita	

joutua jkn hampaisiin	kynsin hampain	
kiristellä hampaita	(makean)hammasta kolottaa	
pitkin hampain		
purra hammasta		
saada hampaat kirskumaan		
sihahtaa hampaiden lomasta		

Taulukko 5. Suomenkielisten idiomien semanttinen analyysi

Suomalaisista idiomeista valtaosa (8/12) on sävyltään selvästi negatiivisiksi luokiteltavia. Ne ilmentävät pelkoa/pelottelua (*joutua jkn hampaisiin*; *hampaat kalisevat*), vihaa/ärtymystä (*purra hammasta*; *kiristellä hampaita*; *saada hampaat kirskumaan*; *sihahtaa hampaiden lomasta*; *hampaat irvessä*) sekä vastentahtoisuutta (*pitkin hampain*). Myös *kynsin hampain* saattaa, kontekstista riippuen, olla tulkittavissa negatiivissävytteiseksi. Toisaalta idiomi itsessään osoittaa ennen kaikkea päättäväisyyttä, mikä taas voidaan nähdä myös positiivisena ominaisuutena. Samoin suoruuteen viittaava *ilman sarvia ja hampaita* voi näkökulmasta riippuen olla hyvä tai huono piirre. Molemmat ilmaukset onkin näin ollen luokiteltu neutraaleiksi. *Ajan hammas* viittaa usein jonkinlaiseen kulumiseen, mikä ei kuitenkaan ole riittävä peruste pitää ilmausta ainoastaan negatiivisena. Neutraaliksi luokitelluista idiomeista neljäs on *(makean) hammasta kolottaa*. Vaikka osa neutraaleiksi katsotuista idiomeista voi toisinaan olla positiivisessa mielessä käytetty, ei aineiston suomenkielisten idiomien joukossa ollut yhtäkään sellaista ilmausta, joka olisi automaattisesti positiiviseksi mielletävä ja siten kolmanteen kategoriaan sopiva.

Ensimmäinen ero, joka voidaan havaita verrattessa taulukoita 5 ja 6 toisiinsa on se, että ranskankielisten idiomien joukosta löytyy yksi selvästi positiiviseksi luokiteltava (*croquer à pleine dent* 'nauttia täysin rinnoin'). Muutoin semanttisen analyysin tulokset vastaavat hyvin pitkälti toisiaan molemmissa kielissä.

negatiivinen	neutraali	positiivinen
avoir la dent dure	avoir les dents longues	croquer à pleine dent

avoir le couteau entre les dents	en dent de scie	
avoir une dent contre qqn	être armé jusqu'aux dents	
casser les dents	murmurer entre ses dents	
claquer des dents	prendre le mors aux dents	
dent contre dent	se faire les dents sur qc	
grincer les dents grincements de dents		
montrer les dents		
péter les dents		
rien à se mettre sous la dent		
serrer les dents		
un coup de dent		

Taulukko 6. Ranskankielisten idiomien semanttinen analyysi

Suomen tavoin myös ranskassa negatiivissävyiset idiomit ilmentävät pelkoa/pelottelua (*avoir la dent dure; montrer les dents*), vihaa/ärtymystä (*grincer les dents/grincements de dents; péter les dents*), kaunaa (*dent contre dent*), kostonhimoa (*avoir le couteau entre les dents*), pahantahtoisuutta (*avoir une dent contre qqn*), mustamaalausta (*un coup de dent*) ja vastentahtoisuutta (*serrer les dents*). Muita negatiivissävytteisiä aiheita, joita aineiston idiomit edustavat, ovat epäonnistuminen (*casser les dents*) ja köyhyys/nälän näkeminen (*rien/qc à se mettre sous la dent*). Hampaiden kalinalla (*claquer des dents*) voidaan viitata sekä palelemiseen että pelkoon.

Neutraaleiksi luokitellut idiomit puolestaan ilmaisevat epäsäännöllistä toimintaa (*en dent de scie*), harjoittelua (*se faire les dents sur qc*), epäselvää puhetta (*murmurer entre ses dents*) sekä kunnianhimoa (*avoir les dents longues*). Kunnianhimo nähdään yleisesti hyvänä ominaisuutena, mutta tulkinta riippuu kontekstista. Piirre ei siis ole automaattisesti positiivinen, joten se on sijoitettu kategoriaan neutraali. Toinen ilmaus, jonka kohdalla kategorian valitseminen ei ollut itsestään selvää, on *prendre le mors aux dents*. Yksi mahdollinen käännös sille on taulukossa 3 käyttämäni *ryhtyä suurella palolla jhk*. Toisaalta innokkuus, johon idiomi viittaa, voi toisinaan mennä jopa hätäilyn puolelle (*Linternaute*). Päädyin siis luokittelemaan ilmauksen neutraaliksi. Idiomi *être armé jus-*

qu'aux dents vastaa niin muodon kuin merkityksenkin osalta suomen idiomeja *olla hampaita myöten aseistautunut*.

Kuten esimerkiksi idiomin funktioita käsittelevässä osiossa (2.5.) todetaan, on negatiivista toimintaa ja tuntemuksia kuvastavia ilmauksia yleisesti enemmän kuin positiivisia tekoja ja tuntemuksia kuvaavia (Habler & Hümmer 2005: 107). Myös tämä tutkimus tukee ajatusta. Ihmisellä lienee luontainen tarve ilmaista ja sanoittaa enemmän ikäviä kuin miellyttäviä asioita. Esimerkiksi pelkoa voidaan ilmaista monilla erilaisilla ilmauksilla, joista vain osa on hammas-idiomeja. Sen sijaan pelon vastakohtaa, turvallisuuden tunnetta, ilmentäviä idiomeja on vaikeampi keksiä. Aineiston idiomeissa ei ole sävyn tai aihepiirien osalta havaittavissa varsinaisia eroja suomen ja ranskan välillä. Ranskankielisessä aineistossa on enemmän eri idiomeja, minkä vuoksi ne myös edustavat useampaa erilaista tunnetilaa/toimintaa kuin suomenkielisen aineiston idiomit.

4.3. Variaatio aineiston idiomeissa

Aineiston idiomeissa ei, etenkin ranskan osalta, esiinny erityisen paljon variaatiota. Erilaisia variantteja on havaittavissa lähinnä suomenkielisiä verbi-idiomeja kuvaavassa taulukossa 1. Olen luokitellut yhdeksi idiomiksi seuraavat neljä versiota: *joutua/jättää jkn hampaisiin*, *olla jkn hampaissa* sekä *päästä jkn hampaista*. Kaikissa tapauksissa on eri verbi, ja samaan listaan voisi lisätä myös lukemattoman määrän muita verbejä, kuten *päättyä*, *jäädä* ja *pelastautua*. Verbin lisäksi *hammas*-sanana taivutus vaihtelee, sillä eri verbit edellyttävät eri rektioita. Edellä mainitusta huolimatta idiomin perimmäinen merkitys säilyy aina samana. Verbin mukana vaihtuvat ainoastaan ilmauksen sävy ja se, kenen näkökulmasta tilannetta katsotaan (esimerkiksi *pelastaa/pelastautua jkn hampaista* tai *jäädä/jättää jkn hampaisiin*). Näin ollen voidaan katsoa kyseessä olevan sama idiomi kaikissa tapauksissa.

