

Teija Alasalmi

Ilmaiset konkordanssiohjelmat Windows-ympäristössä

Tietotekniikan kandidaatintutkielma

16. tammikuuta 2013

Jyväskylän yliopisto

Tietotekniikan laitos

Tekijä: Teija Alasalmi

Yhteystiedot: teikku@jyu.fi

Ohjaaja: Jaana Markkanen

Työn nimi: Ilmaiset konkordanssiohjelmat Windows-ympäristössä

Title in English: Free Concordance Programs for Windows

Työ: Kandidaatintutkielma

Suuntautumisvaihtoehto: Opettajankoulutus

Sivumäärä: 37+1

Tiivistelmä: Tämän tutkimuksen tavoitteena on kartoittaa verkossa saatavilla olevia ilmaisia, yksikielisiä konkordanssiohjelmiä Windows-ympäristössä. Lisäksi tavoitteena on selvittää, millaisia toimintoja ilmaiset konkordanssiohjelmat sisältävät, sekä lähdekirjallisuuden perehtymällä selvittää konkordanssiohjelmien hyödyntämismahdollisuuksia kääntämisessä.

Avainsanat: konkordanssiohjelma, konkordanssi, kollokaatio, frekvenssilista, sanastollisen rikkauden indeksi, korpus, kääntäminen

Abstract: The objective of the present study is to search the Internet for free, monolingual concordance programs that are available for Windows platform. Another aim is to find out what kinds of functionalities these concordance programs include and by using relevant literature, clarify how concordance programs can be utilized in translation.

Keywords: concordance program, concordance, collocation, frequency list, type-token ratio, corpus, translation

Termiluettelo

AAVISTUSTERMI	Kääntäjän oman intuition perusteella muodostunut ehdotus sanan tai ilmauksen käännökseksi.
KLUSTERI	Yleensä vähintään kolmen toistuvasti yhdessä esiintyvän sanan muodostava jono.
KOLLOKAATIO	Sanan merkitykseen perustuvaa taipumusta esiintyä yhdessä tietynlaisten sanojen kanssa, esimerkiksi vakiintuneena ilmauksena.
KOLLOKAATTI	Vierussana, yleensä kahden lähekkäin olevan sanan (tilastollisesti merkitsevä) yhteisesiintymä.
KONKORDANSSI	Tietokoneohjelmalla tehty luettelo sananmuotojen esiintymistä.
KONKORDANSSI-OHJELMA	Tietokoneohjelma, jonka avulla voidaan tarkastella valitun hakusanan tai sanojen esiintymiä tekstiyhteydessään käsiteltävänä olevassa korpuksessa.
KONTEKSTI	Teksti- tai lauseyhteys.
KORPUS	Sähköisessä muodossa oleva kielentutkimukseen käytettävä tekstiaineisto.
KOTEKSTI	Ilmauksen tekstuaalinen lähiympäristö.
KWIC	Keyword in context, avainsanahaku.

Sisältö

1	JOHDANTO	1
2	KORPUSLINGVISTINEN KÄÄNNÖSTUTKIMUS	2
2.1	Korpuslingvistiikka ja käännöstiede	2
2.2	Korpuksen määritelmä	3
2.3	Tietokoneen edut ja haitat korpusten käsittelyssä.....	4
2.4	Korpusten hyödyntäminen kääntämisessä.....	5
3	KONKORDANSSIOHJELMAT	8
3.1	Konkordanssiohjelman määritelmä ja jaottelu	8
3.2	Konkordanssiohjelmien yleiset toiminnot	10
3.2.1	Konkordanssit	10
3.2.2	Kollokaatit	12
3.2.3	Frekvenssilistat	13
3.2.4	Statistiikka	14
3.3	Hyvän konkordanssiohjelman tunnusmerkit	16
4	ILMAISTEN KONKORDANSSIOHJELMIEN KARTOITUS	18
4.1	Aiemmat tutkimukset	18
4.2	Tutkimuksen tavoitteet	19
4.3	Tutkimusongelma ja tutkimuskysymykset	20
4.4	Tutkimusmenetelmä	20
4.5	Tutkimusaineisto	21
4.5.1	aConCorde 0.4.2	22
4.5.2	AdTAT 1.5	22
4.5.3	AntConc 3.2.4.....	23
4.5.4	KWIC Concordance for Windows 5.0	23
4.5.5	Multilingual Corpus Toolkit.....	24
4.5.6	Simple Concordance Program 4.09	24
4.5.7	TextSTAT 2.9	25
4.5.8	WConcord 3.0.....	25
5	HAVAINTOJA ILMAISISTA KONKORDANSSIOHJELMISTA	26
6	YHTEENVETO	29
	LÄHTEET	32
	LIITTEET	38
A	Ilmaisten konkordanssiohjelmien tiedot ja toiminnot	38

1 Johdanto

Nykypäivän informaatioyhteiskunnassa tarve erilaisten tekstien kääntämiselle on kasvanut huomattavasti. Muun muassa erilaisia dokumentteja, käyttöohjeita, sopimuksia, kirjoja, lehtiä, www-sivuja ja puheita käännetään päivittäin kieleltä toiselle. Lisäksi suuri osa käännettävästä aineistosta luodaan, tallennetaan ja säilytetään nykyään yhä enenevässä määrin elektronisessa muodossa, ja siksi ammattikäntäjien tulee hallita elektronisten kieliaineistojen prosessointi erilaisten tietoteknisten työkalujen avulla.

Laajojen, elektronisten kieliaineistojen eli niin sanottujen korpusten prosessointi on kuitenkin helpottunut huomattavasti, sillä tietokoneet ovat entistä helpommin saatavilla, ja ne pystyvät prosessoimaan dataa tehokkaammin kuin aikaisemmin. Lisäksi erityisesti viime vuosikymmeninä erilaisia korpustyökaluja on kehitetty elektronisten kieliaineistojen käyttöä ja analysointia varten. Tämän tutkielman tarkoituksena on kartoittaa ilmaisia, kieliaineiston analysointiin tarkoitettuja ja Windows-ympäristöön suunnattuja yksikielisiä konkordanssiohjelmia sekä niiden yleisiä ominaisuuksia erityisesti käntäjän näkökulmasta. Ilmaisia konkordanssiohjelmia on saatavilla useita, eikä suomenkielisessä kirjallisuudessa ole aiemmin kattavasti esitelty konkordanssiohjelmien toimintaa, niiden yleisiä ominaisuuksia tai kuvattu niiden mahdollisia käyttötapoja käntäjän näkökulmasta.

Tämän tutkielman toisessa luvussa perehdytään korpuslingvistiseen käänöstutkimukseen, ja määritellään korpuksen käsite. Lisäksi luvussa 2 käsitellään tietokoneen käytön etuja ja haittoja korpusten käsittelyssä, sekä kuvataan korpusten hyödyntämistä vieraan kielen kääntämisessä. Luvussa 3 määritellään konkordanssiohjelman käsite, sekä kuvataan konkordanssiohjelmien jaottelutavat ja konkordanssiohjelmien yleiset toiminnot. Lisäksi luvussa 3 määritellään hyvän konkordanssiohjelman kriteerit. Luvussa 4 esitellään ensin lyhyesti aiempia konkordanssiohjelmia käsitelleitä tutkimuksia, ja sen jälkeen kuvataan tämän tutkimuksen tavoitteet, tutkimusongelma ja -kysymykset sekä tutkimusmenetelmä. Lisäksi esitellään tutkimusaineistona toimivat Windows-ympäristöön suunnatut ilmaiset, yksikieliset konkordanssiohjelmat. Luvussa 5 esitellään ilmaisten konkordanssiohjelmien toiminnoista tehtyjä havaintoja. Lopuksi luvussa 6 tehdään yhteenveto tämän tutkimuksen tuloksista, sekä esitetään joitakin jatkotutkimusideoita.

2 Korpuslingvistinen käännöstutkimus

Tässä luvussa kuvataan lyhyesti korpuslingvististä käännöstutkimusta. Lisäksi luvussa määritellään korpuksen käsite, sekä pohditaan tietokoneen tuomia etuja ja haittoja korpuksen käsittelyssä. Lopuksi kuvataan sitä, miten korpuksia voidaan hyödyntää kääntämisessä.

