

Tapaustutkimus monikielisten kirjoittajien kirjoitusprosesseista: Näppäilyntallennus apuna sanastokeskittymien tunnistamisessa

MAARIT MUTTA, PÄIVI LAINE

Turun yliopisto

Tiivistelmä. Artikkeliki käsittelee monikielisten kirjoittajien kirjoittamisprosessien sujuvuutta. Oppiakorpuksen keräämiseen ja analysointiin käytimme graafiteoriaan perustuva näppäilyntallennusohjelmaa GenoGraphiX-Log (GGXLog), jonka visualisointityökalut tarjoavat uuden lähestymistavan kirjoitusprosessin sujuvuuden analysointiin. Koko aineisto koostuu 21 yliopisto-opiskelijan kirjoittamasta suomenkielisestä (L1) ja ranskankielisestä (L2) mielipidekirjoituksesta. Tässä tapaustutkimuksessa keskitymme kuitenkin vain kolmen opiskelijan kirjoitusprosesseihin ja kirjoittajaprofiileihin. Tulokset osoittavat, että kirjoitusprosessien visualisoinnin avulla pystymme tunnistamaan ja havainnollistamaan sujuvaa kirjoittamista edustavia purskeita yksityiskohtaisemmin. Lisäksi kirjoittajan tuottamat purskeet ja sanastokeskittymät voivat paljastaa kirjoittajien yksilöllisiä kirjoittajaprofiileja niin L1- kuin L2-kirjoituksessa. Tulokset auttavat parantamaan pedagogisia käytäntöjä eri koulutusympäristöissä. Lisäksi kirjoitusprosessien analyysi antaa oppijoille mahdollisuuden tunnistaa oman kirjoittajaprofilinsa.

Avainsanat: kirjoittajaprofiilit; näppäilyntallennusohjelma; sujuvus; visualisointi; pedagoginen työkalu; kielenopetus

1. Johdanto

Elämme jatkuvasti muuttuvassa maailmassa, jossa vaaditaan nopeaa ja sujuvaa kielenkäyttöä, esimerkiksi kirjoittamista erilaisilla kirjoitus-alustoilla. Teknologisoituminen on muokannut kirjoittamista entistä enemmän, vaikkei olekaan syrjäyttänyt perinteistä kynällä ja paperilla tapahtuvaa kirjoittamista kokonaan (ks. esim. Lintunen ym. 2017). Jokaisella on saatavilla esimerkiksi kämmenenkokoinen tietokone, puhelin, jolla varsinkin nuori sukupolvi on tottunut hoitamaan arkielämän kirjoittamisen, tiedonhaun, laskujen maksamisen ynnä muun. Puhelimella kirjoitetut viestit ovat tyypillisesti lyhyitä. Lyhyiden viestien kirjoittaminen ja lukeminen vaikuttavat pitkien ja suunnitelmallisten tekstin tuottamiseen ja saattavat vaikuttaa eri tavoin myös yliopisto-opiskelijoiden tapaan tuottaa argumentoivaa tekstiä.

Saapuessaan yliopistoon opiskelemaan vieraita kieliä opiskelijoilla on mukanaan monimuotoinen kielirepertuaari, eli he ovat sillä tavalla monikielisiä, että osaavat useampia kieliä ensikielensä lisäksi (ks. esim. Martin 2016; Mutta & Salminen 2021). Heidän teknologiset taitonsa sekä kirjoitustapansa kuitenkin vaihtelevat suuresti (ks. esim. Lintunen ym. 2017). Jotkut ovat sujuvia viestimään lyhyellä ja puhekielimäisellä tyylillä pikaviestisovelluksissa ja sosiaalisen median alustoilla (esim. WhatsApp, Facebook ja Instagram), kun taas toiset kirjoittavat päiväkirjoja ja jopa pieniä kirjoitelmia omaksi huvikseen vapaa-aikanaan kotitehtävien ohella (opiskelijoiden kirjoittamispuhdistuksista ks. Laine & Mutta tulossa). Lisäksi tietokoneella kirjoitettaessa kymmensormijärjestelmä tai ylipäättään kyky kirjoittaa näppäimistöllä vaikuttaa sujuvaan tekstin tuottamiseen (esim. Van Waes ym. 2021).

Tässä artikkelissa olemme kiinnostuneita kolmen monikielisen kirjoittajan kirjoitusprosessien sujuvuudesta. Tutkimuksessamme olemme kiinnostuneita sekä kirjoittajien ilman taukoa tuottamista jaksoista eli purskeista (engl. *bursts*) että sanastokeskittymistä (esim. *minun mielestäni, à mon avis*). Purskeet määritellään yleensä niin, että kirjoitusrupeaman molemmiin puolin määritellyn tauon tulee olla tietynpituisen

eli vähintään kaksi sekuntia. Tällaista taukoa kutsutaan kognitiiviseksi tauoksi (Chenoweth & Hayes 2001: 83, 88; Kowal 2014; Cislaru & Olive 2018). Purskeet voidaan määritellä tekstiksi, joka tuotetaan kyseisten taukojen välissä ilman korjauksia tai sisältäen niitä (engl. *P-burst* vs. *R-burst*).

Taukojen määrä, pituus ja paikka tekstissä sekä purskeiden pituus ovat keskeisiä kirjoitusprosessin sujuvuuden mittareita; sujuvuus on yleisesti ottaen kirjoitusprosessin vaivatonta etenemistä ilman pitkiä taukoja ja korjauksia (ks. esim. Spellman Miller 2006; Mutta 2020). Purskeiden – ja alla esiteltävien sanastokeskittymien – tuottaminen näyttäisi olevan automatisoitunutta siten, etteivät ne tuota taukoja kirjoitusprosessiin (Cislaru & Olive 2018). Tätä pidetään sujuvan kirjoittamisen yhtenä peruskriteerinä. Silloin kognitiivinen kuormittavuus vähenee ja kielellisiä resursseja vapautuu sisällön suunnitteluun ja organisointiin (esim. Snellings ym. 2002; ks. esim. Mutta 2020). Ensikielellä (L1) kirjoitetaan pidempiä purskeita kuin vieraalla kielellä (L2), ja myös hyvä L2-kielitaito tukee pidempien purskeiden tuottamista (Chenoweth & Hayes 2001). Tauottomien purskeiden tunnistaminen auttaa sanastokeskittymien analysointia, mutta on huomioitava, että yksittäinen purske voi sisältää useamman sanastokeskittymän (ks. luku 4.2).

Tarkastelemme sujuvuutta purskeiden lisäksi etenkin sanastokeskittymien (engl. *formulaic sequence*) avulla. Sanastokeskittymät tai -köntit ovat monisanaisia tai monimorfeemisiä rakenteita, joista käytetään kirjallisuudessa eri nimityksiä. Esimerkiksi Wray (2002: 9) mainitsee termit *amalgams*, *chunks*, *collocations*, *conventionalized forms*, *fixed expressions*, *schemata* ja Perez-Bettan (2015: 36) lisäksi termit *lexical phrases*, *prefabricated routines*, *memorised sentences*, *sentence stems* viitaten eri kielitieteen ja psykolingvistiikan tutkijoihin. Salminen (2020: 13) on tehnyt kattavan koostetaulukon erilaisista nimityksistä, joilla tätä ilmiötä kuvataan.

Sanastokeskittymät voivat siis koostua sanoista, lausekkeista tai konstruktioista. Sanastokeskittymät ovat olennainen osa kirjoittamisen sujuvuutta, ja ne edesauttavat sujuvuutta kaikilla kielitaidon osa-alueilla,

koska ne tuotetaan omina merkityssisältöinä, ja siten niiden aktivoituminen produktiivisesti tai reseptiivisesti on nopeaa ja vaivatonta (ks. esim. Lintunen ym. 2020; myös Segalowitz 2010). Sanastokeskittymien katsotaan varastoituneen pitkäkestoiseen muistiin kokonaisuuksina ja aktivoituvan kerralla (esim. Schmitt 2004; Reiman 2011; Perez-Bettan 2015; Cislaru & Olive 2018; Lieri 2021), koska niiden käyttö on automatisoitunutta.

Sanastokeskittymät ovat kielikohtaisia, mutta niitä voidaan kaikissa kielissä luokitella erilaisiin luokkiin, esimerkiksi tavallisimmin esiintyviin leksikaalisiin ja diskursiivisiin keskittymiin (esim. Wray 2002; Forsberg 2006; Rasmussen 2015; Salminen 2020; Mutta & Salminen 2021). Oppijankielen tutkimuksessa puheen sujuvuutta lisääviä sanastokeskittymiä on tutkittu varsinkin suullisen kielitaidon osana (mm. Wray 2002, 2008; Forsberg 2006; Cordier 2013; Perez-Bettan 2015; Rasmussen 2015). Kirjoitetun kielen tutkimusta on etenkin englannin kielestä, mutta myös muita kielipareja on tutkittu. Esimerkiksi Jaworskan ym. (2015) tutkimuksessa vertailtiin saksaa ensi- ja vieraana kielenä puhuvien sanastokeskittymien käyttöä argumentatiivisessa tekstissä. Heidän tutkimuksessaan hyvin saksaa hallitsevat englantilaiset yliopisto-opiskelijat osasivat käyttää sanastokeskittymiä monipuolisesti, mutta keskittymät eivät olleet samoja kuin natiivien käyttämät. Tässä artikkelissa keskitymme samoin kirjoitettuun kieleen mutta vertaamme samojen kirjoittajien kirjoitusprosesseja sekä suomen että ranskan kielellä.

Käytämme tutkimuksessa apuna nykyteknologian mahdollistamaa aineistonkeruu- ja aineistonanalysointiohjelmaa eli GenoGraphiX-Lognäppäilyntallennusohjelmaa (myöhemmin GGXLog), jonka avulla voidaan tarkastella kirjoitusprosessin eri vaiheita (esim. poistot, lisäykset) sekä taukojen sijaintia ja kestoja. Graafiteoriaan pohjautuva GGXLog-ohjelma mahdollistaa myös kirjoitusprosessien visualisoinnin (Leblay & Caparossi 2015) ja lisää tietoa monikielisten kirjoittajien strategisesta kompetenssista eli tekstin rakenteen suunnittelusta ja tuottamisesta olemassa olevin resurssein ja osana heidän kirjoittajaprofiliaan (profileista ks. esim. Van Waes & Schellens 2003; Mutta 2020). Visualisointeja

vertaamalla voidaan tarkastella myös strategisen kompetenssin siirtymistä eri kielten välillä.