Sen sijaan kahdessa muussa ensimmäisen taulukon tapauksessa merkitys säilyy nähdäkseni täysin samana. Se, *pureeko joku hammasta* vai *pureeko hän niitä (yhteen)* ei muuta lauseen merkitystä, kun ilmausta käytetään kuvaannollisessa merkityksessä. Idiomin variaatiosta on kyse myös silloin, kun toistensa synonyymeja *lomasta* ja *välistä* käytetään vaihdellen.

Taulukossa 2 huomiota herättää ilmaus *irvi hampaassa*. Se esiintyy aineistossa kerran, ja alkuperäiskontekstissaan sitä edeltää verbi *taistella*. Ilmaus on sikäli poikkeuksellisesti

nen, ettei monikaan äidinkielen suomenpuhujia liene törmännyt siihen tuossa muodossa. Sen sijaan varsin yleinen on aineistossa kolmesti toistuva *hampaat irvessä*. Sitä, onko *irvi hampaassa* tahaton tai tahallinen muunnelma kyseisestä tutusta idiomista, on mahdoton tietää. Selvää lienee kuitenkin se, että kirjoittajalla on ollut pyrkimys käyttää ilmausta idiomien tapaan. Koska kirjaimellista tulkintaa ei ole syytä eikä voidakaan tehdä, on ilmaus luokiteltavissa idiomiksi ja rinnastettavissa nimenomaan idiomiin *hampaat irvessä*. Se onkin suomenkielisissä substantiivi-idiomeissa ainoa tapaus, jossa esiintyy minkäänlaista variaatiota.

Kuten edellä käy ilmi, suomenkielisten verbi-idiomien joukossa on muutama tapaus, joissa esiintyy sanastollista vaihtelua. Aineiston ranskankielisissä verbi-idiomeissa sitä ei esiinny, mutta sen sijaan yhdessä tapauksessa variaatiota on artikkelin osalta. Idiomeissa *grincer les/des dents* voidaan käyttää niin määräistä artikkelia *les* kuin epämääräistä artikkelia *des*. Näistä ensimmäinen esiintyy aineistossa ainoastaan kerran, kun taas toinen viisi kertaa. Myös joidenkin muiden idiomien kohdalla olisi mahdollista käyttää muutakin kuin vain aineistossa esiintyvää artikkelia tai pronominia. Esimerkiksi ilmauksessa *montrer ses dents* possessiivipronominin *ses* tilalla näkee toisinaan käytettävän määräistä artikkelia *les*. Mitä ranskankielisiin substantiivi-idiomeihin tulee, ei yhdelläkään neljästä idiomista ole varianttia. Ne esiintyvät aineistossa ainoastaan taulukossa mainituissa muodoissa.

Myös ranskankielisissä ilmauksissa on siis variaatiota, vaikkakin hyvin vähän. Kyse ei kuitenkaan ole sisältösanojen muuttumisesta, joten vaihtelu ei muuta idiomien sisältöä tai edes näkökulmaa, toisin kuin esimerkiksi ilmauksissa *joutua jkn hampaisiin* ja *jättää jkn hampaisiin*.

4.4. Ei-kuvaannolliset tapaukset

Sekä suomessa että ranskassa vain noin viidesosassa aineiston tapauksista *hammas/dent*-sana esiintyy osana idiomia. Näin ollen selvä enemmistö virkkeistä on sellaisia, joissa sana ei ole kuvaannollisessa merkityksessä. Millaisissa yhteyksissä sanoja *hammas* ja *dent* sitten useimmiten käytetään?

Aineiston tarkastelu osoittaa sen, mikä jo etukäteen oli arvattavissa. Valtaosassa virkkeistä aiheena on nimittäin lääketiede tai jokin muu terveyteen liittyvä. Niin suomessa kuin ranskassakin kyse on useimmiten hampaidenhoito-ohjeista, sairaskertomuksesta tai

ihmisten keskustelupalstoilla, toisinaan myös blogeissa, jakamista kokemuksista liittyen esimerkiksi hampaanpoistoon tai ulkomailla tapahtuneeseen hammaslääkärikäyntiin. Hampaidenhoito-ohjeet ovat useimmiten ihmisten, mutta joskus myös eläinten, hampaita koskevia. Yleisiä ovat myös esimerkiksi hammasharja- ja hampaidenvalkaisumainokset.

Molemmissa kielissä esiintyy ainoastaan pariin kertaan virke, jossa viitataan moottori-koneen tai muun työväliseen hampaisiin. *Hammas/dent*-sanon käyttö siinä yhteydessä ei siis tässä aineistossa ole erityisen yleistä kummassakaan kielessä. Ranskankielisessä korpuksessa esiintyy kerran italiasta peräisin oleva ilmaus *al dente*. Kyseinen tapaus on ainoa, jossa *dent* on osana jotakin toista sanaa, vaikka tällaisia tapauksia olisi ollut mahdollista esiintyä huomattavasti enemmänkin, sillä esimerkiksi *al dente* on tunnettu ilmaus monissa kielissä. Voitaneen siis todeta, ettei tapauksella ole juuri vaikutusta lopputuloksen kannalta.

Ainut varsinainen ero, joka ilmenee verrattaessa suomen- ja ranskankielisiä aineistoja toisiinsa idiomia sisältämättömien virkkeiden osalta, liittyy yhteen *dent*-sanon merkityksistä. Sanaa käytetään ranskassa ilmeisen usein viitattaessa vuoren huippuun. Ranskasta myös löytyy useita vuoria, joiden nimeen sana sisältyy. Aineistossa tällaisia ovat *Dent de Morcles*, *Dent du Midi* sekä *Dent d'Orlu*, joista kukin mainitaan kerran omissa virkkeissään. Lisäksi ranskankielisessä aineistossa esiintyy kerran *Temple de la Dent*, joka on suomeksi Pyhän hampaan temppelinä tunnettu srilankalainen temppeli. Mitä erisnimiin muutoin tulee, mainitaan suomen aineistossa kerran suomalainen Hammaslahden kylä, kun taas ranskan aineistossa *Dent* esiintyy kahteen kertaan sukunimenä.

Yhteenvetona todettakoon, että ranskan ja suomen välillä ei ole suurta eroa siinä, millaisissa yhteyksissä sanat *hammas* ja *dent* esiintyvät silloin, kun ne eivät ole osana idiomia. Ranskassa sana toimii nimityksenä vuoren huipulle, ja siitä johtuen se esiintyy osana joidenkin vuorten nimiä, mutta muutoin varsinaisia eroja ei ole. Aineistojen tarkastelu ei myöskään tuottanut yllätyksiä, sillä ylivoimainen enemmistö virkkeistä on otteita hammaslääkärin käsittelevistä teksteistä.

5. Tulokset ja pohdinta

5.1. Tutkimustulokset

Tutkimuksen tarkoituksena oli selvittää:

- 1.) Minkälaisia *hammas/dent*-sanon sisältäviä idiomeja suomessa ja ranskassa esiintyy
- 2.) Onko hammas-idiomeissa tai niiden esiintymismäärissä havaittavissa kieltenvälisiä eroja?

Vastaus ensimmäiseen kysymykseen ilmenee edellä esitetyistä taulukoista, joihin ovat koottuina kaikki aineistoissa esiintyvät hammas-idiomit määrineen. Neljä yleisintä idiomiä suomenkielisessä aineistossa ovat seuraavat: *olla jkn hampaissa* (8) ja *purra hammasta* (5) erilaisine variaatioineen sekä *kynsin hampain* (5) ja *pitkin hampain* (5). Aineistossa esiintyy jonkin verran enemmän suomenkielisiä verbi- kuin substantiivieroja, mutta ero ei ole erityisen suuri.