2.1 Korpuslingvistiikka ja käännöstiede

Yhden näkökulman mukaan **korpuslingvistiikka** (engl. *Corpus Linguistics*) on eräs kielitieteen haara, joka tutkii kieltä ja sen käyttöä korpusten avulla (Kenny 2001: 23). Useimmat tutkijat ovat kuitenkin sitä mieltä, että kielitieteen osa-alueen sijaan korpuslingvistiikka on ennen kaikkea menetelmä, jota voidaan käyttää melkein millä tahansa alalla (Bowker & Pearson 2002: 9; McEnery et al. 2006: 7). **Käännöstiede** (engl. *Translation Studies*) puolestaan on tieteenala, joka tutkii kääntämistä. Se kattaa monia tutkimuskohteita, kuten esimerkiksi kirjallisuuden kääntäminen, tekninen kääntäminen, tulkkaus tai audiovisuaalinen kääntäminen (sis. jälkiäänitys eli dubbaus ja elokuvien ja tv-sarjojen tekstitys). Lisäksi käännöstiede tutkii pedagogiikkaan liittyviä näkökulmia, kuten esimerkiksi kääntäjien koulutusta ja käännösten arviointia. Käännöstiede tutkii myös itseään eli käännöstieteessä käytettyjä menetelmiä, malleja ja teorioita. (Baker 1998: 227–279.)

Korpustutkimuksen ja käännöstutkimuksen integraatio tapahtui 1990-luvun alussa, kun korpuslingvistiikka ja käännöstiede yhdistyivät käännösuniversaalien eli lähde- ja kohdetekstistä riippumattomien käännöskielelle ominaisten piirteiden tutkimuksen merkeissä (Laviosa 2002: 1). Kun nämä kaksi yhdistettiin, saatiin oma tieteenala, **korpuslingvistinen käännöstutkimus**, joka perustuu korpuksiin ja niiden hyödyntämiseen. Korpuksiin perustuvassa käännöstutkimuksessa tutkitaan kääntämistä korpusten avulla hyödyntäen korpuslingvistiikan tarjoamia metodeja ja työkaluja (Kenny 2001: 23). Lisäksi korpuslingvistisessä käännöstutkimuksessa tutkitaan muun muassa korpusteknologian käyttöä kääntäjän apuvälineenä. Yksi keskeinen korpuslingvistiikan ja korpuslingvistisen käännöstutkimuksen käsite on korpus. Korpus voidaan määritellä eri tavoin näkökulmasta riippuen suppeasti tai laajasti. Seuraavassa esitellään joitakin korpuksen määritelmiä.

2.2 Korpuksen määritelmä

Elektronisten korpusten yleistymisen myötä korpustutkimuksesta on viime aikoina tullut yhä keskeisempi tutkimussuuntaus kieli- ja käännöstieteellisessä tutkimuksessa. Kielitieteessä korpuksella (latinan sanasta *corpus*, kokoelma) tarkoitetaan laajassa merkityksessä yksinkertaisesti tiettyä tarkoitusta varten koottua kieliaineistoa, joka voi olla teksti- tai puhemuodossa (Koskeniemi 2004).

Korpuslingvistiikassa ja korpuslingvistikassa käännöstutkimuksessa **korpus** (engl. *corpus*, mon. *corpora*) määritellään kuitenkin suppeammin yleensä suureksi autenttisten tekstien kokoelmaksi, joka on koottu tiettyyn tarkoitukseen ja tiettyjen kriteerien perusteella ja jota voidaan käsitellä elektronisessa muodossa (Baker 1995: 225–226; Bowker & Pearson 2002: 9). Toisin sanoen, korpus koostuu siis aina suuresta määrästä autenttisia tekstejä, jotka ovat esimerkkejä todellisesta ja luonnollisesta kielenkäytöstä kirjoitetussa tai puhutussa muodossa. Korpus ei myöskään ole pelkkä satunnainen tekstikokoelma, vaan sen sisältämä aines on valittu huolella ennalta määrättyjen kriteereiden mukaan siten, että korpus on kattava otos tietyistä kielestä tai kielen variantista. Korpus kootaan myös aina joltain tarkoitusta silmällä pitäen, esimerkiksi tutkimusta, kääntämistä tai opettamista varten. Nykyään korpuksia tallennetaan elektronisessa eli tietokoneen luettavissa olevassa muodossa, jotta niitä voidaan käsitellä tietoteknisten korpustyökalujen avulla. Lisäksi elektronisessa muodossa olevan korpuksen kokoaminen ja käyttäminen on nopeaa ja vaivatonta. (Bowker & Pearson 2002: 9–10.)

Korpuksia voidaan jaotella erilaisiin tyypeihin niiden sisältämän tekstiaineuksen tai käyttö-tarkoituksen mukaan. Korpus voi koostua esimerkiksi yleiskielen tai erikoisalan teksteistä, jolloin puhutaan *yleis-* tai *erikoiskorpuksesta*. Lisäksi korpus voi koostua pelkästään kirjoitetuista teksteistä, jolloin puhutaan *tekstikorpuksesta*, tai vaihtoehtoisesti puhutusta kielestä sisältäen audiomuotoista dataa äänitiedostoina, jolloin kyseessä on *puhekorpus*. Korpus voi myös koostua joko yksi- tai useampikielisistä teksteistä, jolloin puhutaan *yksikielisestä korpuksesta* tai *monikielisestä korpuksesta*. Monikieliset korpuksia voidaan jakaa myös rinnakkaisiin ja verrannollisiin korpuksiin. *Rinnakkaiskorpuksia* sisältävät lähdetekstejä ja niiden käännöksiä yhdelle tai useammalle kielelle, siksi niitä kutsutaan myös *käännöskor-*

*puksiksi. Verrannolliset korpuks*et puolestaan koostuvat erikielisistä teksteistä, jotka on kirjoitettu samasta aiheesta ja joilla on sama kommunikatiivinen funktio. Korpus voi olla myös *synkroninen*, jolloin se edustaa kielenkäyttöä tietyllä aikavälillä, tai *diakroninen*, jolloin se sisältää tekstejä pidemmältä aikaväliltä. Lisäksi korpus voi olla avoin eli ns. *monitorinen*, mikä tarkoittaa sitä, että korpusta päivitetään jatkuvasti lisäämällä siihen uusia tekstejä. Korpukseen voi olla myös liitettynä lingvististä tietoa, jolloin puhutaan *annotoidusta korpuksesta*. (Bowker & Pearson 2002: 11–13.)

2.3 Tietokoneen edut ja haitat korpusten käsittelyssä

Vielä joitakin vuosikymmeniä sitten tekstien käsittely tietokoneella oli ongelmallista. Tietokoneet olivat suuria ja hitaita, ja ne vaativat käyttäjältään erityistä tietoteknistä osaamista. Lisäksi jopa lyhyenkin tekstin käsittely tietokoneella oli epäluotettavaa. Nykyään tilanne on toinen, sillä tietokoneet ovat paitsi helposti saatavilla myös tehokkaampia kuin aiemmin. Siksi esimerkiksi laajojen kieliaineistojen prosessointi on helpottunut huomattavasti.

On selvää, että tietokone tuo korpusten käsittelyyn monenlaisia mahdollisuuksia. Tietokoneella tehdyn analyysin etuna on ennen kaikkea datan käsittelyn nopeus, sillä nykyisillä tietokoneilla voidaan käsitellä nopeasti, tehokkaasti ja johdonmukaisesti suurta määrää teksti- tai puheainesta. Lisäksi tietokoneanalyysi poistaa inhimilliset huolimattomuusvirheet ja tarjoaa mahdollisuuden datan objektiiviseen käsittelyyn, valintaan ja helppoon järjestämiseen sekä datan jatkokäsittelyyn ja tallentamiseen eri muodoissa. Saadut tulokset ovat myös luotettavampia, koska tutkijan omat mieltymykset eivät vaikuta analyysiin. (Barnbrook 1996: 11–12; McEnery et al. 2006: 6.) Korpuksiin voidaan lisäksi liittää esimerkiksi monenlaista metadataa ja lingvististä tietoa (McEnery et al. 2006: 6), mikä voi helpottaa tekstin prosessointia korpustyökalun avulla. Kielen- tai kääntämisentutkimuksen kannalta tietokoneet mahdollistavat myös tutkimusmenetelmän muokkauksen tutkimuksen edetessä, ja tutkija voi tehdä kokeiluja käytössä olevalla tietoteknisellä työkalulla ja tutkimuksen ulkopuolisella aineistolla suhteellisen helposti.