Tapaustutkimuksemme aineisto koostuu suomalaisen yliopiston kolmen ranskan perusopintojen kurssin opiskelijan suomeksi ja ranskaksi kirjoittamista kirjoitelmista. Aineisto on osa laajempaa KISUVI-hankkeen¹ aineistoa. Keskitymme näiden kirjoittajien kirjoitusprosesseihin ja kirjoittajaprofileihin sujuvuuden osalta siten, että tutkimme purskeita ja niiden sisältämiä sanastokeskittymiä. Nämä kirjoittajat osallistuivat vapaaehtoisesti myös kirjoitusprosessin kielentämistä varten tehtyyn haastatteluun, joka jää kuitenkin tässä artikkelissa analyysin ulkopuolelle (kielentämisestä ks. Laine & Mutta tulossa). Pohdimme artikkelissa seuraavaa tutkimuskysymystä: Mitä kirjoitusprosessien näppäilyntalennus kertoo kirjoittajien tuottamista purskeista, sanastokeskittymistä ja sujuvuudesta L1- ja L2-kielissä?

Ennen aineiston keräämisen esittelyä ja analyysiä määrittelemme, mitä sanastokeskittymät ja purskeet ovat ja miten ne liittyvät sujuvuuteen niin ensi- kuin vieraalla kielellä kirjoitettaessa.

2. Sanastokeskittymät ja niiden tunnistaminen

2.1. Oppijankielen sanastokeskittymien aiempi tutkimus

Automaattisesti tuotettuja ilmauksia, kuten sanastokeskittymiä, on kaikissa kielissä, mutta niiden määrittelyt ja luokittelut vaihtelevat kielen mukaan. Wrayn (2002) mukaan sanastokeskittyminä voidaan pitää monisanaisia tai -morfeemisia ilmauksia, joissa on vähintään kaksi sanaa, morfeemia tai kielellistä elementtiä. Ne voidaan jaotella oppija-kohtaisiin (engl. *learner-internal* eli *processing units*) ja kielikohtaisiin (*learner-external* eli *linguistic clusters*) ilmauksiin (Forsberg 2006, 2008; Myles & Cordier 2017; Siyanova-Chanturia & Pellicer-Sánchez 2019; ks. myös kattava kooste ilmiöön liittyvistä nimityksistä Salminen 2020: 13).

¹ Monikielisten kirjoittajien kirjoitusprosessit: sanastokeskittymien ja sujuvuuden visualisointi graafiteorian avulla; hanketta on tukenut Koneen Säätiö.

Kielikohtaiset ilmaukset ovat yleisesti tietyssä kielessä tunnistettavia sanastokeskittymiä (esim. *for example, je pense que* ‘mielestäni/ajattelen, että’), kun taas oppijakohtaiset voivat olla jopa täysin tietylle oppijalla tyypillisiä keskittymiä.

Sanastokeskittymien määrittelyssä on edelleen haasteita, varsinkin typologisesti erilaisten kielten kohdalla. Esimerkiksi suomen kielen kaltaiset, morfologisesti rikkaat kielet ovat haasteellisia sanastokeskittymien tutkimuksessa käytetyille määritelmille. Durrantin (2013) mukaan agglutinoivissa kielissä (esim. turkki ja suomi) sanojen ja ilmausten holistinen prosessointi näyttäisi olevan harvinaisempaa kuin kielissä, joissa ei ole morfologista vaihtelua yhtä paljon (esim. germaaniset kielet). Näyttää siltä, että useampisanaiset sanastokeskittymät tuotetaan esimerkiksi englannissa ja ruotsissa eri tavalla kuin useasta morfeemista koostuvat sanat (ks. myös Mäntylä ym. 2020). Lisää tutkimusta kuitenkin tarvitaan esimerkiksi saman kirjoittajan eri kieliperheisiin kuuluvilla L1- ja L2-kielillä ja eri taitotasoilla, jotta vertailua voitaisiin tehdä.

Indoeurooppalaisissa kielissä, kuten englannissa, ruotsissa ja ranskassa, prepositioilmaukset voidaan katsoa sanastokeskittymiksi, mutta suomen kielessä prepositioilmauksia vastaavien ilmausten määrittelyminen sellaisiksi on osittain ongelmallista: Miten esimerkiksi postpositioilmaukset *koulun luona* ja *repun sisällä* eroavat kirjoitusprosessissa sijamuodoista *koululla* ja *repussa*? Tuotetaanko postpositioilmaukset suomen kielessä yhtä produktiivisesti kuin niitä vastaavat sijamuodot? Entä voiko esimerkiksi sijamuotoja ja niiden muodostamia kongruoivia lausekkeita (esim. *suuressa koulussa*) pitää samanlaisina sanastokeskittyminä kuin suomen kielen adpositiolausekkeita (esim. *ennen koulua* tai *koulun jälkeen*)? (Ks. myös Mäntylä ym. 2020; taivutusparadigmojen produktiivisuudesta tarkemmin ISK § 53.)

Jantunen (2009, 2011) on tarkastellut suomea oppijankielenä (S2) korpusvetoisesti avainsana-analyysin avulla ja tuonut esiin muun muassa joitakin sille tyypillisiä fraseologisia ilmauksia, kuten *minulla on*. Ruotsin osalta voidaan mainita esimerkiksi Prentice (2010) sekä Prentice ja Sköldberg (2010), jotka ovat tutkineet konventionaalisia ilmauksia,

idiomeja ja sanojen yhdistelmiä maahanmuuttajien ruotsissa. Heidän tuloksensa osoittavat, että opiskelijat käyttivät pääasiassa yleisesti hyväksytyjä ilmauksia (esim. *klättra på väggarna*), mutta myös muokattuja tai aivan uusia yhdistelmiä (esim. *mjuk som en muffins*; Prentice & Sköldb-berg 2010: 18–20), ja näistä jälkimmäiset olivat yleisempiä maahanmuuttajataustaisilla opiskelijoilla. Sundman (2011: 331–335) puolestaan tutki 7.–9.-luokkalaisten aineista sanastokeskittymien käyttöä ja päätteli analyysin perusteella, että yleisimmin esiintyvien muotojen käyttö on opittu hyvin (esim. *jag är x år gammal*) ja että eksplisiittinen sanastokeskittymien opettaminen hyödyttää muotojen oppimista.

Mäntylä ym. (2020) havaitsivat kieli- ja oppijakohtaisia eroja sanastokeskittymien tuottamisessa typologisesti erilaisten kielten eli englannin, suomen ja ruotsin oppijoiden välillä, ja eroihin vaikutti osittain oppijan ensikieli (suomi tai ruotsi). Heidän tuloksensa osoittivat, että kirjoittajan ensikieli vaikutti sekä sanastokeskittymien määrään ja laatuun vieraalla kielellä kirjoitettaessa että oppijakohtaisten ja kielikohtaisten ilmausten määrään. Mutta ja Salminen (2021) ovat puolestaan tutkineet tapaustutkimuksessa yliopistotasoisien monikielisten kirjoittajien sanastokeskittymien käytön eroja L2-ranskan ja L2-ruotsin tuottamisessa. Tulokset osoittivat, että opiskelijat tuottivat enemmän sanastokeskittymiä ranskaksi kuin ruotsiksi, mutta sanastokeskittymien käytössä oli paljon yksilöllisiä eroavuuksia sekä tuotoksen että prosessin tasolla. Tutkimuksessa hyödynnettiin retrospektiivisiä haastatteluja kirjoitusprosessien ja -produktion metakielellisen analyysin tukena (ks. myös Toropainen ym. 2020).

Oppijankielen sanastokeskittymätutkimuksen esittelyn jälkeen siirrymme tarkastelemaan, miten sanastokeskittymiä voidaan identifioida kielenoppijoiden tuotoksista.

2.2. Sanastokeskittymien tunnistaminen purskeista

Sanastokeskittymät ovat automaattisesti tuotettuja kokonaisuuksia, ja niiden käyttöä voidaan kirjoittamisessa tutkia kronometrinen menetelmien, kuten prosessien tutkimuksen, sekä ei-kronometrinen menetelmien, kuten metakielellisten haastatteluiden, avulla (esim. Mutta 2017; Toropainen ym. 2020). Sanastokeskittymän tunnistaminen koetaan ongelmalliseksi luvussa 2.1 esitettyjen seikkojen lisäksi siksi, että päätös keskittymäksi identifioimisesta on aina subjektiivinen (esim. Wray 2008; Siyanova-Chanturia & Pellicer-Sánchez 2019).

Objektiivisessa määrittelyssä voidaan käyttää apuna näppäilyntalennusohjelmia, joilla voidaan tarkastella kirjoitusprosessin eri vaiheita, esimerkiksi tekstin poistoa, lisäämistä ja taukojen pituuksia (esim. Lindgren & Sullivan 2019). Kirjoittajan pitämien taukojen ja tekemien korjausten perusteella tuotoksesta pyritään tunnistamaan tietynpituisia kirjoitusrupeamia eli purskeita. Purskeet määritellään yleensä niin, että kirjoitusrupeaman molemmin puolin määritellyn tauon tulee olla vähintään kahden sekunnin pituinen (Chenoweth & Hayes 2001: 83–88; Kowal 2014; Cislaru & Olive 2018). Kahden sekunnin taukoa pidetään kirjoittamisen tutkimuksissa yleisesti hyväksyttynä kognitiivisen tauon minimipituutena, mikä tarkoittaa, että tauko tunnistetaan ajatukselliseksi tauoksi (vs. tekninen tauko, joka tarkoittaa esimerkiksi kirjoittajan tekemää automaattista kirjoitusvirheen korjaamista siten, että tauko on alle kaksi sekuntia).