Yksi suomenkieliseen aineistoon liittyvä huomio on se, että idiomi *makean hammasta kolottaa* esiintyy siinä ainoastaan kerran, eli yhtä usein kuin esimerkiksi *ilman sarvia ja hampaita*. Syynä tähän on todennäköisesti sattuma. Kuten osiosta 3.5 käy ilmi, on kyse todellisuudessa varsin yleisestä ja hyvinkin produktiivisesta idiomista, sillä sanon *makean* pystyy korvaamaan hyvin monella eri vaihtoehdolla vain mielikuvitusta rajana käyttäen. Verrattaessa toisiinsa ilmauksia *x hammasta kolottaa* ja *ilman sarvia ja hampaita* voisi siis olettaa, että ensin mainittu on todellisuudessa käytetympi eikä aineiston muodostama kuva kyseisten idiomien osalta vastaa todellisuutta.

Ranskankielisessä aineistossa neljä yleisintä idiomiä ovat *grincer des/les dents* (6), *croquer à pleine dent* (5), *se faire les dents sur qc* (5) ja *en dent de scie* (4). Joukosta ei ranskakaan kohdalla erotu yhtä selkeästi muita useammin esiintyvää. Toisaalta substantiivi-idiomien taulukkoon sijoitettu *grincements de dents* esiintyy kolmeen kertaan, ja se koostuu samoista aineksista kuin yleisin verbi-idiomi *grincer des/les dents*. Niiden yhteen laskettu määrä aineistossa on yhdeksän, mikä onkin jo varsin suuri osuus siihen nähden, että edellä mainittuja idiomeja lukuun ottamatta ranskankielisen aineiston idiomit esiintyvät siinä kukin vain kerran tai kaksi. Miksi idiomit *grincer des/les dents* ja *grincements de dents* ylipäätään on sijoitettu eri taulukoihin, kun loppujen lopuksi kyse on merkityssisällön ja kielellisten aineiden osalta saman idiomien verbi- ja substan-

tiiviversion? Idiomien jakaminen kahteen eri taulukkoon perustuu tässä nimenomaan siihen, onko kyseessä verbi- vai substantiivi-idioni, ja molemmat esiintyvät aineistossa useaan otteeseen, joten päädyin kaikesta huolimatta myös mainitsemaan ne erikseen kummassakin taulukossa. Myös toisenlainen ratkaisu olisi kuitenkin ollut perusteltu.

Mitä idomien sävyyn tulee, ovat molempien kielten aineistoissa esiintyvät idiomit selvästi enemmän negatiivis- kuin positiivissävytteisiä. Kuten semanttisen analyysin (4.2.) yhteydessä esitetyistä taulukoista käy ilmi, olen luokitellut suomenkielisistä 12 idiomista peräti kahdeksan negatiiviseksi, ja loput neljä neutraaleiksi. Ranskan idiomeista puolestaan olen luokitellut yhden sävyllään positiiviseksi, seitsemän neutraaleiksi ja loput 12 negatiiviseksi.

Toinen tutkimuskysymys on, kummissa kielessä hammas-idiomeja esiintyy enemmän. Alla oleviin taulukoihin on koottu yhteenvetona alaluvussa 4.1. ilmenevät idiomien määrät. Taulukossa 7 eriteltyinä ovat erilaisten verbi- ja substantiivi-idiomien määrät kummassakin kielessä, kun taas taulukosta 8 nähdään, mitkä ovat idiomien kokonaismäärät aineistoissa. Toisin sanoen jälkimmäisessä taulukossa on otettu huomioon kaikki aineistoissa esiintyvät tapaukset, joissa sana *hammas/dent* on osa idiomiä eikä kirjaimellisesti käytettynä.

ERILAISTEN IDIOMIEN MÄÄRÄT AINEISTOISSA		
	suomi	ranska
verbi-idiomit	7	16
substantiivi-idiomit	5	4
Yht.	12	20

Taulukko 7. Erilaisten idiomien määrät aineistoissa

IDIOMIEN KOKONAISMÄÄRÄT AINEISTOISSA		
	suomi	ranska
verbi-idiomit	20	35
substantiivi-idiomit	17	9

Yht.	37/198 (= 19 %)	44/198 (= 22 %)
-------------	-----------------	-----------------

Taulukko 8. Idiomien kokonaismäärät aineistoissa

Taulukoista ilmenevien määrien perusteella voidaan todeta ensinnäkin se, että idiomien kokonaismäärissä ei ole juuri lainkaan eroa suomen ja ranskan välillä. Tutkimushypoteesi piti siis siltä osin paikkansa. Molempia on noin viidesosa aineistojen virkkeistä.

Sen sijaan toinen tutkimushypoteesi, jonka mukaan verbi- ja substantiivi-idiomeja esiintyisi molemmissa kielissä suunnilleen saman verran, ei osoittautunut oikeaksi. Yksittäisten idiomien määriä tarkasteltaessa on nimittäin havaittavissa eroja. Siinä missä suomenkielisestä aineistosta voidaan poimia 12 eri idiomia, on vastaava luku ranskankielisen aineiston osalta 20. Kuten taulukoista 7 käy ilmi, esiintyy ranskankielisessä aineistossa erityisen paljon nimenomaan verbi-idiomeja. Molemmissa kielissä on niin yksittäisten ilmausten kuin idiomien kokonaismäärien tarkastelun valossa enemmän verbi- kuin substantiivi-idiomeja, mutta ranskassa ero on selvempi. On kuitenkin huomioitava se, että tutkimuksessa on kyse vain kirjoitetussa muodossa olevasta tekstistä sekä verrattain pienestä otannasta (198 satunnaista internetissä esiintyvää virkettä). Näin ollen varsinaisia yleistyksiä ei voida tehdä, vaan tuloksia on syytä pitää lähinnä suuntaa antavina.

5.2. Kriittinen näkökulma

Tutkimuksen heikkouksia sivutaan jo aineiston keruuta käsittelevässä osiossa 3.3. Heikkouksiksi voidaan katsoa valitun otoksen pieni koko ja sen välittämä rajallinen kuva *hammas-* ja *dent-*lekseemien käytöstä. Mutta toisaalta, milloin edes voidaan puhua aineistosta, joka olisi riittävän laaja ja huomioisi tarpeeksi monipuolisesti kielellisen variaation? Tutkimustuloksiin on aina syytä suhtautua varauksella, ja harvoin niitä pidetäänkään ehdottomana totuutena. Vaikka tutkimusaineisto koostuisi molemmissa kielissä vajaan 200:n sijasta 2000 virkkeestä, olisi laajuus edelleen suhteellista. Toki tutkimuksen luotettavuus ja tulosten yleistettävyyks kasvavat aina otannan mukana. Tutkimuksilla on kuitenkin omat lähtökohtansa, ja tässä tapauksessa lähes 400 virkkeen, joista puolet on vieraskielisiä, yksittäinen tarkastelu ja osassa tapauksista niiden alkuperäis-kontekstin selvittäminen oli pro gradu -tutkielman laajuudessa toteutettavissa oleva analyysimäärä.