Kielen- ja kääntämisentutkimuksen kannalta korpusten käsittelyssä on myös haittoja. Jotta tekstejä voisi ylipäättään käsitellä ja analysoida tietokoneella, tulee aineisto ensin muuttaa tietokoneen ymmärtämään elektroniseen muotoon korpukseksi. Tästä aiheutuu usein aikaa vievää lisätyötä. Lisäksi käytettävien korpustyökalujen sopivuuden testaus tai uuden työkalun kehittäminen vie ylimääräisiä resursseja. Tietokoneilta puuttuu myös ihmisellä oleva taustatieto kielestä ja sen ominaisuuksista, mikä voi vaikeuttaa tekstien analysoimista ja prosessointia. (Barnbrook 1996: 12.) Lisäksi kielen tai kääntämisen tutkija voi usein joutua tyytymään olemassa oleviin aineistoihin ja ohjelmiin sekä helppoihin ratkaisuihin sen sijaan, että käyttäisi omaa luovuuttaan esimerkiksi datan analysoinnissa. Tutkijan omat tietotekniset taidot voivat myös olla riittämättömät, joten kieliaineksen prosessointiin ja analysointiin tarkoitettujen työkalujen toimintoja ei saada siten hyödynnettyä tarpeeksi tehokkaasti. Pelkästään tietoteknisten työkalujen käyttöönotto voi olla hankalaa, ellei niiden käyttöä ole opiskeltu jo koulutuksen aikana.

Koska tietokoneet mahdollistavat korpusten tehokkaamman ja monipuolisemman käytön, on korpuksia viime vuosina käytetty apuna muun muassa vieraan kielen ja kirjoittamisen opettamisessa. Seuraavassa kartoitetaan korpusten mahdollisuuksia käännöstyössä.

2.4 Korpusten hyödyntäminen kääntämisessä

Korpuksia voidaan hyödyntää monin tavoin käännettäessä tekstiä kieleltä toiselle. Tekniisiin tai erikoisalan teksteihin erikoistunut kääntäjä voi hyödyntää korpuksia esimerkiksi tietyn alan termeihin tutustumisessa, terminologisen tutkimuksen teossa, aiemmin käytettyjen käännösstrategioiden tarkastelussa tai tekstin tyypin ja tyylillisten seikkojen arvioimisessa. Kirjallisuuden kääntäjä voi edellä mainittujen seikkojen lisäksi hyödyntää korpuksia esimerkiksi tarkastellessaan lähdetekstin kirjoittajan tyyliä ja kirjoittajalle ominaisten sanamuotojen käyttöä. (Olohan 2004: 176.)

Kuten Kübler (2011: 67–68) huomauttaa, korpuksia voidaan hyödyntää myös jokaisessa kääntämisen vaiheessa riippumatta siitä, mitä käännetään. Ensinnäkin korpuksia voidaan hyödyntää dokumentointivaiheessa, jossa kääntäjä perehtyy tekstin aiheeseen ja sisältöön sekä lähtö- ja kohdeteksteissä käytettyyn terminologiaan. Korpuksia voidaan käyttää hy-

väksi myös varsinaisessa käännösvaiheessa, jossa kääntäjä kääntää lähtötekstin kohdekielelle ja etsii ratkaisuja erityisiin käännösongelmiin. Korpuksia voidaan hyödyntää myös viimeisessä tarkistusvaiheessa, jolloin kääntäjä tarkistaa lopullisen käännöksen ja siinä esiintyvän terminologian.

Erilaiset korpustyytit mahdollistavat myös niiden erilaisen hyödyntämisen kääntämisessä. Rinnakkaisia korpuksia voidaan käyttää kaksikielisen sanakirjan tavoin apuna kääntämisessä etenkin silloin, kun halutaan tarkastella tai etsiä yksittäisen termin potentiaalisia käännösvastineita. Tällöin voidaan hyödyntää esimerkiksi ns. kaksikielistä konkordanssiohjelmaa syöttämällä avainsana yhdellä kielellä, jolloin konkordanssiohjelma näyttää kaikki avainsanan esiintymät konteksteineen sekä näitä vastaavat käännökset, yleensä rinnakkain (Bowker & Pearson 2002: 194). Kuten Evans (2012) huomauttaa, rinnakkaiskorpus on kaksikielistä sanakirjaa parempi käännösapu, sillä rinnakkaiskorpusta käyttämällä avainsanaa voidaan tarkastella laajemmassa kontekstissa yhdessä sen kontekstin, kollokaattien ja käännösvastineiden kanssa. Tällöin kääntäjä voi paitsi tehdä helpommin päätöksen avainsanan oikeasta käännöksestä, myös saada tietoa avainsanan käytöstä ja merkityksestä (Olohan 2004: 177).

Yksikielisiä korpuksia voidaan rinnakkaisten korpusten tavoin myös hyödyntää kääntämisen apuvälineenä. Yksikieliset korpukset ovat käyttökelpoisia esimerkiksi silloin, kun kääntäjä haluaa todentaa ns. aavistustermin käännöksen paikkansapitävyyden. Yksikielistä korpusta voidaan hyödyntää myös käännösvastineiden etsimisessä rinnakkaisen korpuksen tapaan. Tällöin voidaan käyttää yksikielistä konkordanssiohjelmaa ja ns. kontekstihakua, jolla voidaan tarkastella termin käyttöä kontekstissaan. Yksittäiselle sanalle synonyymiä valittaessa voidaan hyödyntää myös ns. frekvenssilistoja. (Bowker & Pearson 2002: 198–209.)

Olohanin (2004: 13) mukaan käännöstieteellisessä tutkimuksessa korpuksia on hyödynnetty vielä toistaiseksi suhteellisen vähän (ks. alan pioneeri Baker 1993). Myöskään ammattikäntäjät eivät ole innostuneet korpusten käytöstä varsinaisessa käännöstyössä. Küblerin (2011: 63–64) mukaan tähän on useita syitä. Ensinnäkin, kaikki kääntämistä opettavat opilaitokset eivät ole sisällyttäneet korpusten ja korpustyökalujen käytön opettelemista

osaksi kääntäjäkoulutusta. Toiseksi korpuksia ja korpustyökaluja ei välttämättä ole aina saatavilla tietylle kielelle, ja vaikka korpustyökalut ovat kehittyneet viimeisten vuosikymmenien aikana, ne eivät ole välttämättä käytettäviä, ja lisäksi niiden käyttäminen vaatii aina tietynlaista erityisosaamista. Korpusten ja korpustyökalujen osaamista ei myöskään yleensä edellytetä esimerkiksi kääntäjille suunnatuissa työpaikkailmoituksissa.

Yksi syy siihen, miksi kääntäjät eivät käytä korpuksia ja korpustyökaluja käännöstyössään voi olla se, ettei heillä ole aiempaa kokemusta korpusten ja korpustyökalujen käytöstä. Kuten Olohan (2004: 176) huomauttaa, vaikka kääntäjät ovat ymmärtäneet korpusten ja korpustyökalujen potentiaalin, heillä ei välttämättä ole työnsä ohessa aikaa tutustua korpuksiin ja korpustyökaluihin tai käyttää niitä laajasti käännöstyössään, sillä korpusten koaminen ja käyttö on aikaa vievää. Näin ollen kääntäjät eivät välttämättä edes tiedä, millaisia korpuksia ja korpustyökaluja on saatavilla ja millaisia toimintoja ne sisältävät. Toinen syy korpusten käytön vähyyteen käytännön käännöstyössä voi olla myös siinä, ettei korpustyökaluja alun perin oltu suunniteltu kääntäjiä vaan muun muassa kielentutkijoita ja kieltenopettajia varten (Bowker 2002: 47; Wilkinson 2006). Lisäksi korpuslingvistiikassa käytettyjä tietoteknisiä työkaluja ei myöskään ole arvioitu kääntäjän näkökulmasta. Siksi kääntäjillä ei välttämättä ole selkeää kuvaa siitä, miten korpustyökalut ylipäättään sopivat tekstin kääntämiseen kielestä toiseen.