Kirjoittajille voidaan myös määritellä henkilökohtaisia tauon pituuksia esimerkiksi tilastollisten mallien, kuten Markovin mallin, avulla (ks. esim. Mutta 2017; 2020). Tauon pituus voi olla myös määritelty lyhyemmäksi (1,5 sek), kuten Mäntylän ym. (2020: 113–114) tutkimuksessa. Aiemmat tutkimukset osoittavat, että purskeet ovat pidempiä L1-teksteissä kuin L2-teksteissä, koska vieraalla kielellä kirjoittaessa tauot saattavat sijoittua sanojen sisällekkin, kun taas ensikielessä ne ovat pääosin lauseiden ja kappaleiden välissä (Spelman Miller 2006: 141). Purskeiden pituuden on myös havaittu korreloivan kielitaitotason kanssa (Chenoweth & Hayes 2001; ks. myös Palviainen ym. 2012).

Purskeet voivat koostua yksittäisistä sanoista tai yhdestä tai useammasta sanastokeskittymästä. Esimerkiksi purskeesta [---] *meistä ei pysty äärimmäisyyksiin ja loppuviimein ylisuorittaminen ja mahdollinen loppuunpalaminen* *kostautuu* [---] voi erottaa kolme sanastokeskittymää: *ei pysty*, *loppuviimein* ja *mahdollinen loppuunpalaminen* (ks. taulukko 3). On huomattava, että kaikki samassa purskeessa esiintyvät sanat eivät siis muodosta sanastokeskittymää, esim. *mun mielestä* on sanastokeskittymä mutta *kissa istuu pöydällä* ei ole. Oletuksena on, että sanastokeskittymät ovat analysoimattomia kiteymiä, eikä niissä ei pidetä taukoja tai tehdä korjauksia sanastokeskittymän sisällä, vaan tauot pidetään juuri niitä ennen ja niiden jälkeen (esim. Mäntylä ym. 2020; Mutta & Salminen 2021).

3. Tutkimuksen aineistonkeruu- ja analysointimenetelmät ja aineisto

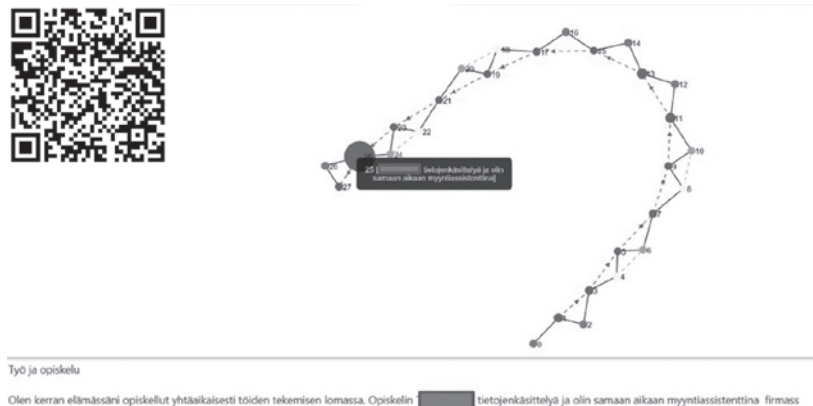
3.1. Näppäilyntallennusohjelma ja visualisointityökalu

Aiemmissa tutkimuksissa ei ole yhdistetty kirjoitusprosessien sujuvuustutkimusta, purskeita ja sanastokeskittymiä sekä niiden analyysiä matemaattista graafiteoriaa hyödyntävän ohjelman avulla. Kompleksisen kirjoitusprosessin kuvaaminen on hankalaa, koska kirjoitusprosessi ei etene lineaarisesti pisteestä A pisteeseen B. Perinteinen lineaarisen tekstin etenemisen kuvaus esimerkiksi ScriptLog-näppäilyntallennusohjelmalla antaa esimerkin 1 kaltaisen näkymän (yksi GGXLogin monista toiminnoista; ks. Usoof ym. 2020: 37).

- (1) <MOUSE EVENT>Työ ja opiskelu<RETURN><RETURN><3857>
Olen kerran ep<BACKSPACE>lämässäni oll<BACKSPACE2>piskellut
yhtäaikaaisesti<2416>töiden tekemisen lomassa<3009>.<3152>
Opiskelu<BACKSPACE>in t<BACKSPACE>X:lla tietojenkäsittelyä
ja olin samaan aikaan myyntiasistenttina<6704> firmass <BACK-
SPACE8>

Esimerkki 1 kuvaa kirjoittajan etenemistä näyttämällä kirjoitetun tekstin, siirtymiset hiirellä, tekstin poistamisen ja tauot sekunteina, mutta toimintojen suhteita toisiinsa on vaikea havaita. Matemaattiseen graafiteoriasta pohjautuva visualisointi tuo erilaisen perspektiivin kirjoitusprosessien kuvaukseen ja avaa kirjoitusprosessia uudella tavalla sekä kirjoittajille että opettajille ja tutkijoille (Leblay & Caparossi 2015; Bécotte-Boutin ym. 2019; Leblay ym. 2022). Graafiteoriaan pohjautuvassa ohjelmassa eri toimintojen suhteita, kuten kirjoitusprosessien yhteyksiä ja taukojen pituuksia, visualisoidaan erilaisten värien, kaarien ja toimintojen kestoja kuvaavien solmujen tai pallojen (engl. *nodes*) avulla. Ne kertovat, miten teksti rakentuu ja millaisia ovat yhtenäiset sujuvat purskeet. Prosessikuvio helpottaa tekstin rakentumisen observointia ja täydentää muita analyysejä, joita saadaan ohjelman tuottamista taulukoista ja piirakkakuvioista. Eri palloja tarkastelemalla saadaan esimerkiksi näkyviin, mitä kirjoittaja on kohdassa tuottanut tai poistanut. Seuraavassa esimerkin 1 visualisointi esittää kirjoitusprosessin alkupuolelta tuotetun lineaarisen tekstin kuviona (kuvio 1). Värilliset kuviot ovat saatavilla QR-koodin takaa.

Kuviossa 1 näkyvä suurin pallo 25 on monisanainen purske *[X:llä] tietojenkäsittelyä ja olin samaan aikaan myyntiassistenttina*. GGXLogin



KUVIO 1. Kirjoitusprosessin eri vaiheiden visualisointi GGXLog-ohjelmassa

visualisointi havainnollistaa kirjoituksen etenemistä viivoin siten, että se kuvaa purskeiden, korjausten ja taukojen pituutta erikokoisin ja -värisin palloin (ks. Usoof ym. 2020). Punaiset pallot (1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27) havainnollistavat tuotettuja purskeita, vihreät (0, 2, 12, 14, 16, 26) taukoja ja siniset (6, 10, 20, 24) ja keltaiset (4, 8, 18, 22) poistoja ja korjauksia. Yhtenäinen viiva kuvaa kirjoittajan etenemistä ja katkoviiva valmista tuotosta. Käyrän muodolla ja suunnalla ei ole graafissa merkitystä, vaan progressiivinen kuvio etenee tilan mukaan. Huomioitavaa on, että kuvio on yksittäisessä analyysissä aina samanlainen mutta että kuvio ei kerro visuaalisena kuviona kognitiivisesta prosessista. Prosessista kertovat vain pallot ja käyrät ja niiden väliset suhteet (esim. taukojen pituudet, korjaukset, purskeet).

Kirjoittamisen sujuvuutta tässä artikkelissa ilmaisevan mittarin (= kuinka monta merkkiä minuutissa kirjoitetaan keskimäärin purskeen sisällä) määrittelyssä käytimme kognitiivista kahden sekunnin taukoa ja niiden välissä ilman korjauksia tuotettua merkkimäärää minuutissa (ks. *Burst and Efficiency* Usoof ym. 2020: 42–43; ks. myös Spelman Miller ym. 2008). Laskimme purskeiden ilmoittaman sujuvuuden merkkimäärästä emmekä sanamäärästä, koska se sopii paremmin vertailuun morfologisesti erilaisissa kielissä (vrt. Palvianen ym. 2012: 51). Ohjelma mahdollistaa toki myös muiden tilastojen tarkastelun, kuten kokonaismerkkimäärä minuutissa, tekstin pituus ja poistojen määrän prosessin aikana.

Tässä tapaustutkimuksessa keskityimme kirjoitusprosessin sujuvuuteen yhtäjaksoisten purskeiden ja sanastokeskittymien tarkastelun avulla. Tallensimme teksteistä kaikki kognitiivisten taukojen (vähintään kaksi sekuntia) välissä tuotetut purskeet. Sen jälkeen erottelimme jokaiselta kirjoittajalta kahdesta kolmeen pisintä pursketta ja analysoimme niistä sanastokeskittymät luokittelemalla ne diskursiivisiin, idiosynkraattisiin, kieliopillisiin, leksikaalisiin ja temaattisiin luokkiin (esim. Wray 2002; Forsberg 2006; Rasmussen 2015; Salminen 2020). Näiden sujuvuuspiirteiden osalta vertaamme kolmen osallistujan kirjoitusprofileja suomessa ja ranskassa.

3.2. Aineiston kuvaus

Artikkelin aineistona ovat kolmen yliopisto-opiskelijan kirjoittamat tekstit sekä suomeksi (L1) että ranskaksi (L2). Tapaustutkimukseen valittujen opiskelijoiden P2, P6 ja P17 tuotokset ovat osaa laajempaa tutkimusaineistoa (n = 21), joka kerättiin syksyllä 2021 suomalaisen yliopiston ranskan kielen perusopintojen kurssilla. Nämä kolme opiskelijaa valikoituvat joukosta siksi, että he osallistuivat vapaaehtoisesti myös kirjoittamisesta tehtyyn kielentämiseen eli kirjoittamisen verbalisointihaastatteluun. Tutkimukseen osallistuneet ovat antaneet tietosuoja-asetuksen (GPDR) mukaiset tutkimusluvut, ja kirjoittajat on anonymisoitu. Kirjoittajien taustatiedot on koottu taulukkoon 1.