Tutkimukseen tulee herkästi subjektiivisuutta silloin, kun tutkija joutuu tekemään omia ratkaisuja esimerkiksi luokittelun suhteen. Tässä tapauksessa yksi valinta oli muun mu-

assa ilmausten *joutua jkn hampaisiin*, *jättää jkn hampaisiin*, *olla jkn hampaissa* ja *päästä jkn hampaista* luokittelu yhdeksi ja samaksi idiomiksi, jolloin siitä tuli suomenkielisen aineiston yleisin idiomi. Varianttien lisäksi ratkaisuja oli tehtävä myös joidenkin epäselvien tapausten kohdalla.

Analyysia edeltäneessä aineiston keruuvaiheessa yllätyksenä tuli ranskankielisten osu-
mien määrä. Siinä missä suomenkielisiä *hammas*-sanana sisältäviä virkkeitä on käyttölii-
tymässä tällä hetkellä 479, joista olisi saanut koostettua huomattavasti nykyistä laajem-
mankin aineiston, esiintyy vastaavassa ranskankielisessä aineistossa ainoastaan 198
hammas-idiomia. Etukäteen päätetty määrä 200 ei siis kuitenkaan jäänyt kuin kahden
virkkeen verran vajaaksi, mikä ei lopputuloksen kannalta ole merkittävä määrä.

6. Yhteenveto

Tämän tutkielman aiheena ovat suomen- ja ranskankieliset idiomit, joihin sisältyy sana *hammas* tai sen ranskankielinen vastine *dent*. Tutkielman teoreettisessa viitekehyksessä tarkastellaan idiomien käsitettä useasta näkökulmasta. Keskeisiä näkökulmia ovat idiomien tyypilliset piirteet sekä idiomien tunnistaminen ja funktiot kielessä.

Aineistoa ja metodeita käsittelevässä luvussa tarkastellaan tämän tutkimuksen aineiston sekä sen keruu- ja luokittelumenetelmien lisäksi internet-aineistoa yleisesti. Lisäksi luvussa kuvaillaan tutkielman aihetta eli suomen ja ranskan hammas-idiomeja. Metodiosio sisältää katsauksen kielten vertailun teoriaan.

Neljäs luku koostuu aineiston analyysistä, ja siinä esitellään ja kuvataan kaikki aineistossa esiintyvät idiomit kieli- ja rakennekohtaisiin taulukoihin koottuina. Idiomien rakenteen, semantiikan ja variaation lisäksi luvussa käsitellään myös ei-kuvaannollisia tapauksia, joita suurin osa aineiston virkkeistä edustaa. Viides luku sisältää tulosten esittelyn sekä pohdintaa tutkimuksen heikkouksista.

Tutkimuksen tarkoituksena oli selvittää:

- 1.) Minkälaisia *hammas/dent*-sanan sisältäviä idiomeja suomessa ja ranskassa esiintyy internet-aineistossa?
- 2.) Onko hammas-idiomeissa tai niiden esiintymismäärissä havaittavissa kieltenvälisiä eroja?

Sekä suomessa että ranskassa noin viidesosassa aineiston virkkeistä *hammas/dent*-sanaa käytetään kuvaannollisessa merkityksessä, osana idiomia. Sitä, minkälainen tilanne muilla sanoilla on tai missä määrin tämän tutkimuksen tuloksista voidaan tehdä yleistyksiä muiden tekstiaineistojen osalta, on mahdoton sanoa. Kuitenkin esimerkiksi Nenosen (2002: 115) tutkimuksen perusteella monet muut ruumiinosanimet, etenkin *pää*, *silmä* ja *käsi*, esiintyvät huomattavasti useammin ja useammassa eri idiomissa kuin *hammas*.

Mitä tulee verbillisten ja verbittömien idiomien suhteeseen, suomen aineistossa eri verbi-idiomeja on vain aavistuksen enemmän kuin substantiivi-idiomeja. Samoin kaikkien suomenkielisten idiomien yhteenlasketut määrät ovat melko samoissa lukemissa verbi-

idiomien määrän ollessa 20 ja substantiivi-idiomien 17. Varsinaisesta erosta ei siis voida puhua. Sen sijaan ranskankielisessä aineistossa eri verbi-idiomeja on jopa nelinkertainen määrä substantiivi-idiomeihin nähden. Samoin kaikkien ranskan verbi-idiomien yhteenlaskettu määrä on nelinkertainen substantiivi-idiomien kokonaismäärään verrattuna (35 vs. 9). Tutkimus herättää siten kysymyksen, onko ranskassa enemmän verbillisiä kuin verbittömiä hammas-idiomeja.

Ensimmäiseen tutkimuskysymykseen, eli siihen minkälaisia *hammas/dent*-sanan sisältäviä idiomeja suomessa ja ranskassa esiintyy, voidaan vastata erittelemällä idiomeja niiden rakenteen perusteella (4.1.) sekä mainitsemalla kummassakin kielessä aineiston perusteella yleisimmät idiomit. Lisäksi aineiston idiomeja voidaan lähestyä semantiikan ja variaation näkökulmista, kuten alaluvuissa 4.2. ja 4.3. on tehty.

Jotta saataisiin luotettavampaa tietoa sanojen *hammas* ja *dent* esiintymisestä osana idiomia, tulisi aineiston olla laajempi ja monipuolisempi. Internet-tekstien lisäksi tarkasteluun olisi hyvä ottaa myös painettuja tekstejä sekä puhuttua kieltä ja eri-ikäisiä puhujia. Tulosten yleistettävyyttä edellyttäisi myös tarkkojen todennäköisyyksien laskemista. Tällöin aineistoa voitaisiin tarkastella kunnolla myös määrällisestä näkökulmasta, joka tässä tutkimuksessa jäi laadullisen tarkastelun varjoon. Esiintymistiheyteen liittyvät tulokset ovat siis tässä tutkielmassa ainoastaan suuntaa antavia, ja niihin on näin ollen suhtauduttava varauksella.

Idiomien käyttö eri tilanteissa olisi mielenkiintoinen, vaikkakin myös paljon vaivannäköä vaativa tutkimuskohde. Siinä, mitä idiomeja ihmiset suosivat puheessa ja kirjoituksessa, saattaa hyvinkin olla eroavaisuuksia. Niiden selvittäminen ja tulosten yleistettävyyttä edellyttäisivät kuitenkin varsin laajaa ja monipuolista otantaa.

Toinen kiinnostava, idiomeihin liittyvä aihe ovat niiden käyttöön vaikuttavat syyt. Aihetta sivutaan lyhyesti tämän tutkielman alaluvussa 2.5., jossa tarkastellaan idiomien funktioita. Useissa tapauksissa idiomien välittämän asian voisi ilmaista myös toisin. Missä määrin kielenkäyttäjät tekevät tietoisin valinnan esimerkiksi kertoessaan tehneensä jotakin *hammasta purren* eikä *vastentahtoisesti*? Taustalla saattaa olla vaikkapa halu käyttää elävää ja humoristista kieltä tai vaihtoehtoisesti muiden vaikutuksesta omaksuttu tapa.

Idiomeja voi tutkimusaiheena lähestyä hyvin monesta eri näkökulmasta. Niihin liittyvää teoriaa on tarjolla runsaasti, kuten tämänkin tutkimuksen laajasta teoriaosuudesta voidaan päätellä. Aineksia laajempaankin käsittelyyn olisi ollut tarjolla, mutta kuten tutkimuksissa monesti, myös tässä tapauksessa raja oli vedettävä johonkin. Sen lisäksi, että idiomeja voi tarkastella jo pelkästään yhden kielen sisällä monin eri tavoin, tarjoavat muut kielet ja kielten välinen vertailu vielä huomattavan määrän uusia mahdollisuuksia ja näkökulmia. Tällöin käännosten vertailu ei suinkaan ole ainut vaihtoehto, kuten tämänkin tutkimus osoittaa.