Jotta kääntäjät voisivat hyödyntää korpuksia, he tarvitsevat sopivia työkaluja korpusten käsittelemiseksi. Seuraavassa luvussa tutustutaan korpustyökaluiksi luettaviin yksinkertaisiin konkordanssiohjelmiin. Lisäksi luvussa kuvaillaan konkordanssiohjelmien jaottelua sekä sellaisia yleisiä ominaisuuksia, joita lähdekirjallisuuden mukaan voi löytää useimmista konkordanssiohjelmista. Lopuksi pohditaan lähdekirjallisuuteen nojaten konkordanssiohjelmien sisältämien toimintojen hyödyntämistä kääntämisessä sekä määritellään hyvän konkordanssiohjelman tunnusmerkkejä.

3 Konkordanssiohjelmat

Korpusten käyttöä ja analysointia varten on vuosien saatossa kehitetty monenlaisia tietoteknisiä työkaluja. **Korpustyökalut** (engl. *corpus tools*) ovat tietokoneohjelmia, jotka mahdollistavat suurten datamäärien automaattisen prosessoinnin suhteellisen nopeasti, tarkasti ja vaivattomasti. Korpustyökalut tarjoavat erilaisia keinoja korpuksessa olevan aineiston uudelleenjärjestämiseen käyttäjäystävällisempään ja analysoitavampaan muotoon. Lisäksi korpustyökalujen avulla voidaan muodostaa monenlaista tilastotietoa korpuksen sisältämästä datasta. (Evans 2012.) Korpustyökalut voidaan jakaa karkeasti kahteen pääryhmään: laajempiin, yleensä maksullisiin korpusten analysointiohjelmiin ja pienempiin, yleensä ilmaisena saatavilla oleviin konkordanssiohjelmiin.

Tässä luvussa tarkastellaan konkordanssiohjelmia määrittelemällä ensin konkordanssiohjelman käsite. Tämän jälkeen luvussa kuvataan konkordanssiohjelmien jaotteluperusteita sekä konkordanssiohjelmissä yleensä esiintyviä toimintoja, sekä pohditaan toimintojen hyödyntämiskeinoja kääntämisessä. Lopuksi määritellään lähdekirjallisuuteen nojaten hyvän konkordanssiohjelman kriteerit.

3.1 Konkordanssiohjelman määritelmä ja jaottelu

Konkordanssiohjelma (engl. *concordancer, concordance program*) on yksinkertainen tietokoneohjelma, jonka avulla voidaan tarkastella valitun hakusanan tai sanojen esiintymiä tekstiyhteydessään käsiteltävänä olevassa korpuksessa (Hunston 2002: 39). Konkordanssiohjelmalla voidaan usein tarkastella myös esimerkiksi sanojen kombinaatioita eli kollokaatteja (engl. *collocates*) tai klustereita eli toistuvasti yhdessä esiintyvien sanojen jonoja (engl. *clusters*) (Olohan 2004: 63). Lisäksi konkordanssiohjelmillä voidaan tarkastella sanojen esiintymistiheyksiä sekä muodostaa sanalistoja (Biber et al. 1998: 257).

Konkordanssiohjelmia on kahta tyyppiä. Ohjelma voi olla verkosta ladattavissa kotikoneelle, jolloin ohjelma on saatavilla joko zip-pakettina tai suoritettavana exe-tiedostona. Esimerkki tällaisesta konkordanssiohjelmasta on *Multilingual Corpus Toolkit*, joka on vapaasti ladattavissa zip-pakettina ohjelman kotisivulta (Piao 2009). Kotikoneelle asennettavien

konkordanssiohjelmien lisäksi on olemassa sovelluksia, jotka ovat käytettävissä ainoastaan verkon kautta. Tällainen sovellus koostuu Blochin (2009: 61) mukaan korpuksesta ja www-käyttöliittymästä, jonka avulla voidaan tehdä valittujen kriteereiden mukaisia hakuja tietystä korpuksesta. Hyviä esimerkkejä tällaisista www-käyttöliittymän avulla käytettävistä korpuksista ovat *WebCorp Live*, joka käyttää koko Internetiä korpuksena (Birmingham City University 2012), sekä *British National Corpus*, jonka www-käyttöliittymällä voi hakea konkordansseja 100 miljoonan sanan korpuksesta, jossa on kirjoitetun ja puhekielen tekstejä useista eri lähteistä (University of Oxford 2010).

Konkordanssiohjelmiä on lisäksi saatavilla joko kaupallisina tai täysin ilmaisversioina. Kaupalliset konkordanssiohjelmat ovat yleensä jonkun yksityisen henkilön tai henkilöiden kehittämiä. Kaupallisista konkordanssiohjelmissä on useimmiten tarjolla ilmainen kokeiluversio, jonka käyttöä on joko rajattu tai josta on karsittu pois joitakin maksullisessa ohjelmassa olevia toimintoja. R. J. C. Wattin kehittämä *Concordancer* on hyvä esimerkki tällaisesta konkordanssiohjelmasta, sillä kyseisestä ohjelmasta on saatavilla 30 päivän kokeiluversio maksullisen version ohella (Watt 2010). Joskus yksinkertaisesta konkordanssiohjelmasta on monen vuoden jatkokehityksen tuloksena voitu saada aikaan laajempi korpustyökalu, joka on uusien ominaisuuksiensa myötä muuttunut maksulliseksi tuotteeksi, ja johon on saatavilla uusia päivityksiä ja käyttäjätukea. Mike Scottin *WordSmithTools* on hyvä esimerkki kaupallisesta korpustyökalusta, joka on kehittynyt alun perin yksinkertaisesta konkordanssiohjelmasta (*MicroConcord*) monipuoliseksi korpusten analysointiohjelmaksi (Scott 2008).

Maksullisten konkordanssiohjelmien tapaan myös ilmaiset konkordanssiohjelmat ovat yksityisen henkilön, yleensä kielitieteilijän tai muun alan tiedeyhteisön jäsenen kehittämiä. Erona maksullisiin konkordanssiohjelmiin on paitsi hinta myös esimerkiksi se, mitä aineiston käsittelyn ja -analysoinnin keinoja ohjelmat tarjoavat, mikä on ohjelmien suorituskyky, missä käyttöjärjestelmässä ohjelman voi ajaa, millä ohjelmointikielellä konkordanssiohjelma on toteutettu, onko ohjelmapäivityksiä, dokumentaatiota tai käyttöohjeita saatavilla ja tarjotaanko sovelluksen käyttäjille tukea esimerkiksi sähköpostin välityksellä. Ilmaisten konkordanssiohjelmien kehitys on myös voitu rahoittaa esimerkiksi yliopiston tuella. Esimerkkinä tällaisesta ilmaisesta konkordanssiohjelmasta on *AntConc*, jonka on kehittänyt

Wasedan yliopistossa professorina työskentelevä Laurence Anthony. Kyseisen ohjelman kehitys on rahoitettu muun muassa Wasedan yliopiston tarjoaman apurahan turvin (Anthony 2012a: 9).

Konkordanssiohjelmat voidaan myös jaotella sen mukaan, minkälaista korpusta niillä käsitellään. Yksikieliset konkordanssiohjelmat (engl. *monolingual concordancer*) pystyvät käsittelemään vain yksikielisiä korpuksia. Kaksikieliset konkordanssiohjelmat (engl. *parallel concordancer*) puolestaan pystyvät käsittelemään yksi-, kaksi- tai monikielisiä korpuksia. Tällöin käyttöliittymässä erikieliset tekstit näytetään yleensä rinnakkain omissa ikkunoissaan. (Bowker & Pearson 2002: 120.) Hyvä esimerkki yksikielisestä konkordanssiohjelmasta on *MonoConc Pro* ja kaksikielisestä konkordanssiohjelmasta *ParaConc*. Molemmat ovat Athelstanin kehittämiä kaupallisia konkordanssiohjelmiä. (Athelstan 2012.)