TAULUKKO 1. *Tapaustutkimuksen taustatiedot (kirjoittajat P2, P6 ja P17)*

Kirjoittaja	P2	P6	P17
Ikä	19	24	24
Äidinkielen ylioppilaskokeen arvosana	C	M	E
Kuinka pitkään opiskellut ranskaa?	6 vuotta	10 vuotta	6 vuotta, tauon jälkeen aloittanut uudelleen
Onko opiskellut ranskan kielisessä maassa?	Ei	Kyllä, yhteensä yli 12 kk	Ei
Oma arvio ranskan kielen taidosta	A2	B1/B2	A1
Vahvin vieras kieli (oma arvio)	englanti	englanti	englanti

Yliopisto-opiskelijat kirjoittivat tietokonehuokassa kaksi tekstiä: toisen suomeksi ja toisen ranskaksi. Molempien tekstien kirjoittamiseen oli aikaa 30 minuuttia. Suomeksi tehtävänä oli kirjoittaa työn ja opiskelun yhdistämisestä: *Mitä mieltä olet töiden tekemisestä opiskelun ohessa? Mitä hyviä ja huonoja puolia siinä on?* Ranskaksi opiskelijat kirjoittivat vaihtoon lähtemisestä ja siihen liittyvistä hyvistä ja huonoista puolista:

Qu'est-ce que vous pensez des études en échange? Quels sont les avantages? Quels sont les inconvénients? Kirjoitustehtävien aiheet liittyvät opiskelijoiden arkeen, jotta heidän olisi mahdollisimman helppo kirjoittaa aiheesta argumentoiva teksti, vaikeivat olisi itse olleet vielä vaihdossa tai työskennelleet opintojen ohella. Aiheita on käytetty myös aiemmissa kirjoittamisen prosessien tutkimuksessa osittain eri kieliyhdistelmällä (ks. Salminen 2020; Mutta & Salminen 2021).

Opiskelijat jaettiin tunnin aikana kahteen ryhmään niin, että toiset aloittivat suomeksi (tapaustutkimuksen P6 ja P17) ja toiset ranskaksi (tapaustutkimuksen P2). Ryhmät vaihtoivat tehtävänantoa puolen tunnin jälkeen. Molemmat kirjoitelmat kirjoitettiin tavallisessa teksti-editori-ikkunassa. Ohjelma tallensi kaiken, mitä editorissa tehtiin näppäimistöllä ja hiirellä. Lisäksi ohjelma tallentaa tauot ja niiden kestot, poistot, lisäykset ym. kirjoitusprosesseihin liittyvät ilmiöt.

4. Monikielisten kirjoittajien purskeet ja sanastokeskittymät kirjoitusprosessissa

4.1. Kirjoitusprosessien pohjalta rakentuvat kirjoittajaprofiilit

Seuraavaksi tarkastelemme kolmen opiskelijan kirjoittajaprofiileja. Taulukkoon 2 on koottu GGXLogiin automaattisesti tallentuneet tiedot kirjoittajien sanamäärästä, kirjoittamiseen käytetystä ajasta sekä kognitiivisten taukojen (vähintään 2 sekuntia) osuuksista suomessa (L1) ja ranskassa (L2). Tiedot kuvaavat kirjoitusprosessien sujuvuutta tässä artikkelissa. On huomioitava, että kirjoitusprosessien sujuvuuteen voi kuulua muitakin mittareita, jotka jäävät tässä analyysin ulkopuolelle (ks. myös Cislaru & Olive 2018; Mutta 2020; Van Waes ym. 2021).

Opiskelija P2 aloitti ranskaksi ja kirjoitti sen jälkeen suomeksi. Hän käytti molempien tekstien kirjoittamiseen koko annetun ajan. Ensikielellä tekstiä syntyi hieman enemmän: suomenkielisen tekstin sanamäärä on 160, ranskankielisen 109. Ranskaksi kirjoittaminen oli aavistuksen hitaampaa kuin suomeksi, mutta molemmilla kielillä kirjoittaja piti

TAULUKKO 2. Kirjoittajien P2, P6 ja P17 sanamäärät ja kognitiivisten taukojen osuudet suomessa (L1) ja ranskassa (L2)

Kirjoittaja	Tekstin kieli	Kirjoittamiseen käytetty aika	Sanamäärä	Kognitiivisten taukojen (vähintään 2 sek) osuus koko kirjoitusajasta
P2	suomi (L1)	30:11 min	160	60,84 %
	ranska (L2)	28:11 min	109	78,26 %
P6	suomi (L1)	26:11 min	347	43,3 %
	ranska (L2)	28:50 min	321	52,03 %
P17	suomi (L1)	24:54 min	473	28,13 %
	ranska (L2)	28:57 min	95	68,87 %

työskentelyajastaan huomattavan osan taukoa: suomenkielisen tekstin kirjoitusajasta kognitiivisten taukojen osuus on noin 61 prosenttia ja ranskankielisen tekstin kirjoitusajasta noin 78 prosenttia. P2:n kirjoittaja-profiili on molemmilla kielillä samantapainen, eli hän pitää kirjoittaessaan paljon suunnittelu- ja mietintätauvoja sekä ensikielillä että vieraalla kielellä kirjoittaessaan. Siksi kirjoittaja on profiiltaan lähinnä fragmentaarinen kirjoittaja, eli hän liikkuu tekstissä edestakaisin muokaten tekstiään paljon (ks. esim. Van Waes & Schellens 2003; Mutta 2020).

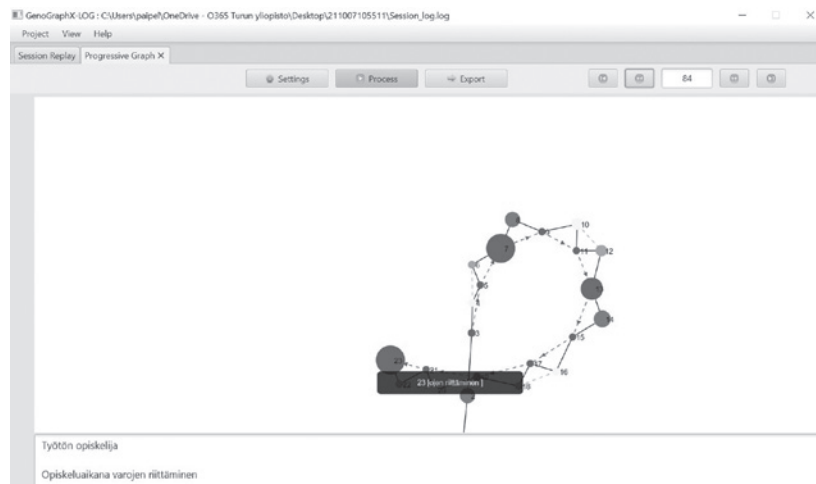
Kirjoittaja P6 aloitti suomenkielisestä tehtävästä ja teki ranskankielisen tekstin toisena tehtävänä (ks. taulukko 1). P6 on varsin tuottelias kirjoittaja molemmilla kielillä: suomenkielisessä tekstissä sanoja on 347 ja ranskankielisessä 321. Hän hallitsee kymmensormijärjestelmän ja tekee hyvin vähän lyöntivirheitä (ks. Laine & Mutta tulossa). P6 kirjoittaa lyhyessä ajassa niin paljon, että molemmissa teksteissä myös kognitiivisten taukojen osuus on pienempi kuin P2:lla: suomenkielisessä tekstissä taukojen osuus on noin 43 prosenttia ja ranskankielisessä noin 52 prosenttia. Kirjoittaja on profiiltaan lähinnä nonstop-kirjoittaja, eli hän pystyy tuottamaan tekstiä yhtäjaksoisesti paljon ja korjaa samalla melko vähän (ks. esim. Van Waes & Schellens 2003; Mutta 2020), ja kymmensormijärjestelmän hallinta näkyy hänen tuotteliaassa tavassaan kirjoittaa tekstiä (Van Waes ym. 2021).

Kolmas kirjoittaja eli P17 aloitti P6:n tapaan suomenkielisestä tekstistä (ks. taulukko 1). Hänen kirjoittajaprofilinsa eroaa suomeksi ja ranskaksi huomattavasti. Suomeksi P17 on sujuva ja nopea kirjoittaja: hän tuottaa puolessa tunnissa 473 sanan tekstin ja pitää kirjoittessaan hyvin vähän kognitiivisia taukoja, vain noin 28 prosenttia koko kirjoitusajasta (ks. myös Mutta 2017). Ranskaksi kirjoittaminen puolestaan on P17:lle vaikeaa. Hän tuottaa puolessa tunnissa 95 sanan tekstin ja pitää ajasta taukoja noin 69 prosenttia. Hänen ranskantaitonsa ei vielä riitä sellaisten kirjoitustaitojen ja -strategioiden käyttöön, joita hän voi käyttää ensikielessä (ks. myös Palviainen ym. 2012). Esimerkiksi kymmensormijärjestelmän käytöstä ei ranskan kirjoittamisessa vielä ole juurikaan hyötyä, vaikka välineenhallinta sujuvoittaa kirjoittamista ensikielellä (Laine & Mutta tulossa). Kirjoittaja on profiililtaan lähinnä nonstop-kirjoittaja ensikielellä ja fragmentaarinen ranskan kielellä (ks. esim. Van Waes & Schellens 2003; Mutta 2020; myös Van Waes ym. 2021).

4.2. Purskeet ja sanastokeskittymien identifiointi

Seuraavaksi tarkastelemme tarkemmin kirjoittajien P2, P6 ja P17 purskeita eli ilman kognitiivista taukoa kirjoitettuja jaksoja suomessa (L1) ja ranskassa (L2). Myös purskeiden osalta kirjoittajat noudattavat edellä profiloitua kuvaa, sillä kaikilla informanteilla purskeet ovat kirjoittajaprofiilien mukaisia. Opiskelija P2 kirjoittaa sekä suomeksi että ranskaksi hyvin lyhyitä purskeita, koska hän pitää molemmilla kielillä kirjoittessaan melko paljon taukoja. Keskimäärin hänellä on suomenkielisessä tekstissä purskeissa noin 11 merkkiä ja ranskankielisessä tekstissä noin 6 merkkiä. Nämä tarkoittavat purskeita, joissa on yhdestä kolmeen sanaa kielestä riippuen. Kuviossa 2 on visualisoitu yksi P2:n esimerkkipurske (15 merkkiä) [*var*]ojen riittäminen.