Lähteet

Aineslähteet

Turku NLP Groupin käyttöliittymä. Viitattu 10.–11.4.2018. Saatavissa: [bionlp-
www.utu.fi/dep_search/](http://www.utu.fi/dep_search/)

Muut lähteet

Anhava, Jaakko 2002: Sanontojen elämää. *Kielikello 2/2002*. Viitattu 15.10.2017. Saatavissa: <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=1343>

Asheghi, Noushin Rezapour — Sharoff, Serge — Market, Katja 2016: Crowdsourcing for web genre annotation. *Language Resources and Evaluation 3/50*, 603–641. Viitattu 30.1.2018. Saatavissa: <https://dl.acm.org/citation.cfm?id=2990183>

Baker, Mona 1992: *In other words: A coursebook on translation*. Routledge, London. Saatavissa: <https://www.docdroid.net/9ai3/in-other-words-by-mona-baker.pdf#page=6>

Bennett, Paul 1997: Compositionnalité et figement des locutions: étude comparative du français et l'anglais. Teoksessa Fiala, Pierre — Lafon, Pierre — Piguët, Marie-France (toim.), 11–17. *La locution : entre lexicque, syntaxe et pragmatique: Identification en corpus, traitement, apprentissage*. Klincksieck, Paris.

Benson, Morton 1985: Collocations and idioms. Teoksessa Robert Ilson (toim.), 61–68. *Dictionaries, lexicography and language learning*. Pergamon Press, London.

Biber, Douglas — Conrad, Susan — Reppen, Randi 1998: *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press, Cambridge.

Cacciari, Cristina 2014: Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*. Università degli studi di Modena e Reggio Emilia, Modena.

Cacciari, Christina 1993: The place of idioms in a literal and metaphorical world. Teoksessa Cacciari, Christina & Tabossi, Patricia (toim.), 25–55. *Idioms: Processing, structure, and interpretation*. Lawrence Erlbaum Associates, New Jersey. Viitattu 15.3.2018. Saatavissa: https://www.researchgate.net/publication/238123412_The_place_of_idioms_in_a_literal_and_metaphorical_world

Cacciari, Cristina — Glucksberg, Sam 1991: Understanding Idiomatic Expressions: The Contribution of Word Meanings. Teoksessa G.B. Simpson (toim.), 217–240. *Understanding word and sentence*. Elsevier Science Publishers, Amsterdam.

Chesterman, Andrew 1998: *Contrastive Functional Analysis*. John Benjamins, Amsterdam; Philadelphia.

GitHub – Universal Dependencies/UD_French-GSD. Viitattu 30.1.2018. Saatavissa: https://github.com/UniversalDependencies/UD_French-GSD

- Granger, Sylviane — Paquet, Magali 2008: *Disentangling the phraseological web*. Teoksessa Granger, Sylviane — Meunier, Fanny (toim.), 27–49. *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing Company, Amsterdam.
- Habler, Gerda — Hümmer, Christiane 2005: Figement et défigement polylexical: L'effet des modifications dans des locutions figées. *Linx*, 53 s. 103–119. Viitattu 10.1.2018. Saatavissa: <https://linx.revues.org/266>
- Heine, Bernd — Kuteva, Tania 2002: *World Lexicon of Grammaticalization*. Cambridge University Press, Cambridge.
- Heinonen, Tarja Riitta 2013a: *Idiomien leksikaalinen kuvaus kielenkäytön ja vaihtelun näkökulmasta*. Helsingin yliopisto, Helsinki.
- Heinonen, Tarja Riitta 2013b: Tsunami vesilasissa — leikittelyä sanonnoilla. *Kielikello 2/2013*. Viitattu 20.11.2018. Saatavissa: <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=2610>
- Järventausta, Marja 2013: *Kontrastiivinen tutkimus vertailevan kielentutkimuksen kentässä*. Teoksessa Kolehmainen, Leena — Miestamo, Matti — Nordlund, Taru (toim.), 96–134. *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seura, Helsinki.
- Kari, Erkki 1993: *Naulan kantaan: nykysuomen idiomisanakirja*. Otava, Helsinki.
- Kielitoimiston sanakirja. Viitattu: 10.4.2018. Saatavissa: <https://www.kielitoimistonsanakirja.fi/>
- Kilgarriff, Adam — Grefenstette, Gregory 2003: Introduction to the special issue on the web as corpus. *Computational Linguistics* 29(3), s. 333–347. Viitattu 30.1.2018. Saatavissa: <https://dl.acm.org/citation.cfm?id=964752>
- Korhonen, Jarmo 2018: *Fraseologia: Kiinteiden sanayhtymien tutkimus*. Finn Lectura, Helsinki.
- Kortelainen, Kristiina 2011: *Ei ole kaikki muumit laaksossa*. Tutkimus suomen kielen idiomikonstruktion produktiivisuudesta. Pro gradu -tutkielma. Turun yliopisto, Turku.
- Laippala, Veronika — Luotolahti, Juhani — Kyröläinen, Aki-Juhani — Salakoski, Tapio — Ginter, Filip 2017: Creating register sub-corpora for the Finnish Internet Parsebank. *Proceedings of the 21th Nordic Conference of Computational Linguistics (NODALIDA 2017)*. Viitattu 10.2.2018. Saatavissa: <http://aclweb.org/anthology/W17-0218>
- Lamiroy, Béatrice 2008: Les expressions figées: À la recherche d'une définition. Teoksessa Blumenthal, Peter — Mejri, Salah (toim.), 85–98. *Les séquences figées: Entre langue et discours*. Franz Steiner Verlag, Stuttgart. Viitattu 15.11.2017. Saatavissa: <http://www.ling.arts.kuleuven.be/franitalco/papers/Lamiroy2008.pdf>
- Lamiroy, Béatrice — Klein, Jean René 2005: Le problème central du figement est le semi-figement. *Linx*, 53, 135–154. Viitattu 15.11.2017. Saatavissa: <https://linx.revues.org/271>