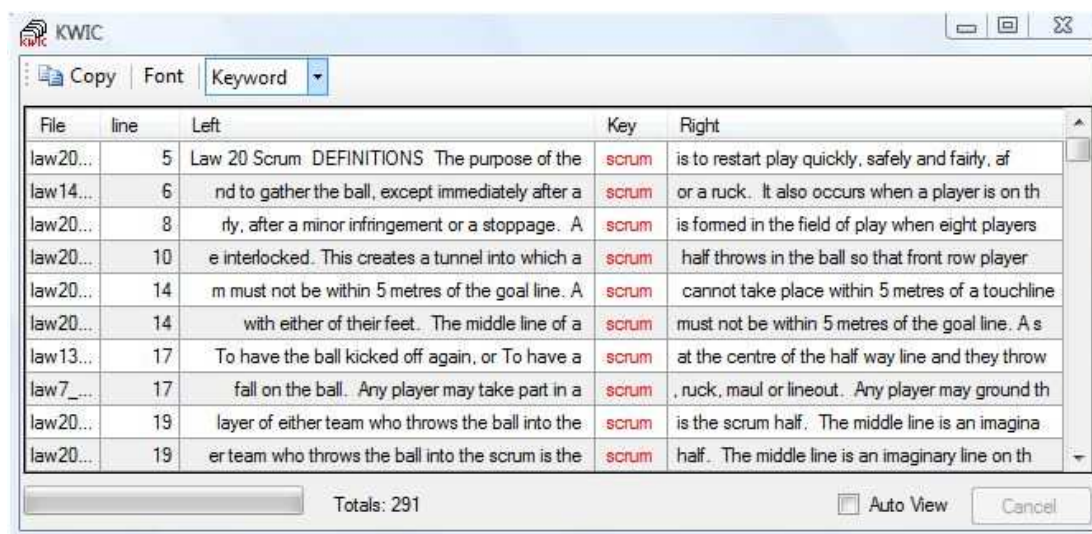
3.2 Konkordanssiohjelmien yleiset toiminnot

Maksullisille ja ilmaisille konkordanssiohjelmit on yhteistä se, että niissä on usein samanlaisia toimintoja. Yleisimpiä konkordanssiohjelmien toimintoja ovat konkordanssit, kollokaatit, frekvenssilistat ja statistiikkaan liittyvät toiminnot (Bowker and Pearson 2002: 13). Alla oleva esitys ei pyri olemaan mitenkään kaiken kattava, vaan tarkoitus on esitellä sellaisia ominaisuuksia ja toimintoja, jotka löytyvät useimmista konkordanssiohjelmita ja joita voidaan hyödyntää kääntämisessä kielestä toiseen.

3.2.1 Konkordanssit

Konkordanssiohjelman yksi keskeinen toiminto on ns. konkordanssitoiminto, jonka avulla voidaan hakea **konkordansseja** (engl. *concordances*) käsiteltävänä olevasta tekstistä tai tekstikorpuksesta. Konkordanssilla tarkoitetaan halutun merkkijonon esiintymää kontekstin kanssa (Olohan 2004: 199). Haku toimii siten, että käyttäjä syöttää hakukenttään yhden tai useamman hakusanan ja suorittaa haun. Tämän jälkeen konkordanssiohjelma näyttää luettelomuodossa kaikki hakusanan tai hakusanojen esiintymät keskellä, ja niiden ympärillä molemmin puolin halutun määrän kontekstia. Jotkut konkordanssiohjelmat näyttävät lisäksi hakusanan korostettuna esimerkiksi tietyllä värillä. Lisäksi konkordanssirivit voi usein

järjestää ohjelman avulla eri tavoin esimerkiksi aakkosjärjestykseen, ja tallentaa tekstitiedostoon. (Bowker & Pearson 2002: 121–124.) Kuviossa 1 esitetään esimerkki *KWIC Concordance for Windows* -konkordanssiohjelman tulostamasta konkordanssiotteesta.



The screenshot shows the KWIC Concordance for Windows application window. It features a menu bar with 'File', 'Copy', 'Font', and 'Keyword'. Below the menu is a table with five columns: 'File', 'line', 'Left', 'Key', and 'Right'. The 'Key' column contains the word 'scrum' in red text, which is highlighted in the original image. The table lists concordance results for the keyword 'scrum' across various files and line numbers. At the bottom of the window, there is a status bar showing 'Totals: 291' and buttons for 'Auto View' and 'Cancel'.

File	line	Left	Key	Right
law20...	5	Law 20 Scrum. DEFINITIONS The purpose of the	scrum	is to restart play quickly, safely and fairly, af
law14...	6	nd to gather the ball, except immediately after a	scrum	or a ruck. It also occurs when a player is on th
law20...	8	rtly, after a minor infringement or a stoppage. A	scrum	is formed in the field of play when eight players
law20...	10	e interlocked. This creates a tunnel into which a	scrum	half throws in the ball so that front row player
law20...	14	m must not be within 5 metres of the goal line. A	scrum	cannot take place within 5 metres of a touchline
law20...	14	with either of their feet. The middle line of a	scrum	must not be within 5 metres of the goal line. A s
law13...	17	To have the ball kicked off again, or To have a	scrum	at the centre of the half way line and they throw
law7_...	17	fall on the ball. Any player may take part in a	scrum	, ruck, maul or lineout. Any player may ground th
law20...	19	layer of either team who throws the ball into the	scrum	is the scrum half. The middle line is an imagina
law20...	19	er team who throws the ball into the scrum is the	scrum	half. The middle line is an imaginary line on th

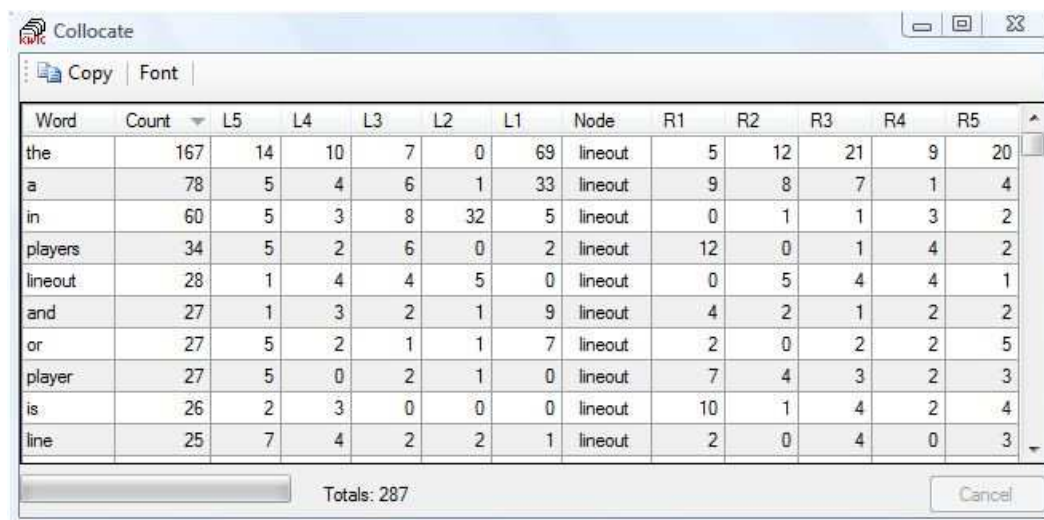
Kuvio 1. *KWIC Concordance for Windows* -konkordanssiohjelman tulostama konkordanssiote, jossa hakutermi 'scrum' on korostettu punaisella värillä.

Konkordanssihausta käytetään usein myös nimitystä avainsanahaku, eli **KWIC-haku** (engl. *keyword in context*) (Olohan 2004: 63). Konkordanssitoiminto löytyy yleensä mistä tahansa korpustyökalusta tavalla tai toisella toteutettuna (O'Donnell 2008: 108), mikä kertoo toiminnon hyödyllisyydestä.

Konkordansseja voi käyttää kääntämisessä esimerkiksi sanakirjan rinnalla, kun halutaan tarkastella sanan tai ilmauksen käyttöä kontekstissaan. Konkordanssin avulla saadaan siten hyvä kokonaiskuva sanan erilaisista käyttötavoista ja näin saada kuva tekstissä esiintyvistä termeistä. Lisäksi konkordansseja voi käyttää esimerkiksi käännösvastineiden etsimiseen. (Gawronska 2008.)