GGXLog-ohjelmassa kuvataan värien avulla, millaisia toimintoja kirjoittaja tekee. Samoja värejä käytetään myös graafeissa, joita ohjelma tuottaa automaattisesti (ks. kuvio 3). Kuviossa 3 havainnollistetaan

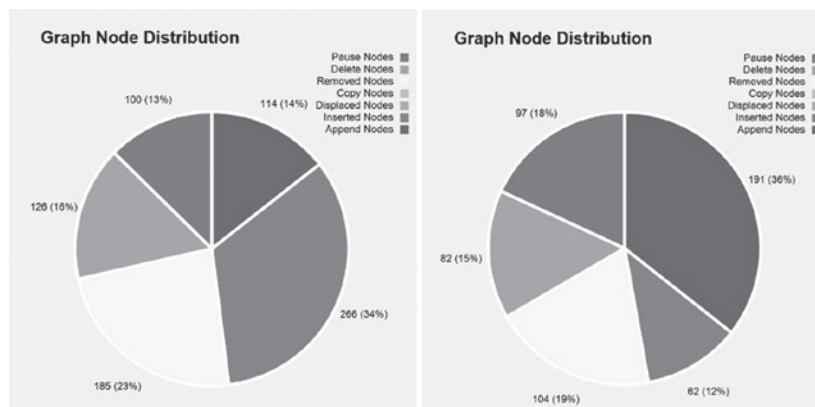


KUVIO 2. Esimerkki kirjoittajan P2 purskeen visualisoinnista GGXLog-ohjelmalla (L1, suomi)

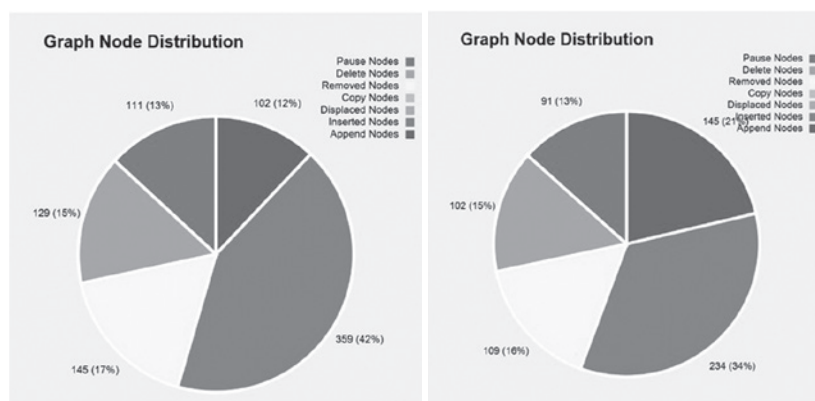
palloja, joita prosessissa kuvataan (lisäykset, poistot, korjaukset ja tauot; ks. myös kuvio 1 yllä). Kuvion prosenttiosuudet ovat suhteessa kuhunkin kirjoitusprosessiin.

Kuviossa 3 näkyy, että P2 käyttää suomeksi kirjoittaessaan eniten tekstin sisälle lisäämistä (266 toimintoa, 34 %). Hän korjaa esimerkiksi lyöntivirheitä ja jatkaa siksi kirjoittamista keskeltä sanaa tai lausetta. Tämä näkyy myös poistamisen (126, 16 %) ja korjaamisen (185, 23 %) osuuksina. Lisäksi P2 kirjoittaa uutta tekstiä loppuun (114, 14 %). Ohjelmassa vihreä väri (tässä kohta 100, 13 %) osoittaa kognitiivisten taukojen osuutta prosessissa. Ranskaksi P2 kirjoittaa hyvin samalla tavalla fragmentaarisesti, mutta tekstin sisälle lisättyjen osuuksien (62, 12 %) määrä on pienempi kuin tekstin loppuun lisättävien (191, 36 %). Lisäksi taukojen osuus on hieman suurempi (97, 18 %) kuin ensikielessä (100, 13 %).

Suomeksi ja ranskaksi kirjoittajaprofiililtaan hyvin samankaltainen kirjoittaja P6 kirjoittaa sekä suomeksi että ranskaksi pitkiä purskeita. Suomeksi purskeet ovat keskimäärin 20 merkkiä ja ranskaksi 11 merkkiä. Kirjoittajan P6 prosessin osat on kuvattu kuviossa 4.



KUVIO 3. GGXLog-ohjelman analyysi kirjoittajan P2 kirjoitusprosessista: vasemmalla L1 (suomi) ja oikealla L2 (ranska)



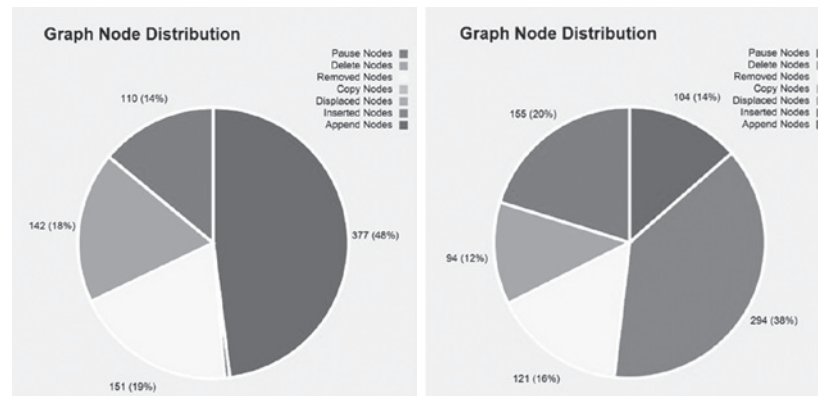
KUVIO 4. GGXLog-ohjelman analyysi kirjoittajan P6 kirjoitusprosessista: vasemmalla L1 (suomi) ja oikealla L2 (ranska)

Olennessin ero on P6:lla siinä, miten hän kirjoittaa uutta tekstiä: ranskaksi teksti näyttäisi syntyvän hieman valmiimpana kuin suomeksi, sillä kirjoittaja lisää uutta tekstiä enemmän loppuun (145 toimintoa, 21 %). Suomeksi suurempi osa uusista lisäyksistä tulee tekstin keskelle (359, 42 %), mikä on osoitus tekstin samanaikaisesta muokkaamisesta.

Molemmissa teksteissä kuitenkin taukojen (13 % ja 13 %) ja korjausten (15 % ja 15 % sekä 17 % ja 16 %) osuudet ovat melko samanlaiset.

Suomeksi ja ranskaksi kirjoittajaprofiilit erosivat toisistaan eniten kirjoittajalla P17. Myös purskeet ovat hänellä suomeksi ja ranskaksi hyvin erilaisia: Kun P17 kirjoittaa suomeksi, hänen purskeensa ovat pitkiä eli keskimäärin 18 merkkiä. Ranskaksi purskeet taas ovat todella lyhyitä eli vain 4,6 merkkiä, mikä tarkoittaa yhden tai kahden sanan purskeita. Suomeksi teksti syntyy hyvin virheettömästi, ja se näkyy kuviossa 5 tekstin loppuun tehtyjen lisäysten suurena osuutena (377 toimintoa, 48 %). Ranskaksi erilainen kirjoitusprosessi näkyy P17:llä myös kuviossa 5: etenkin taukojen (155, 20 %) osuus on suurempi kuin suomeksi, mutta kirjoittaminen on myös enemmän lisäämistä (294, 38 %) kuin suomenkielisessä tekstissä.

Tekstien purskeiden erottamisen jälkeen tarkastelemme purskeiden sisällä olevia sanastokeskittymiä siten, että analysoimme, mitä kunkin kirjoittajien **pisimmät** kirjoitusrupeamat pitävät sisällään (taulukko 3). Jokaiselta on valittu kahdesta kolmeen pisintä pursketta molemmista kielistä. Luokittelu seuraa aiempia luokittelumalleja (ks. yllä; myös esim. Wray 2002; Forsberg 2006; Rasmussen 2015; Salminen 2020).



KUVIO 5. GGXLog-ohjelman analyysi kirjoittajan P17 kirjoitusprosesseista: vasemmalla L1 (suomi) ja oikealla L2 (ranska)

Yhtäjaksoisesti tuotetut purskeet ovat taulukossa lihavoituina (niiden eteen ja jälkeen on merkitty tauon pituus millisekunteina, esim. <2000> = 2 sekuntia), mutta jos samassa kontekstissa on useita purskeita, ne on merkitty alleiviivauksella (ensimmäinen purske) ja *kursiivilla* (toinen purske). Joissain kohdin kontekstia on lyhennetty ja tämä merkitty tekstiin hakasulkeisiin ([---]).

Edellä määrittelimme P2:n kirjoittajaksi, joka pitää pitkiä taukoja ja kirjoittaa molemmissa kielissä melko lyhyitä purskeita. Suomessa purske-esimerkit vahvistavat visualisoinnin havaintoja kirjoittajan profiilista: P2:n visualisoinnista erottuvat pisimmät purskeet ovat melko lyhyitä (ks. taulukko 3) ja rakentuvat esimerkiksi muutaman sanan sanastokeskittymästä ja lauseen predikaatista, kuten purskeessa *perinteinen opiskeljaruokavalio koostuu*. Kirjoittaja käyttää sanastokeskittyminä vakiintuneita lausekerakenteita, kuten nominatiivimuotoisia substantiivilausekkeita (NP), jossa substantiivia edeltää adjektiivimäärite (*perinteinen opiskeljaruokavalio*), tai genetiivin ja sanan *ohella* muodostamaa lauseketta (*työmäärän ohella*).

Lisäksi P2:n sanastokeskittymiä ovat vakiintuneet lausemuotit, kuten *siihen asti että*, sekä verbirakenteet, kuten *ottaa kontolleen*. L2-ranskassa purskeet ovat P2:lla vielä ensikieltä lyhyempiä ja niihin sisältyy L1-suomen tapaan muutamia melko lyhyitä sanastokeskittymiä, kuten omaa mielipidettä ilmaiseva diskursiivinen *je pense* 'ajattelen' tai leksi-kaalinen *en échange* 'vaihdossa', joka voidaan määritellä myös teemaattiseksi, koska ilmaisu liittyy kirjoitelman aiheeseen ja esiintyy suoraan tehtävänannossa. P2:n sanastokeskittymät eivät ole ranskassa vielä täysin automatisoituneita, koska hänellä on prosessissa idiosynkraattisia eli virheellisiä muotoja, joista osan hän on osaa korjata (*pend* > *pense*), mutta joista osa jää lopulliseen tekstiin (**pres*, po. *près* 'lähellä').