- Lastunen, Henrietta 2016: *Être comme pain et beurre – comment se traduisent les idiomes français en finnois?* Turun yliopisto, Turku
- Lindquist, Hans — Levin, Magnus 2008: *Foot and mouth: The phrasal patterns of two frequent nouns*. Teoksessa Granger, Sylviane — Meunier, Fanny (toim.), 143–158. *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing Company, Amsterdam.
- Linternaute. Viitattu 20.10.2018. Saatavissa: <http://www.linternaute.com/expression/langue-francaise/350/se-casser-les-dents/>
- Luotolahti, Juhani — Kanerva, Jenna — Laippala, Veronika — Pyysalo, Sampo — Ginter, Filip 2015: Towards Universal Web Parsebanks. *Proceedings of the Third International Conference on Dependency Linguistics* (Depling 2015) s. 211–220. Viitattu 30.1.2018. Saatavissa: <http://www.aclweb.org/anthology/W15-2124>
- Mejri, Salih 2005: Figement absolu ou relatif: La notion de degré de figement. *Linx*, 53, 183–196. Viitattu 30.11.2017. Saatavissa: <https://journals.openedition.org/linx/283>
- Miestamo, Matti 2013: *Kielten vertailu kielityologisessa tutkimuksessa*. Teoksessa Kolehmainen, Leena — Miestamo, Matti — Nordlund, Taru (toim.), 27–55. *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seura, Helsinki.
- Moon, Rosamund 1998: *Fixed expressions and idioms in English. A corpus-based approach*. Clarendon Press, Oxford.
- Muikku-Werner, Pirkko 2009: Moninaisuus idiomeissa. *AFinLA-e Soveltavan kielitieteen tutkimuksia 1*, 77–90. Viitattu 1.11.2017. Saatavissa: <https://journal.fi/afinla/article/view/2616>
- Nenonen, Marja 2002: *Idiomit ja leksikko. Lausekeidiomien syntaktisia, semanttisia ja morfologisia piirteitä suomen kielessä*. Joensuun yliopiston humanistisia julkaisuja 29, Joensuu.
- Nykyajan kielenopas. Viitattu 30.1.2018. Saatavissa: <http://jkorpele.fi/kielenopas/all.html>
- Ojutkangas, Krista 2001: *Ruumiinosannimien kieliopillistuminen suomessa ja virossa*. Suomalaisen Kirjallisuuden Seura, Helsinki.
- Omazic, Marija 2008: *Processing of idioms and idiom modifications: A view from cognitive linguistics*. Teoksessa Granger, Sylviane — Meunier, Fanny (toim.), 67–79. *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing Company, Amsterdam.
- Orlikowski, Wanda — Yates, Joanne 1994: Genre repertoire: The structuring of communicative practices in organizations. *Administrative science quarterly*, 541–574.
- Palmer, Frank Robert 1981: *Semantics*. Toinen painos. Cambridge University Press, Cambridge.

- Piirainen, Elisabeth 2008: *Figurative phraseology and culture*. Teoksessa Granger, Sylviane — Meunier, Fanny (toim.), 207–228. *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing Company, Amsterdam.
- Rey, Alain — Chantreau, Sophie 2003: *Dictionnaire des expressions et locutions*. Toinen painos. Le Robert, Paris.
- Rey, Alain — Morvan, Danièle — Firmin, Gilles 2008: *Le Robert de poche 2009*. Uusittu painos. Sejer, Pariisi.
- Saarinen, Elina 2015: *"Entre chien et loup": Expressions figées en français et en finnois*. Turun yliopisto, Turku.
- Svensson, Maria Helena 2004: *Criteria de figement: L'identification des expressions figées en français contemporain*. Umeå University, Uumaja. Viitattu 12.11.2017. Saatavissa : <http://www.diva-portal.org/smash/get/diva2:143138/FULLTEXT01.pdf>
- Svensson, Maria Helena 2008: *A very complex criterion of fixedness: Non-compositionality*. Teoksessa Granger, Sylviane — Meunier, Fanny (toim.), 81–93. *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing Company, Amsterdam.
- Tervomaa, Raija 2002: *Päin mäntyä: Suomen kielen sanontoja*. Laatusana, Helsinki.
- Tieteen termipankki. Viitattu 8.5.2018. Saatavissa: <http://tieteentermipankki.fi/wiki/Nimitys:fraasi>
- Turku NLP Group -sivusto. Viitattu 8.1.2018. Saatavissa: <https://turkunlp.github.io/>
- Universal Dependencies -sivusto. Viitattu 30.12.2017. Saatavissa: <http://universaldependencies.org/>
- VISK = Hakulinen, Auli — Vilkuna, Maria — Korhonen, Riitta — Koivisto, Vesa — Heinonen, Tarja Riitta — Alho, Irja 2004: *Iso suomen kielioppi, verkkoversio*. Suomalaisen Kirjallisuuden Seura, Helsinki. Viitattu 1.3.2018. Saatavissa: <http://scripta.kotus.fi/visk>

1. Introduction

Le langage contient énormément d'idiomes et nous les utilisons souvent — que ce soit conscient ou inconscient. L'utilisation est consciente, par exemple, dans la situation où le locuteur vise à faire rire les auditeurs ou à adoucir l'expression. Les idiomes, ainsi que les autres types d'expressions figées, rendent la langue vivante et permettent la créativité. De plus, beaucoup d'idiomes sont culturellement spécifiques.

Le sujet de cette recherche est formé des idiomes finnois et français contenant le mot *dent* (en finnois *hammas*). La motivation pour ce sujet vient de l'intérêt pour les idiomes dont traite également notre mémoire de licence. Pour limiter le sujet, nous avons choisi un mot qui apparaît dans de nombreux idiomes. Les expressions somatiques, alors les idiomes incluant un nom d'une partie du corps (Korhonen 2018 : 131), sont courantes, constate Nenonen (2002 : 114). Pourtant, le mot *dent* n'est pas grammaticalisé de même manière que, par exemple, les mots *main* et *tête* (Ojutkangas 2001). La probabilité des cas ambigus est plus petite avec le mot *dent* qu'avec les mots grammaticalisés qui ont plusieurs sens. *Dent* fait normalement référence, soit à *dent humaine* ou *animale*, soit il fait partie d'un idiomme.

Le corpus est formé de 198 phrases finnoises et de 198 phrases françaises qui existent sur Internet et qui contiennent ce mot. La raison pour laquelle le nombre de phrases est 198 au lieu de 200 sera expliquée plus tard. Nous examinons chaque phrase pour trouver s'il s'agit d'un idiomme ou si le mot *dent* est utilisé dans un sens non-figuratif.

Après avoir examiné le corpus, nous appliquons le classement morphosyntaxique de Korhonen (2018) et divisons les idiomes finnois et français en deux groupes, selon leur forme : les idiomes contenant un verbe (idiomes verbaux) et les idiomes sans verbe (idiomes substantivés). Les questions de recherche sont alors les suivantes :

- 1.) Quels types d'idiome comporte le mot *dent* ?
- 2.) Y a-t-il des différences entre le finnois et le français concernant les idiomes et leur quantité dans le corpus ?

Pour répondre à ces questions, nous examinons, non seulement la quantité de différents idiomes, mais aussi leurs caractéristiques structurelles et sémantiques.

2. Idioms

2.1. Terminologie et caractéristiques d'idiome

Les études d'idiome font partie de la phraséologie. Selon Korhonen (2018 : 13), le terme *phraséologie* peut faire référence à deux concepts. Premièrement, il peut être compris comme une entité contenant toutes les expressions figées et lexicalisées d'une langue. Deuxièmement, il fait référence à une partie de la linguistique qui étudie les expressions figées et lexicalisées.

La terminologie du domaine de la phraséologie est diverse, ce qui s'explique partiellement par les différences linguistiques. Ainsi, en linguistique germanique l'*hyperonyme* d'autres types d'expressions est souvent un *phrasème* (Korhonen 2018 : 14). En linguistique anglaise, en revanche, le concept de *collocation* est particulièrement accentué (Granger & Paquot 2008 : 27–28). Pourtant, beaucoup de chercheurs sont d'accord sur le concept d'*expression figée*. Par exemple, Svensson (2004) définit l'*expression figée* comme *hyperonyme* d'idiome, *locution*, *collocation*, *phrasème*, *proverbe* et *gallicisme*.