3.2.2 Kollokaatit

Jotkut konkordanssiohjelmat tarjoavat myös mahdollisuuden saada selville, minkä sanojen kanssa hakusanaa yleensä käytetään, toisin sanoen mitkä ovat ko. sanan tavanomaiset **kollokaatit** (engl. *collocates*). Haku toimii samalla tavalla kuin konkordanssihaku, mutta lisäksi yleensä määritetään mahdollisten sanojen lukumäärä hakutermien välillä. Tyypillisenä raja-arvona käytetään viittä sanaa hakutermistä sekä vasemmalle että oikealle. (Bowker & Pearson 2002: 124; Olohan 2004: 82.) Ohjelmasta riippuen kollokaattihaun tulokset näytetään taulukkomuodossa tai konkordanssiotteessa. Kuviossa 2 havainnollistetaan *KWIC Concordance for Windows* -konkordanssiohjelman tapaa esittää löydetty kollokaatit taulukossa.



The screenshot shows the 'Collocate' window of the KWIC Concordance for Windows software. It displays a table of collocates for the word 'lineout'. The table has columns for Word, Count, L5, L4, L3, L2, L1, Node, R1, R2, R3, R4, and R5. The data is as follows:

Word	Count	L5	L4	L3	L2	L1	Node	R1	R2	R3	R4	R5
the	167	14	10	7	0	69	lineout	5	12	21	9	20
a	78	5	4	6	1	33	lineout	9	8	7	1	4
in	60	5	3	8	32	5	lineout	0	1	1	3	2
players	34	5	2	6	0	2	lineout	12	0	1	4	2
lineout	28	1	4	4	5	0	lineout	0	5	4	4	1
and	27	1	3	2	1	9	lineout	4	2	1	2	2
or	27	5	2	1	1	7	lineout	2	0	2	2	5
player	27	5	0	2	1	0	lineout	7	4	3	2	3
is	26	2	3	0	0	0	lineout	10	1	4	2	4
line	25	7	4	2	2	1	lineout	2	0	4	0	3

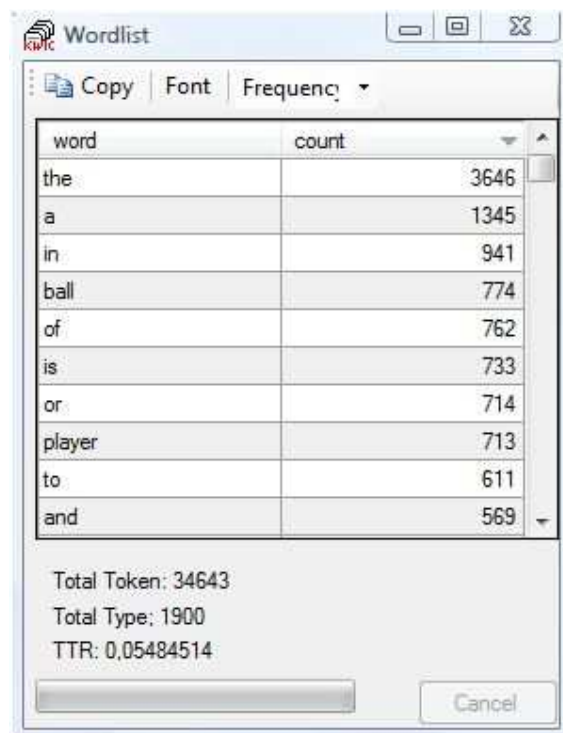
At the bottom of the window, it shows 'Totals: 287' and a 'Cancel' button.

Kuvio 2. *KWIC Concordance for Windows* -konkordanssiohjelman tulostama kollokaattitaulukko sanalle 'lineout'.

Kollokaatteja voidaan hyödyntää kääntämisessä esimerkiksi silloin, kun tutkitaan lähdetekstissä esiintyviä merkityksiä etenkin usein yhdessä esiintyvien sanojen osalta (Olohan 2004: 82). Näin saadaan tietoa paitsi itse lähdetekstistä, myös tietyn vakiintuneen ilmauksen mahdollisista käännösvastineista.

3.2.3 Frekvenssilistat

Kolmas useimmissa konkordanssiohjelmissa esiintyvä toiminto on ns. **frekvenssilistojen** (engl. *frequency lists*) luonti. Yksinkertaisimmillaan frekvenssilista sisältää listan kaikista sanoista ja niiden esiintymiskerroista käsiteltävänä olevassa tekstiaineistossa (Olohan 2004: 77). Frekvenssilista voidaan luoda konkordanssiohjelmalla yleensä helposti esimerkiksi vain yhtä painiketta painamalla. Tällöin konkordanssiohjelma tunnistaa sanojen esiintymät, laskee identtisten sanojen esiintymät yhteen ja tulostaa lopuksi sanat listana esiintymiskerrat mukaan lukien joko näytölle tai tiedostoon (Barnbrook 1996: 43). Yleensä frekvenssilistan voi lisäksi järjestää konkordanssiohjelman avulla esimerkiksi sanojen esiintymismäärän tai järjestyksen mukaan, tai aakkosjärjestykseen laskevasti tai nousevasti (Olohan 2004: 77). Kuviossa 3 esitetään esimerkki frekvenssilistasta *KWIC Concordance for Windows* -konkordanssiohjelman tulostamana.



word	count
the	3646
a	1345
in	941
ball	774
of	762
is	733
or	714
player	713
to	611
and	569

Total Token: 34643
Total Type: 1900
TTR: 0,05484514

Cancel

Kuvio 3. *KWIC Concordance for Windows* -konkordanssiohjelman tulostama frekvenssilista.

Jotkut konkordanssiohjelmat mahdollistavat frekvenssilistan luomisen myös **avainsanoista** (engl. *key words*). Tällöin ohjelma tunnistaa ne sanat, jotka esiintyvät useimmin käsiteltävänä olevassa korpuksessa suhteessa esimerkiksi johonkin toiseen korpukseen, ja muodostaa kyseisistä sanoista listan. Jotkut konkordanssiohjelmat myös sallivat ns. **estolistan** (engl. *stop list*) käytön frekvenssilistoja muodostettaessa. Tämä tarkoittaa sitä, että käyttäjä voi muodostaa listan niistä sanoista, joita ei haluta ottaa mukaan frekvenssilistaan. (Bowler & Pearson 2002: 113–115.)

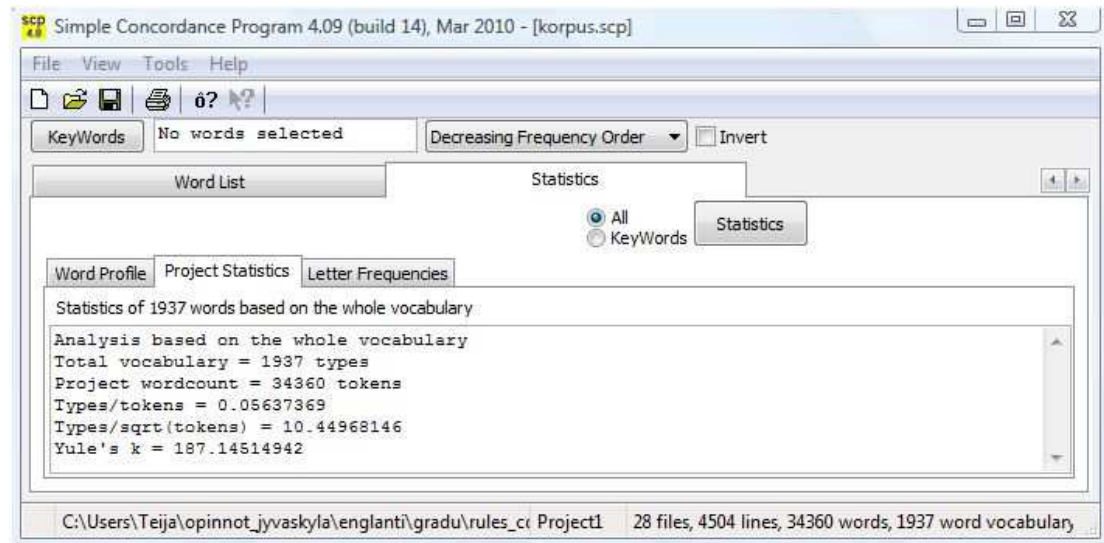
Frekvenssilistoja voidaan hyödyntää kääntämisessä monin tavoin. Frekvenssilistojen avulla saadaan esimerkiksi hyödyllistä tietoa käsiteltävänä olevan tekstin pääpiirteistä ja tekstissä yleisimmin esiintyvistä sanoista. Lisäksi frekvenssilistojen avulla voidaan valita helpommin mielenkiintoisimmat sanat lähemmän tarkastelun kohteeksi. (Barnbrook 1996: 47.) Frekvenssilistoista erityisesti avainsanalistat ovat kääntäjän kannalta hyödyllisiä, sillä avainsanalista auttaa kääntäjää selvittämään, mistä teksti kertoo ja mitä käsitteitä siinä käytetään. Mitä korkeampi sanan frekvenssi on, sitä keskeisempi käsite kyseinen sana on tekstissä. Lisäksi avainsanalista kertoo kääntäjälle, mihin erikoisalaan teksti sijoittuu, miten tekstissä liitetään sisältöjä toisiinsa ja millaisia metaforia tai epätyypillisiä sanoja tekstin kirjoittaja käyttää. (Kemppanen 2008.)