P6 kirjoittaa taukojen ja purskeiden keskipituuksien mukaan pitkiä purskeita, ja tämä näkyy myös purske-esimerkeissä (ks. taulukko 3). Suomenkielisessä tekstissä purskeet ovat usein koko lauseen mittaisia tai rakentavat jopa useampilauseisia virkkeitä. Lauseita muodostavissa pitkissä purskeissa on siten mukana useita sanastokeskittymiä samassa

purskeessa. Kirjoittaja hyödyntää sujuvan kirjoittamisen apuna lausekemalleja, esimerkiksi nominatiivimuotoisia NP-lausekkeita, joissa substantiivisia määrittää adjektiivi, kuten *mahdollinen loppuunpalaminen* ja *kultainen keskitie*. Muita toistuvia sanastokeskittymärakenteita ovat erilaiset vakiintuneet verbirakenteet, kuten genetiivisubjekti ja *on hyvä olla* (jokaisen *on hyvä olla*) sekä genetiivisubjekti ja *on tehtävä* -verbiliitto (*opiskelijan on muistettava*).

Myös ranskankielisessä tekstissä on pitkiä, lauseenmittaisia purskeita, joissa on useita sanastokeskittymiä. Ne jakautuvat eri luokkiin, kuten diskursiiviseen luokkaan kuuluvat *c'est* 'se on', joka on tekstuaalinen sanastokeskittymä, ja omaa asemoitumista ilmaiseva *quand même* 'kuitenkin'. Kieliopilliset muodot ovat eri tavoin määrää ilmaisevia ilmauksia, kuten adverbi *un peu* 'vähän' ja (*dans*) *la plupart des cas* 'suurimmassa osassa tapauksista', joka tekstissä on idiosynkraattisessa muodossa **la plupart les cas*. Preposition yhteydessä tämä voitaisiin myös laskea leksikaaliseen kategoriaan, kuten suomenkielinen käännös osoittaa. P6:n kirjoittamat purskeet ja niiden sisältämät sanastokeskittymät osoittavat näiden muotojen olevan hyvin pitkälle automatisoituneita hänen mielessään tai ainakin niiden tuottaminen on hänelle kognitiivisesti helppoa.

Ensikielen ja vieraan kielen kirjoittajaprofili eroaa eniten opiskelijalla P17. Hän kirjoittaa suomeksi keskimääräisen merkkimäärän mukaan pitkiä purskeita. Esimerkkien perusteella hänen purskeensa ovatkin lähes yhtä pitkiä kuin kirjoittajalla P6. Molemmilla purskeet ovat lauseenmittaisia tai jopa lauseita yhdistäviä (ks. taulukko 3) ja rakentuvat useista sanastokeskittymistä. P17:llä tyypillisiä monisanaisia rakenteita ovat esimerkiksi erilaiset vakiintuneet verbirakenteet, kuten **pärjätä arjesta, joutuvat uhraamaan, saattaisivat olla* ja genetiivisubjektin ja verbirakenteen yhdistelmä *minun käy (häntä myös) sääliksi*. Sanastokeskittyminä on myös substantiivilausekkeita (*kolme päivää viikossa*) ja diskursiivisia tekstiä kytkeviä rakenteita (*toisaalta*). Ranskankielisessä tekstissä P17:n purskeet ovat vastaavasti hyvin lyhyitä, ja tekstin pisimpien purskeiden kirjoittamisessa P17 on käyttänyt apuna

TAULUKKO 3. Kirjoittajien P2, P6 ja P17 purske-esimerkit ja niissä esiintyvät sanastokeskittymät L1/L2-kirjoitusprosessissa

Kirjoittaja ja tekstin kieli (L1/L2)	Konteksti ja yhtäjaksoisesti tuotettu purske	Sanastokeskittymät
P2 suomi (L1)	[---] että <5023> kelatut riittävät vielä siihen asti, että uudet kilahtavat <5137> kuun alussa tilille. Muun muassa <3809> perinteinen opiskelijaruokavalio koostuu <3840> nuudeista tai makaronista [---] jaksako <24161> yliopiston työmäärän ohella ottaa kontolleen vielä työpaikan <5249> paineita.	leksikaalinen (Adv.): <i>siihen asti että</i> leksikaalinen (NP/Adj.): <i>perinteinen opiskelijaruokavalio</i> 1) leksikaalinen (+ temaattinen) (N/Gen. + ohella): <i>(yliopiston) työmäärän ohella</i> 2) leksikaalinen (V): <i>ottaa kontolleen</i>
P2 ranska (L2)	<21659> Ce que je pense* <2958> <i>des études en échange</i> <6910> 'Mitä ajattelen opiskelusta vaihdossa' En avion, <3297> je voudrais aller <2321> en échange 'lentokoneella haluaisin mennä vaihtoon' Je voudrais voyager au pays <31985> qui ne sont pas <3009> <i>pres de moi</i> <10321> 'haluaisin matkustaa *maahan jotka eivät ole lähellä minua' <10225> Ketään meistä ei pysty äärimmäisyyksiin ja loppuviimein ylisuorittaminen ja mahdollinen loppuunpalaminen kostautuu, joskus jopa pitkäksi ajaksi. Onko tästä siis hyötyä <5008>	1) diskursiivinen (oma asemituminen): <i>je pense 'ajattelen'</i> 2) leksikaalinen (prep. + N)/temaattinen: <i>en échange 'vaihdossa'</i> leksikaalinen (V): <i>je voudrais aller 'haluaisin mennä'</i> idiosynkraattinen (leksikaalinen, prep. + pron.): <i>*pres de moi 'lähellä minua'</i> 1) leksikaalinen (V): <i>ei pysty</i> 2) leksikaalinen (N/Adv.): <i>loppuviimein</i> 3) leksikaalinen (NP/Adj.): <i>mahdollinen loppuunpalaminen</i> 4) leksikaalinen (V+N): <i>onko hyötyä</i>
P6 suomi (L1)	<6721> Vanhän viisauden ja ohjeen mukaan, kannattaa muistaa säilyttää kultainen keskitie, jotta jokaisen on hyvä olla <2336>.	1) leksikaalinen (N/Gen. + mukaan): <i>vanhan viisauden [---] mukaan</i> 2) leksikaalinen (V): <i>kannattaa muistaa</i> 3) leksikaalinen (NP/Adj.): <i>kultainen keskitie</i> 4) leksikaalinen (Pron./Gen. + V/Adj.): <i>jokaisen on hyvä olla</i>

	<p><12448> Opiskelijan on ensimmäisenä muistettava oma hyvinvointi ja jaksaminen. Ihminen tarvitsee <2144> aktiivista tekemistä, [---]</p> <p><2240> C'est un peu un cliché mais dans plupart les cas, c'est la réalité qu' <6641> 'Se on hieman klisee mutta suurimmassa osassa tapauksista, on totta että'</p>	<p>1) leksikaalinen (NP/Gen. + V): <i>opiskelijan on muistettava</i></p> <p>2) leksikaalinen (NP/Adj.): <i>oma hyvinvointi ja jaksaminen</i></p> <p>1) diskursiivinen (tekstuaalinen): <i>c'est se on'</i></p> <p>2) kieliopillinen: <i>un peu 'hieman'</i></p> <p>3) idiosynkraattinen (kieliopillinen, prep. + NP/Gen.): <i>* dans la plupart les cas 'suurimmassa osassa tapauksista'</i></p>
P6 ranska (L2)	<p><3376> Je voulais quand même apprendre la langue française <3575> <i>et je me suis inscrit pour aller en échange. Au début, c'était</i> <5617> 'Halusin kuitenkin oppia ranskan kieltä ja ilmoittauduin vaihtoon lähtöön. Aluksi se oli'</p> <p>Onko heidän mahdollista <2351> päriätä arjesta <3600> heidän aikataululla? Kuinka paljon he joutuvat uhraamaan sen edestä? [---] <2304></p>	<p>1) diskursiivinen (oma asemoituminen): <i>quand même 'kuitenkin'</i></p> <p>2) leksikaalinen (V/NP): <i>apprendre la langue française 'oppia ranskan kieltä'</i></p> <p>3) leksikaalinen (Prep. + V + Prep./N) / temaattinen: <i>pour aller en échange 'vaihtoon lähtöön'</i></p> <p>4) idiosynkraattinen (leksikaalinen, Prep. + N): <i>* au début 'aluksi'</i></p> <p>5) diskursiivinen (tekstuaalinen): <i>c'était 'se oli'</i></p> <p>1) idiosynkraattinen (leksikaalinen, V/NP): <i>* päriätä arjesta</i></p> <p>2) kieliopillinen (Adv.): <i>kuinka paljon</i></p> <p>3) leksikaalinen (V): <i>joutuvat uhraamaan</i></p>
P17 suomi (L1)	<p><2801> Koulua minulla oli tällöin noin kolme päivää viikossa, joten sain työnteon <3041> sisällytettyä aikatauluuni ilman ongelmia.</p> <p><4769> Toisaalta minun käy heitä myös sääliksi, vaikka he itse saattaisivat olla <2416> onnellisia tilanteessaan.</p>	<p>1) leksikaalinen (V/omistuslause+ NP)/temaattinen: <i>koulua minulla oli</i></p> <p>2) leksikaalinen (N/Adv.): <i>kolme päivää viikossa</i></p> <p>1) diskursiivinen (tekstuaalinen): <i>toisaalta</i></p> <p>2) leksikaalinen (NP/Gen. + V/NP): <i>minun käy sääliksi</i></p> <p>3) leksikaalinen (V): <i>saattaisivat olla</i></p>
P17 ranska (L2)	<p><27849> Quels sont les avantages? <2065> 'Mitkä ovat edut?'</p> <p><2065> Et quels sont les inconvénients? <5912> 'Ja mitkä ovat haitat?'</p> <p><2496> Je ne sais pas <7104> je veux étudier en échange une jour. 'En tiedä haluaisin opiskella vaihdossa jonakin päivänä.'</p>	<p>Temaattinen, kopioitu tehtävänannosta: <i>Quels sont les avantages?</i></p> <p>Temaattinen, kopioitu tehtävänannosta: <i>Et quels sont les inconvénients?</i></p> <p>Diskursiivinen (oma asemoituminen): <i>je ne sais pas 'en tiedä'</i></p>

* Kirjoittaa tässä kohdin **Ce que je pend* (luokiteltu idiosynkraattinen), mutta korjaa myöhemmin oikean kirjoitusasun.

tehtävänannon kopioimista. Näiden temaattisten ilmaisujen lisäksi hänellä on yksittäisiä diskursiivisia oman mielipiteen ilmaisuja, kuten *je ne sais pas* ‘en tiedä’. Hänen ranskan kielen kirjoitustaitonsa ei ole vielä automatisoituneella tasolla.