Pour définir ce qu'est un idiome, il faut examiner ses caractéristiques. Tout d'abord, l'idiome est *polylexical*. C'est-à-dire qu'un idiome se compose toujours d'au moins deux parties (Heinonen 2013a : 8). Plus précisément, ces parties doivent être des mots lexicaux, pas grammaticaux, ajoute Korhonen (2018 : 33). Quant à son fonctionnement sémantique, l'idiome est typiquement non-compositionnel et opaque. Le premier réfère au fait que le sens d'un idiome n'est pas la somme des sens des mots qui le composent (Lamiroy 2008 : 6). Cela est une des qualités qui différencie *idiome* et *collocation*. En ce qui concerne l'opacité, les idiomes sont toujours plus ou moins opaques. Par exemple, le sens de l'idiome anglais *bury the hatchet* est plus facile de comprendre que celui de l'expression *les carottes sont cuites*. La connaissance de l'histoire et des métaphores aide à déduire le sens du premier idiome (Cacciari & Glucksberg 1991 : 230).

Souvent, les idiomes peuvent être compris de deux manières. Par exemple, l'expression mentionnée ci-dessus peut être, en lien au contexte, soit une phrase transparente dans son sens concrète, soit une phrase opaque n'ayant rien à faire avec les carottes (Svensson 2004 : 86). Dans ce cas, l'adjectif *opaque* pourrait être remplacé par *figuratif*. Il est typique qu'un idiomme puisse être interprété symboliquement et littéralement. Une personne peut vraiment *avoir les dents longues*, mais souvent l'interprétation littérale serait si absurde ou du moins si improbable que le sens figuratif serait plus convenant (Svensson 2004 : 74).

De la même manière que les idiomes sont plus ou moins opaques, ils sont plus ou moins figés (Mejri 2005 : 193). Autrement dit, un idiomme typique ne permet que peu, ou pas du tout, de modifications. Souvent le changement d'un mot, par exemple, n'est pas possible sans que le sens d'un idiomme change trop et qu'il ne puisse plus être considéré comme un idiomme. En outre, il existe le concept de *variation*, ce qui n'est pas la même chose que *modification*. Il y a un peu de variation parmi les idiomes. Korhonen (2018 : 38–40) distingue la variation lexicale et la variation morphosyntaxique. Il s'agit du premier mentionné si un mot lexical peut varier, comme dans l'expression finnoise *lyödä rukkaset pöytään/tiski* ('abandonner' ; litt. 'frapper les gants sur la table'). Dans la variation morphosyntaxique, par contre, la variante est un mot grammatical (*grincer des/les dents*).

Comme les autres unités lexicales, les idiomes sont conventionnels et ainsi lexicalisés. Selon Korhonen (2018 : 47), ils sont stockés dans la mémoire collective d'une communauté linguistique. Un idiomme peut être d'abord temporaire, et plus tard, devenir une partie du lexique (ibid.).

2.2. Fonctions d'idiome

Comme déjà mentionné dans l'introduction, une des fonctions de l'idiome est d'amuser. Selon Kortelainen (2011 : 50), les idiomes rendent la conversation plus légère, amusante et détendue. Une manière de jouer avec la langue et d'utiliser un idiomme qui peut être interprété à la fois figurément et littéralement (Muikku-Werner 2009 : 86).

D'après Muikku-Werner (2009 : 84), en utilisant des idiomes, le locuteur peut régler la formalité et le style de son texte. Les expressions représentent différents degrés de formalité : depuis *vulgaire* jusqu'à *très formel* voire *sophistiqué* (Korhonen 2018 : 156–

160). Un idiomme peut aussi révéler l'attitude du locuteur (Muikku-Werner 2009 : 85). Comme exemple, Muikku-Werner mentionne la phrase finnoise *Pekka ei sylje lasiin*. Elle se traduit littéralement par 'Pekka ne crache pas dans le verre', ce qui veut dire que la personne boit beaucoup d'alcool. Il y énormément d'idiomes plus ou moins neutres en finnois pour exprimer le fait que quelqu'un boit (beaucoup). En choisissant cet idiomme, le locuteur fait le choix d'exprimer sa désapprobation.

Les sujets délicats et les émotions difficiles sont souvent exprimés par des euphémismes (Muikku-Werner 2009 : 86–87). Par conséquent, l'ivresse est un sujet qui apparaît dans beaucoup d'idiomes dans différentes langues. Kortelainen (2011 : 6) ajoute que la stupidité est clairement un sujet de discussion fréquent, car il existe particulièrement beaucoup d'idiomes finnois pour l'exprimer.

Une raison pour utiliser des idiomes est leur économie (Habler & Hümmmer 2005 : 107). Même un idiomme court peut avoir plusieurs sens et nuances. Pour les exprimer autrement, par une paraphrase, il faudrait utiliser plusieurs phrases. En plus des fonctions présentées, nous ajoutons que les idiomes ont un rôle communautaire important. Il y a des idiomes linguistiquement et culturellement spécifiques dont le sens, bien qu'il soit traduit, ne peut pas être compris que par un certain groupe de gens.

3. Corpus et collecte des idiomes

Internet permet d'étudier beaucoup de phénomènes linguistiques. Kilgarriff et Grefenstette (2003 : 333) le décrivent comme un corpus énorme auquel tout le monde a accès et où sont représentés différentes langues et de différents dialectes. Un grand avantage des corpus en ligne est naturellement leur taille (Heinonen 2013a : 27). Un autre avantage, d'après Heinonen (ibid. p. 26), est la diversité. De tels corpus contiennent des messages laissés sur des forums et aussi des articles médicaux, ainsi que des publicités et des textes juridiques, par exemple.

Quant aux désavantages du corpus en ligne, ils contiennent probablement plus d'erreurs que les publications imprimées. Cela s'explique par la large gamme d'auteurs (Kilgarriff & Grefenstette 2003 : 342). Il est plus probable de trouver une expression mal utilisée sur un blog ou un forum de discussion que sur un journal, par exemple. Un autre désavantage, selon Kilgarriff et Grefenstette (2003 : 341), est la difficulté ou même l'impossibilité de classer des textes en ligne par genre.

Notre corpus est composé de phrases finnoises et françaises existant sur Internet et contenant le mot *hammas/dent*. Elles sont prises sur une interface utilisateur générale (bionlp-www.utu.fi/dep_search/), maintenue par Turku NLP Group. Il s'agit d'un groupe pluridisciplinaire de chercheurs dont le domaine est le traitement automatique des langues.

L'interface utilisateur contient des bases de données en plus de 60 langues différentes. Notre intention était d'examiner les 200 premières phrases autant en finnois qu'en français. Mais comme le nombre des prises françaises n'a été que 198, nous avons choisi 198 phrases également en finnois. La collecte de données a été menée du 10 au 11 avril 2018. Si la collecte était répétée, les résultats pourraient varier, car le corpus peut être modifié plus tard.

4. Analyse

4.1. Analyse structurelle

Nous examinons chaque phrase pour trouver s'il s'agit d'un idiomme ou si le mot *dent* est utilisé dans un sens non-figuratif. Ensuite, nous divisons les idiomes en quatre groupes : idiomes verbaux finnois, idiomes substantivés finnois, idiomes verbaux français et idiomes substantivés français. La catégorisation se base sur la classe lexicale de l'idiome qui est syntaxiquement dominant. Selon Korhonen (2018 : 50), il y a quatre classes lexicales possibles : verbe, substantif, adjectif et adverbe. En ce qui concerne les idiomes de notre corpus, ils sont tous soit des idiomes verbaux, soit des idiomes substantivés. Par exemple *avoir les dents longues* fait partie des premiers mentionnés, alors qu'*un coup de dent* des derniers.