3.2.4 Statistiikka

Joillakin konkordanssiohjelmissa voi myös analysoida käsiteltävänä olevaa tekstiaineistoa määrällisesti, sillä jotkut konkordanssiohjelmat osaavat muodostaa monenlaista tilastotietoa. Konkordanssiohjelmien muodostama tilastotieto voi sisältää tietoa esimerkiksi korpuksen kokonaissanamäärästä, toisistaan eroavien sanojen kokonaismäärästä, sanojen ja lauseiden pituudesta ja määrästä, kappaleiden määrästä sekä muista tekstin sanastollisista piirteistä. (Olohan 2004: 78).

Eräs joidenkin konkordanssiohjelmien laskema tilastotieto on Eskolan (2002: 93) käyttämin termin **sanastollisen rikkauden indeksi** (engl. *type-token ratio*, *TTR*), joka on yksi hyödyllisimmistä tekstin mittareista. *Type-token ratio* -termissä *token* viittaa sanojen kokonaismäärään käsiteltävässä tekstissä, kun *type* puolestaan viittaa toisistaan eroavien sanojen

kokonaismäärään. Näiden välinen suhde lasketaan jakamalla toisistaan eroavien sanojen määrä sanojen kokonaismäärällä. (Olohan 2004: 78–80; Williamson 2009.) Kuviossa 4 esitetään *Simple Concordance Program* -konkordanssiohjelman muodostama tilastotieto eräästä korpuksesta. Ohjelman laskema sanastollisen rikkauden indeksi esitetään Project Statistics -välilehdellä neljännellä rivillä.



Kuvio 4. Simple Concordance Program -konkordanssiohjelman muodostama tilastotieto käsiteltävästä korpuksesta.

Sanastollisen rikkauden indeksin avulla voidaan analysoida tekstin leksikaalista vaihtelevuutta eli sitä, kuinka paljon eri sanamuotoja tekstissä on käytetty verrattuna tekstin kokonaissanamäärään. Toisin sanoen, mitä enemmän tekstissä on toisistaan eroavia sanoja verrattuna kaikkiin sanoihin, sitä vaihtelevampaa on tekstin sanasto. Kääntämisessä sanastollisen rikkauden indeksiä voidaan käyttää apuna esimerkiksi vertailtaessa eri tekstityyppejä keskenään, tai vertailtaessa lähtö- ja kohdetekstejä keskenään. (Baker 1995: 236–237.)

3.3 Hyvän konkordanssiohjelman tunnusmerkit

Konkordanssiohjelmaa on olemassa maksullisia ja ilmaisia, ne voivat sisältää hyvin erilaisia toimintoja, ja niillä voidaan käsitellä erilaisia korpuksia. Tästä syystä konkordanssiohjelmaa valittaessa on hyvä pitää mielessä, mitä hyvältä konkordanssiohjelmalta voidaan odottaa. Barnbrookin (1996: 68–78) mukaan hyvä konkordanssiohjelma löytää kaikki halutun hakusanan esiintymät nopeasti ja esittää haun tulokset sopivassa muodossa. Lisäksi hyvä konkordanssiohjelma sallii avainsanan monipuolisen määrittelyn joko perusmuodossa, sanaparin (kollokaatin) kanssa tai säännöllisenä lausekkeena, sekä sallii kontekstin pituuden säätämisen joko tekstikappaleiden, rivien, sanojen tai merkkien mukaan. Hyvä konkordanssiohjelma myös hyödyntää frekvenssitietoa tulosten rajaamisessa esimerkiksi esittämällä yleisimmät hakutulokset ensin. (Barnbrook 1996: 68–78.)

Edellä mainittujen ominaisuuksien lisäksi hyvä konkordanssiohjelma sallii hakutuloksen monipuolisen järjestämisen erilaisten kriteerien mukaan, esimerkiksi avainsanan tai avainsanaa edeltävän tai sitä seuraavan kontekstin perusteella. Hyvä konkordanssiohjelma tarjoaa käyttäjälleen myös mahdollisuuden lisätä yksittäisille konkordanssiriveille omia kommentteja, joiden mukaan konkordanssirivejä voi myös järjestää. (Bergenholtz & Tarp 1995: 35–36.)

Monipuolisen haun ja hakutulosten esittämisen lisäksi hyvän konkordanssiohjelman tulisi pystyä käsittelemään sanastollisen rikkauden indeksiin liittyvät ongelmat. Konkordanssiohjelman tulisi tehdä ero esimerkiksi sille, etsitäänkö haussa myös hakusanan johdannaismuotoja ja minkä sanaluokan sanana hakusanaa käsitellään. (Antworth & Valentine 1998: 175.) Esimerkiksi suomen kielen sana 'kurki' tarkoittaa substantiivina lintua, mutta sana esiintyy myös verbinä kieltomuodossa 'Älä kurki'. Tällöin ohjelman on tiedettävä, kumpaa sanaluokkaa käyttäjä haluaa hakea.

Wilkinson (2012) lisää hyvän konkordanssiohjelman kriteereihin lisäksi seuraavia käyttäjäystävällisyyteen liittyviä näkökulmia. Wilkinsonin mukaan hyvän konkordanssiohjelman tulisi:

1. olla helposti opittava ja nopea käyttää
2. mahdollistaa käsiteltäväksi tarkoitetun korpuksen suoraviivaisen valinnan
3. mahdollistaa hakusanan tai sanojen helpon syöttämisen
4. mahdollistaa jokerimerkkien käyttö hakutoiminnoissa
5. sisältää nopea KWIC-haku, jossa hakutermi tulisi näkyä keskellä siten, että kontekstia näkyy riittävä määrä hakutermiin molemmin puolin
6. esittää KWIC-hakutulos selvässä ja siistissä muodossa
7. mahdollistaa konkordanssirivien helpon järjestämisen esimerkiksi aakkosjärjestykseen tai hakutermiin mukaan
8. mahdollistaa konkordanssirivin tarkastelu laajemmassa kontekstissa, esimerkiksi kokonaisen lauseen tai kappaleen tasolla
9. olla vakaa, toisin sanoen se ei saisi jähmettyä tai kaatua ja lisäksi
10. olla edullinen.

Yhteenvetona voisi todeta, että hyvä konkordanssiohjelma on paitsi edullinen, vakaa ja käytettävä, se mahdollistaa monipuolisen konkordanssien hakemisen, järjestämisen ja tarkastelun sekä hyödyntää frekvenssitietoa tai mahdollistaa sen muodostamisen. Lisäksi hyvä konkordanssiohjelma huomioi käyttäjänsä erityistarpeet.

Hyvän konkordanssiohjelman valitseminen ei ole aina helppoa. Internetissä on kuitenkin saatavilla lukuisia potentiaalisia Windows-ympäristöön suunnattuja ilmaisia konkordanssiohjelmia, ja niiden kartoitusta kuvataan seuraavaksi. Siten seuraavassa luvussa kuvataan tähän tutkimukseen liittyvä tausta, tavoitteet, tutkimuskysymykset, tutkimusmenetelmä ja tutkimusaineisto.