6. Lopuksi

Artikkelimme esittelee KISUVI-projektiin kuuluvaa tapaustutkimusta. Tarkastelimme kolmen yliopisto-opiskelijan kirjoitusprosessien sujuvuutta purskeiden ja sanastokeskittymien avulla. Näiden määrittelyssä käytimme apuna GGXLog-näppäilyntallennusohjelman visualisointityökaluja. Olemme pyrkineet vertaamaan jokaisen kirjoittajan ensikielen (L1-suomi) ja vieraan kielen (L2-ranska) visualisointeja ja niiden rakentamaa kuvaa kirjoittajan tuottamista purskeista ja purskeissa esiintyvistä sanastokeskittymistä sekä näiden piirteiden vaikutuksesta sujuvuuteen. Muodostimme näiden sujuvuuspiirteiden pohjalta heidän kirjoittajaprofilinsa. Sujuva, tauoton kirjoittajaprofiili vertautuu aiempiin tutkimustuloksiin (profiileista ks. esim. Van Waes & Schellens 2003; Mutta 2020).

Tutkimustulokset ovat saman suuntaisia kuin aiemmissa tutkimuksissa myös siten, kirjoittajat ovat sujuvampia ensi- kuin vieraalla kielellä kirjoitettaessa, mutta tässä on paljon taitotasoon liittyvää variaatiota (vrt. Palviainen ym. 2012). Myös tottumus kirjoittaa koneella sujuvoitti kirjoittamista (vrt. Van Waes ym. 2021). Aiemmissä tutkimuksissa L2-tuotoksia on usein vertailtu natiivi kielen puhujien tapaan tuottaa kieltä (Lintunen ym. 2020), mutta tässä tutkimuksessa verrattiin saman kirjoittajan kirjoittamista L1- ja L2-kielellä. Myös kielipari suomi–ranska poikkeaa useista aiemmista tutkimuksista, joissa L2-englanti on yleinen tutkittu vieras kieli. Sujuvuutta on monesti tutkittu vain joko tuotoksesta (esim. tekstin sidonnaisuuden kannalta; vrt. Segalowitzin 2010 *perceived fluency*) tai prosessista (esim. kirjoitusnopeus, engl. *mean transition times*, Kowal 2014), mutta GGX-Log-ohjelman avulla voidaan tutkia näitä molempia.

Näppäilyntallennuksen avulla kirjoittajien tapaa kirjoittaa pystyttiin kuvaamaan niin, että erot ja yhtäläisyydet suomeksi ja ranskaksi kirjoittamisessa tulivat näkyviksi. Visualisointien luomien kirjoittajaprofilien pohjalta oli nähtävissä, että kielitaito vaikutti kirjoitusprosessin sujuvuuteen etenkin ranskassa (L2). Molemmissa kielissä sujuva kirjoittaja (P6) pystyi hyödyntämään samankaltaisia kirjoittamisstrategioita sekä ensikielessä että vieraassa kielessä. Vastaavasti heikko kielitaito esti ensikielessä hyväksi havaittujen kirjoitusstrategioiden hyödyntämistä täysipainoisesti ja haittasi ylipäätään sujuvaa kirjoittamista kirjoittaja P17:lla. Samansuuntaisia tuloksia on löydetty aiemmista tutkimuksista (esim. Palviainen ym. 2012).

Sujuva kirjoittaja (P6) tuotti pitkiä purskeita sekä ensikielellä että vieraalla kielellä ja pystyi hyödyntämään kirjoittamisen apuna rakenne-malleja, joiden tuottaminen oli siinä mielessä automaattista, ettei kirjoittamiseen tullut purskeiden sisällä kognitiivisia taukoja. Sanastokeskittymien tarkka analyysi vaatii kuitenkin laajemman aineiston analyysiä. Nyt purskeissa korostuivat tietynlaiset rakenteet, mutta niiden esiintyminen muilla kirjoittajilla jäi vielä tässä tutkimuksessa selvittämättä. Sanastokeskittymän määritelmän mukaan osa löydetystä monisanaisista ilmauksista voi olla siis oppijakohtaisia, hänelle ominaisia ilmauksia ja osa kielikohtaisia (ks. esim. Forsberg 2006, 2008; Myles & Cordier 2017; Siyanova-Chanturia & Pellicer-Sánchez 2019). Purskeiden rajaaminen kahden sekunnin kognitiivisen tauon sisään kuitenkin kuvasi kirjoittajien tekstin tuottamista ja tuottamisen eroja hyvin. Lisätutkimusta tarvitaan suomen kaltaisten morfeemirikkaiden kielten sanastokeskittymien määrittelyä. Esimerkiksi *koulussa* ja *repussa* -tyyppiset ilmaisut voivat olla suomi toisena kielenä -oppijalla (S2) automatisoituneita holistisesti prosessoituja ja tuotettuja ilmaisuja, kun ensikielellä ne saatetaan analysoida taivutusmuotoina.

GGXLog-näppäilyntallennusohjelma ja sen tekemä kirjoitusprosessien tallennus sekä ohjelmasta automaattisesti saatavat tilastot ja visualisoinnit tarjoavat kirjoitusprosessista runsaasti sellaista tietoa, jota voidaan hyödyntää kielenoppimisen tutkimuksessa. Ohjelman

tuottamat tilastot pohjautuvat tiettyihin kriteereihin ja määriteltyihin oletuksiin, jotka on hyvä ottaa huomioon ja mainita metodin ja tulosten esittelyssä, jotta muiden tutkijoiden on mahdollista verrata tuloksia omiin tutkimustuloksiinsa.

Ohjelmaa voidaan kuitenkin käyttää myös opetuksen apuna. Teknologian käyttö kirjoittamisessa on muuttanut sekä itse kirjoittamista että prosessia ja sen suunnittelua jo aika paljon. Meillä ei kuitenkaan ole kovin tarkkaa kuvaa siitä, miten esimerkiksi itse kirjoitamme tietokoneella ja käytämmekö kirjoittamisen apuna esimerkiksi prosessikirjoittamisen ohjeita suunnittelusta tai tekstin jäsentämisestä. Kirjoittamisen tallentaminen antaa suunnittelun roolista hyödyllistä tietoa sekä tutkijoille että opettajille.

Olemme testanneet oman kirjoitusprosessin kielentämistä tässä artikkelissa analysoitujen kolmen opiskelijan (ks. Laine & Mutta tulossa) lisäksi esimerkiksi opettajan työkaluna yläkoulussa (Laine 2019). Sekä yläkoululaiset että yliopisto-opiskelijat ovat olleet motivoituneita katsomaan omasta kirjoitusprosessista tehtyä tallennetta. Opettaja voi-kin käyttää visualisointeja apuna antaessaan palautetta kielenoppijalle ja eriyttäessään kirjoittamisen ja kielen rakenteiden opetusta kunkin oppijan tarpeiden ja taitotason mukaisesti. Opettaja pystyy suuntaamaan palautetta esimerkiksi rakenteisiin, jotka juuri sillä hetkellä ovat vahvistumassa tai kehittymässä automaattisiksi sanastokeskittymiksi. Ongelmallisten kohtien pohdinta yhdessä auttaa myös opettajaa tiedostamaan, minkälaisia ovat kirjoituksen esteet, mitä oppija pohtii ja miten häntä voi tukea kohti sujuvaa kirjoittamista.

Lopuksi voidaan todeta, että yleisesti tunnetut, aiemmissa tutkimuksissa käytetyt luokittelut (esim. Myles & Cordier 2017; Wray 2002) näkyvät tutkimuksessa deduktiivisena lähestymistapana, varsinkin lopputuotosten tasolla, mutta kirjoitusprosessien purskeiden tutkimus GGXLog-ohjelman analysoinnin avulla mahdollistaa induktiivisen lähestymistavan. Tarvitaan toki tätä artikkelia suurempi aineisto, jotta voidaan nähdä, millaista uutta luokittelutapaa voitaisiin ehdottaa typologisesti erilaisiin kieliin. Purskeet ja sanastokeskittymät ovat yksi avain

sujuvaan kirjoittamiseen, ja visualisoinnin avulla pääsemme lähemmäksi näiden ilmiöiden ymmärtämistä.

Lähteet

- Bécotte-Boutin, Hélène-Sarah, Gilles Caporossi, Christophe Leblay, Alain Hertz 2019. Writing and rewriting: Keystroke logging's colored numerical visualization. – Eva Lindgren, Kirk P. H. Sullivan (eds.). *Observing Writing: Insights from Keystroke Logging and Handwriting*. Leiden: Brill, 96–124. https://doi.org/10.1163/9789004392526_006
- Chenoweth, N. Ann, John R. Hayes 2001. Fluency in writing: generating text in L1 and L2. – *Written Communication* 18 (1), 80–98. <https://doi.org/10.1177/0741088301018001004>
- Christiansen, Morten H., Inbal Arnon 2017. More than words: The role of multiword sequences in language learning and use. – *Cognitive Sciences* 9 (3), 542–551. <https://doi.org/10.1111/tops.12274>
- Cislaru, Georgeta, Thierry Olive 2018. Le processus de textualisation: Analyse des unités linguistiques de performance écrite. Louvain-la-Neuve: De Boeck Supérieur. <https://doi.org/10.3917/dbu.cisla.2018.01>
- Cordier, Caroline 2013. The Presence, Nature and Role of Formulaic Sequences in English Advanced Learners of French: A Longitudinal Study. Newcastle: Newcastle University. <http://hdl.handle.net/10443/2265>
- Durrant, Philip 2013. Formulaicity in an agglutinating language: The case of Turkish. – *Corpus Linguistics and Linguistic Theory* 9 (1), 1–38. <https://doi.org/10.1515/cllt-2013-0009>
- Forsberg, Fanny 2006. *Le langage préfabriqué en français parlé L2: étude acquisitionnelle et comparative*. Stockholm: Université de Stockholm.
- ISK = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen, Irja Alho 2004. *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seuran toimituksia 950. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Jantunen, Jarmo Harri 2009. Minulla on aivan paljon rahaa. Fraseologiset yksiköt suomen kielen opetuksessa. – *Virittäjä* 113 (3), 356–381.
- Jantunen, Jarmo Harri 2011. Avainsana-analyysi annotoidun oppijankieliaineiston tutkimisessa: alustavia havaintoja. – *AFinLA-e: Soveltavan kielitieteen tutkimuksia* 3, 48–61.
- Jaworska, Sylvia, Cédric Krummes, Astrid Ensslin 2015. Formulaic sequences in native and non-native argumentative writing in German. – *International Journal of Corpus Linguistics* 20 (4), 500–525. <https://doi.org/10.1075/ijcl.20.4.04jaw>