Nous présentons plusieurs exemples, mais pas tous les idiomes individuels. Toutes les expressions du corpus sont cependant listées dans les tableaux 1–4 dans la partie finnoise de cette étude.

Les nombres d'idiomes différents sont visibles dans le tableau ci-dessous. Le corpus finnois contient alors 12 idiomes divers dont 7 sont des idiomes verbaux et 5 des idiomes substantivés. Quant aux idiomes français, les mêmes nombres sont 20, 16 et 4. Ce qui est remarquable, c'est le grand nombre de différents idiomes verbaux français. Il

faut pourtant noter ici ce qui concerne les résultats de cette étude : de grandes généralisations ne peuvent pas être faites, car l'échantillonnage est relativement petit.

Tableau 1. Nombres d'idiomes différents.

Nombres d'idiomes différents		
	finnois	français
Idiomes verbaux	7	16
Idiomes substantivés	5	4
Au total	12	20

Les idiomes verbaux finnois apparaissant le plus souvent dans le corpus sont : *olla jkn hampaissa* ('être dominé par quelqu'un') (8) et *purra hammasta* (*serrer les dents*) (5). Le chiffre entre les dernières parenthèses présente le nombre des cas où l'idiome est mentionné dans le corpus. Les idiomes substantivés les plus fréquentes sont *kynsin hampain* ('de toute manière') (5), *pitkin hampain* ('involontairement') (5) et *hampaat irvessä* ('grimacant d'involonté') (4). Tous les autres idiomes du corpus finnois y apparaissent seulement une ou deux fois.

Dans le corpus français, les idiomes verbaux les plus fréquentes sont *grincer les dents* (6), *croquer à pleine dent* (5) et *se faire les dents sur qc* (5). Quant aux idiomes substantivés, les plus communs sont *en dent de scie* (4) et *grincements de dents* (3). Le dernier mentionné pourrait être combiné avec *grincer les dents*, mais comme il s'agit particulièrement d'une catégorisation morphosyntaxique, nous avons décidé de le classer ainsi. Les expressions françaises non mentionnées n'apparaissent qu'une fois dans le corpus.

Bien que la classe lexicale soit toujours la même, il y a de la variation parmi quelques idiomes. En français, le seul cas est *grincer des/les dents*. En finnois, il y en a plus et il s'agit de variation lexicale.

Dans le deuxième tableau sont présentés les nombres totaux d'idiomes, en dépit des idiomes individuels et de leurs fréquences. Même si la comparaison des nombres

d'idiomes verbaux (fi 20 vs fr 35) et substantivés (fi 17 vs fr 9) indique des différences, le nombre total est assez similaire dans les deux langues (fi 37 vs fr 44).

Tableau 2. Nombres totales d'idiomes.

Nombres totales d'idiomes		
	finnois	français
Idiomes verbaux	20	35
Idiomes substantifs	17	9
Au total	37/198 (= 19 %)	44/198 (= 22 %)

4.2. Analyse sémantique

L'examen des idiomes montre que, dans les deux langues, il n'y a que peu ou pas du tout d'idiome avec le mot *dent* qui auraient un sens appréciatif. Quelques expressions sont plutôt neutres, mais la plupart ont un ton dépréciatif. Par conséquent, cette étude conforte l'idée de Habler et Hümmer (2005 : 107) selon laquelle il existe plus d'idiomes exprimant des émotions et actions dépréciatives.

Le seul idiome que nous considérons comme appréciatif, c'est l'idiome français *croquer à pleine dent*. La majorité des expressions sont dépréciatives dans les deux langues : 8/12 en finnois et 13/20 en français. Les émotions dépréciatives auxquelles font référence les idiomes de notre corpus sont : la peur et l'intimidation (*avoir la dent dure; montrer les dents*), la haine et l'irritation (*grincer les dents/grincements de dents; péter les dents*), le ressentiment (*dent contre dent*), la vindicativité (*avoir le couteau entre les dents*), la malveillance (*avoir une dent contre qqn*) et la réticence (*serrer les dents*). De plus, les idiomes réfèrent, par exemple, à l'échec (*casser les dents*) et à la pauvreté (*rien à se mettre sous la dent*). Il y a beaucoup d'idiomes dans notre corpus qui sont les mêmes en finnois et en français et, à cause de cela, ce n'est pas surprenant que les sujets des expressions soient principalement les mêmes.

Des idiomes neutres, d'après nous, sont, par exemple, les idiomes *en dent de scie*, *se faire les dents sur qc*, *x hammasta kolottaa* ('avoir envie de (manger) x'), *olla hampaita myöten aseistautunut* et sa version française *être armé jusqu'aux dents*. Dans cette catégorie, nous classifions également quelques idiomes dont le sens peut, dépendant du contexte, être interprété comme positif/dépréciatif. Par exemple *avoir les dents longues* et *kynsin hampain* ('de toute manière') expriment l'ambition et la détermination qui ne sont pas clairement de bonnes ou mauvaises caractéristiques.

5. Résultats

Nos questions de recherche sont les suivantes :

- 1.) Quels types d'idiome comporte le mot *dent* ?
- 2.) Y a-t-il des différences entre le finnois et le français concernant les idiomes et leur quantité dans le corpus ?

aux deux questions est visible dans les tableaux 1 et 2. Sur 198 phrases finnoises contenant le mot *dent*, il y a 37 phrases où il s'agit d'un idiome. En français, ce nombre est 44 ; c'est-à-dire qu'environ un cinquième dans les deux cas. Par contre, en comparant les idiomes verbaux et substantivés, les différences sont plus claires. En finnois, le rapport entre les deux catégories est assez égal, alors qu'en français il y a quatre fois plus d'idiomes verbaux que substantivés. Quant aux nombres d'idiomes différents, chacun apparaissant d'une à huit fois dans le corpus, les rapports sont les mêmes. Sur 12 idiomes finnois différents, 5 sont verbaux et 7 substantifs. Sur 20 idiomes français différents, même 16 sont verbaux et seulement 4 substantivés.

Pour répondre à la première question, nous pouvons aussi nommer les idiomes les plus fréquents dans notre corpus. En finnois, ce sont *olla jkn hampaissa* ('être dominé par quelqu'un'), *purra hammasta* (*serrer les dents*), *kynsin hampain* ('de toute manière'), *pitkin hampain* ('involontairement') et *hampaat irvessä* ('grimacant d'involonté'). En français, les idiomes les plus fréquents sont *grincer les dents*, *croquer à pleine dent*, *se faire les dents sur qc*, *en dent de scie* et *grincements de dents*.

les résultats sont assez similaires dans les deux langues. En français, il y a un idiome clairement appréciatif (*croquer à pleine dent*). Autrement, le ton des idiomes est soit neutre soit dépréciatif, le dernier mentionné étant le plus fréquent.

Comme mentionné auparavant, le nombre des prises dans cette étude est relativement petit, alors les résultats ne sont pas tout à fait généralisables. Pourtant, quelques conclusions prudentes peuvent être faites. Le grand nombre des idiomes verbaux français peut indiquer qu'ils sont plus fréquents que les idiomes substantivés en français et aussi plus fréquents que les idiomes verbaux en finnois. Les résultats montrent également que le sens de ce type d'idiome est le plus souvent dépréciatif.