- Kowal, Iwona 2014. Fluency in second language writing: A developmental perspective. – *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 131, 229–246.
- Laine, Päivi 2019. Nuoret ovat taitavia analysoimaan omaa kirjoittamistaan. <http://www.suomenkielenseura.fi/fi/nuoret-ovat-taitavia-analysoimaan-omaa-kirjoittamistaan> (13.10.2022).
- Laine, Päivi, Maarit Mutta (tulossa). Kirjoitusprosessin kielentäminen pedagogisena välineenä. – *Ainedidaktisia tutkimuksia* 21.
- Leblay, Christophe, Gilles Caporossi 2015. A graph theory approach to online writing data visualization. – Georgeta Cislaru (eds.). *Writing(s) at the Crossroads: The Process-Product Interface*. Amsterdam: John Benjamin, 171–181. <https://doi.org/10.1075/z.194.09leb>
- Leblay, Christophe, Hakim Usoof, Gilles Caporossi 2022. Kirjoittamisprosessin dynaaminen visualisointi. – *Kieli, koulutus ja yhteiskunta* 13 (3). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2022/kirjoittamisprosessin-dynaaminen-visualisointi> (13.10.2022).
- Lieri, Kristiina 2021. *Studies of Multilingual Finnish-Norwegian Adolescents' Mastering of Multi-word Units*. Turun yliopiston julkaisuja. *Annales Universitatis Turkuensis*, sarja B, osa 553. Turku: Turun yliopisto. <http://urn.fi/URN:ISBN:978-951-29-8602-6>
- Lindgren, Eva, Kirk P. H. Sullivan (eds.) 2019. *Observing Writing: Insights from Keystroke Logging and Handwriting*. *Studies in Writing* 38. Leiden: Brill. <https://doi.org/10.1163/9789004392526>
- Lintunen, Pekka, Maarit Mutta, Pauliina Peltonen (eds.) 2020. *Fluency in L2 Learning and Use*. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781788926317>
- Lintunen, Pekka, Maarit Mutta, Sanna Pelttari 2017. Profiling language learners in hybrid learning contexts: Learners' perceptions. – *The EuroCALL Review* 25 (1), 61–75. <https://doi.org/10.4995/eurocall.2017.7145>
- Martin, Maisa 2016. Monikielisyys muutoksessa. – *Kieli, koulutus ja yhteiskunta* 7 (5). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-lokakuu-2016/monikielisyys-muutoksessa> (13.10.2022).
- Mutta, Maarit 2017. Pausal behavior in the writing processes of foreign and native language writers: The importance of defining the individual pause length. – Sylvie Plane et al. (eds.). *Research on Writing: Multiple Perspectives. International Exchanges on the Study of Writing*. Fort Collins, Colorado: The WAC Clearinghouse and CREM, 511–529. <https://doi.org/10.37514/INT-B.2017.0919.2.28>

- Mutta, Maarit 2020. L2 fluency and writer profiles. – Pekka Lintunen, Maarit Mutta, Pauliina Peltonen (eds.). *Fluency in L2 Learning and Use*. Bristol: *Multilingual Matters*, 63–80.
- Mutta, Maarit, Susanna Salminen 2021. Les séquences préfabriquées dans la production écrite dans le cas de scripteurs finnophones de français et suédois L2. – *Synergies pays riverains de la Baltique* 14/20, 11–26.
- Myles, Florence, Caroline Cordier 2017. Formulaic sequences (FS) cannot be an umbrella term in SLA: Focusing on psycholinguistic FSs and their identification. – *Studies in Second Language Acquisition* 39 (1), 3–28. <https://doi.org/10.1017/S027226311600036X>
- Mäntylä, Katja, Sinikka Lahtinen, Outi Toropainen 2020. Processing L2 vocabulary in writing. – Paul Booth, Jon Clenton (eds.). *First Language Influences on Multilingual Lexicons*. New York: Routledge, 107–126. <https://doi.org/10.4324/9780429031410-9>
- Palviainen, Åsa, Paula Kalaja, Katja Mäntylä 2012. Development of L2 writing: fluency and proficiency. – *AFinLA-e: Soveltavan kielitieteen tutkimuksia* 4, 47–59.
- Perez-Bettan, Annie 2015. *Apprentissage et utilisation du langage préfabriqué chez des apprenants de français langue étrangère*. Paris: Université Vincennes Saint-Denis. <https://tel.archives-ouvertes.fr/tel-01714531/document> (13.10.2022).
- Prentice, Julia 2010. På rak sak. Om ordförbindelser och konventionaliserade uttryck bland unga språkbrukare i flerspråkiga miljöer. Göteborg: University of Gothenburg.
- Prentice, Julia, Emma Sköldberg 2010. Klättra på väggarna eller bara vara ett med soffan? Om figurativa ordförbindelser bland ungdomar i flerspråkiga miljöer. – *Språk och stil* 20, 5–35.
- Rasmussen, Laura 2015. *L'usage des séquences préfabriquées dans le langage parlé d'étudiants finnophones du français L2 au niveau universitaires: Une étude de cas*. *Julkaisematton pro gradu -tutkielma*. Turku: Turun yliopisto.
- Reiman, Nina 2011. *Transitiivikonstruktio ikkunana syntaksin kehitykseen: Infinitiiviset rakenteet ja passiivi taidon indikaattoreina S2-oppijoiden teksteissä*. – *AFinLA-e: Soveltavan kielitieteen tutkimuksia* 3, 142–157.
- Salminen, Susanna 2020. *Les séquences préfabriquées dans les productions écrites des étudiants du français L2 et du suédois L2*. *Julkaisematton pro gradu -tutkielma*. Turku: Turun yliopisto. <https://www.utupub.fi/handle/10024/149713>

- Schmitt, Norbert (ed.) 2004. *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.9>
- Segalowitz, Norman 2010. *The Cognitive Bases of Second Language Fluency*. New York: Routledge. <https://doi.org/10.4324/9780203851357>
- Siyanova-Chanturia, Anna, Ana Pellicer-Sánchez 2019. Formulaic language: Setting and scene. – Anna Siyanova-Chanturia, Ana Pellicer-Sánchez (eds.). *Understanding Formulaic Language*. New York: Routledge, 1–15. <https://doi.org/10.4324/9781315206615-1>
- Snellings, Patrick, Amos van Gelderen, Kees de Glopper 2002. Lexical retrieval: An aspect of fluent second language production that can be enhanced. – *Language Learning* 52 (4), 723–754. <https://doi.org/10.1111/1467-9922.00202>
- Spelman Miller, Kristyan 2006. Pausing, productivity and the processing of topic in online writing. – Kirk P. H. Sullivan, Eva Lindgren (eds.). *Computer Key-Stroke Logging and Writing*. *Studies in Writing* 18. Leiden: Brill, 131–156. https://doi.org/10.1163/9780080460932_009
- Spelman Miller, Kristyan, Eva Lindgren, Kirk P. H. Sullivan 2008. The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. – *TESOL Quarterly* 42 (3), 433–454. <https://doi.org/10.1002/j.1545-7249.2008.tb00140.x>
- Sundman, Marketta 2011. Helsekvenser i finska grundskoleelevers skrivande. – *Svenskans beskrivning* 31, 327–336.
- Toropainen, Outi, Sinikka Lahtinen, Anne-Maj Åberg 2020. Metaspråklig kunskap i andraspråksskribenters reflektioner. – *Nordand* 15 (2), 111–126. <https://doi.org/10.18261/issn.2535-3381-2020-02-05>
- Usoof, Hakim, Christophe Leblay, Gilles Caporossi 2020. *GenoGraphiX-Log Version 2.0 User Guide*. – *Les Cahiers Du GERAD G-2020-68*, 1–63.
- Van Waes, Luuk, Mariëlle Leijten, Jens Roeser, Thierry Olive, Joachim Grabowski 2021. Measuring and assessing typing skills in writing research. – *Journal of Writing Research* 13 (1), 107–153. <https://doi.org/10.17239/jowr-2021.13.01.04>
- Van Waes, Luuk, Peter Jan Schellens 2003. Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. – *Journal of Pragmatics* 35 (6), 829–853. [https://doi.org/10.1016/S0378-2166\(02\)00121-2](https://doi.org/10.1016/S0378-2166(02)00121-2)
- Wray, Alison 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511519772>
- Wray, Alison 2008. *Formulaic Sequences: Pushing Boundaries*. Oxford: Oxford University Press.

A case study of multilingual writers' writing processes: Using keystroke logging to identify formulaic sequences

MAARIT MUTTA, PÄIVI LAINE

University of Turku

This article deals with the fluency of multilingual writers' writing processes. To collect and analyze the learner-corpus of our case study, we used the graph theory-based keystroke logging software GGXLog, whose visualization functions provide a new approach to the analysis of writing process fluency. The material consists of three writings in Finnish (L1) and French (L2) written by university students. In this article, we focus on the writing processes and profiles of these three students, especially regarding their bursts and formulaic sequences during the writing process. The results show that the statistics and visualization allow us to identify and illustrate bursts of fluent writing in a more detailed way. Furthermore, sequences of fluent writing can reveal the patterns of individual writer profiles, which increases the knowledge of L1 and L2 writing processes. The results in L2 writing are closely related to students' proficiency level. These results can contribute to enhancing pedagogical practices in different educational settings.

Keywords: writer profiles; keystroke logging; fluency; visualisation; pedagogical tool; language teaching

Maarit Mutta

Turun yliopiston kieli- ja käännöstieteiden laitos
FI-20014 University of Turku, Finland
maamut@utu.fi

Päivi Laine

Turun yliopiston kieli- ja käännöstieteiden laitos
FI-20014 University of Turku, Finland
paipel@utu.fi