4 Ilmaisten konkordanssiohjelmien kartoitus

Tässä luvussa esitellään ensin lyhyesti aiempia konkordanssiohjelmiä käsitelleitä tutkimuksia. Lisäksi kuvataan tämän tutkimuksen tavoitteet, tutkimusongelma ja tutkimusky-symykset, tutkimusmenetelmä sekä tutkimusaineisto.

4.1 Aiemmat tutkimukset

Aiemmat konkordanssiohjelmiin liittyvät tutkimukset ovat olleet lähinnä ohjelmien arvosteluja tai katsauksia, jotka on julkaistu jossakin tieteellisessä julkaisussa, esimerkiksi kielitiedettä ja teknologiaa käsittelevissä lehdissä. Näissä arvosteluissa ennalta valittuja konkordanssiohjelmiä on arvioitu ja joskus myös vertailtu keskenään etupäässä niiden tarjoamien toimintojen näkökulmasta, pyrkien antamaan suosituksia sen hetken parhaasta ohjelmasta riippumatta ohjelman käyttäjän todellisista tarpeista. Konkordanssiohjelmiä arvioivat ja vertailevat tutkimukset ovat keskittyneet etupäässä kuvailemaan joko yksittäistä konkordanssiohjelmaa (Bauer 1992; Stevens 2002; Forchini 2005), vertailemaan vain muutamaa konkordanssiohjelmaa keskenään (Diniz 2005; Roberts et al. 2005; Anagnostou & Weir 2006) tai vertailemaan ja kuvailemaan yksinomaan kaupallisten konkordanssiohjelmien ominaisuuksia (Reppen 2001). Lisäksi muutamassa konkordanssiohjelmiä vertailevassa tutkimuksessa on keskitytty vertailemaan vain tiettyä konkordanssiohjelman ominaisuutta, kuten konkordanssiohjelman kykyä prosessoida arabiankielistä tekstiä (Roberts et al. 2005), käyttäjäystävällisyyttä ja tehokkuutta (Ari 2006) tai ohjelman tarjoamaa annotointiominaisuutta (Smith et al. 2008).

Ainoastaan Wiechmann ja Fuhs (2006) ovat arvioineet kattavammin yleisimpiä ilmaisia konkordanssiohjelmiä. Wiechmann ja Fuhs arvioivat kolmen kaupallisen ohjelman lisäksi seitsemää ilmaista konkordanssiohjelmaa; näistä yksi (*ConcApp*) on sittemmin muuttunut maksulliseksi. Wiechmann ja Fuhs arvioivat konkordanssiohjelmiä niiden toiminnallisuuden, tehokkuuden ja käyttäjäystävällisyyden kannalta ohjelmakohtaisesti, ja antoivat suosituksia käytettävästä ohjelmasta etupäässä tutkijoille. Wiechmann ja Fuhs eivät kuitenkaan huomioineet kääntäjiä yhtenä kohderyhmänä suositellessaan ohjelmia.

Kaikkien edellä mainittujen aiempien tutkimusten heikkous on kuitenkin se, että ne ovat keskittyneet ainoastaan konkordanssiohjelmien sisältämiin toimintoihin, mutta yhdessäkään tutkimuksessa ei ole kartoitettu tai arvioitu valittujen konkordanssiohjelmien sopivuutta kääntäjän työhön tai kääntämiseen. Lisäksi käytettävyyden arviointi on sivuutettu lähes kokonaan. Neljässä edellä mainituista konkordanssiohjelmiä arvioivissa tutkimuksissa (Stevens 2002; Anagnostou & Weir 2006; Ari 2006; Wiechmann & Fuhs 2006) käytettävyyttä on arvioitu ainoastaan käyttäjäystävällisyyden näkökulmasta, mutta muut käytettävyyden osa-alueet on jätetty tyystin huomioimatta. Lisäksi konkordanssiohjelmat kehittyvät kaiken aikaa, joten osa edellä mainituissa tutkimuksissa esitellyistä ohjelmista on jo vanhentunut.

Syynä aikaisempien tutkimusten heikkouksiin voi olla että kaiken kattavia, luotettavia tai yleisesti hyväksyttyjä korpustyökalujen ja etenkin yleensä kääntämisessä apuna käytettävien ohjelmien arviointikriteereitä ei yksinkertaisesti ole tällä hetkellä olemassa (Daelemans & Hoste 2008). Toinen syy voi olla että tutkimusten tekijät eivät välttämättä myöskään omaa tarpeeksi tietoteknistä taustaa voidakseen arvioida monipuolisesti esimerkiksi ohjelmien käytettävyyttä. Lisäksi jokainen tiedejulkaisu, jossa edellä mainitut tutkimukset on julkaistu asettaa omat vaatimuksensa esimerkiksi julkaistavan artikkelin pituudelle, joten ohjelmien arvioinnissa on voitu joutua karsimaan arvioitavia tekijöitä.

4.2 Tutkimuksen tavoitteet

Tämän tutkimuksen tavoitteena on kartoittaa verkossa saatavilla olevia Windows-ympäristöön suunnattuja ilmaisia, yksikielisiä konkordanssiohjelmiä sekä niiden yleisimpiä ominaisuuksia. Tavoitteena on aineistoa analysoimalla selvittää, millaisia toimintoja ilmaiset konkordanssiohjelmat sisältävät, sekä lähdekirjallisuuteen perehtymällä kartoittaa konkordanssiohjelmien hyödyntämismahdollisuuksia kääntämisessä. Tutkimuksessa ei ole tarkoitus vertailla valittuja konkordanssiohjelmiä kuitenkaan keskenään, vaan pyrkiä kuvaamaan sitä, millaisia toiminnallisuuksia kyseiset ohjelmat tarjoavat.

Ilmaisia konkordanssiohjelmiä esittelevä tutkimus on hyödyllistä paitsi kielitieteen, kieliteknologian ja käännöstieteen, myös tietojenkäsittelytieteen näkökulmasta. Aihe on ajan-

kohtainen käännöstieteessä, ja lisäksi tietotekniikan osaajat ja kielitieteilijät työskentelevät yhä enenevässä määrin yhteistyössä uusien, entistä monipuolisempien korpustyökalujen kehittämiseksi. Konkordanssiohjelmien avulla voidaan opettaa esimerkiksi vieraan kielen kirjoittamista ja kääntämistä, joten aiheeseen liittyy myös vahvasti opetuksellinen näkökulma. Oletuksena on, että tämä tutkimus tuo paitsi uutta tietoa konkordanssiohjelmissa ja niiden sisältämistä ominaisuuksista, myös niiden sopivuudesta kääntäjän työhön.

4.3 Tutkimusongelma ja tutkimuskysymykset

Tutkimuksella pyritään vastaamaan seuraaviin kysymyksiin:

1. Millaisia ilmaisia konkordanssiohjelmissa on saatavilla Windows-ympäristöön?
2. Millaisia toimintoja ilmaiset konkordanssiohjelmat tarjoavat?
3. Miten konkordanssiohjelmissa voidaan hyödyntää kääntämisessä?

Ensimmäiseen tutkimuskysymykseen pyritään löytämään vastaukset etsimällä sopivia konkordanssiohjelmissa verkosta. Toiseen tutkimuskysymykseen pyritään löytämään vastauksia paitsi kirjallisuuskatsaukseen nojaten, myös testaamalla ohjelmien toimintaa käytännössä Windows-ympäristössä ja kirjaamalla ohjelmissa esiintyneet toiminnot ylös ohjelmakohtaisesti. Kolmanteen tutkimuskysymykseen pyritään löytämään vastaukset tämän tutkimuksen kirjallisuuskatsauksella.

4.4 Tutkimusmenetelmä

Tutkimusaineisto analysoidaan testaamalla konkordanssiohjelmien toimintaa käytännössä 32-bittisessä Windows Vista Home Premium -käyttöjärjestelmässä, johon on myös ladattu Service Pack 2 -käyttöjärjestelmäpäivitys. Yksikieliset konkordanssiohjelmat ladataan tai asennetaan Hewlett-Packard -merkkiselle kannettavalle tietokoneelle, jonka suoritusnopeus on 1,7 Ghz ja jossa on 2 gigatavua RAM-muistia. Ohjelmissa testataan yksitellen, jonka jälkeen ohjelmissa olevat toiminnallisuudet kirjataan ylös ohjelmakohtaisesti. Testauksessa apuna

käytetään tutkimuksen tekijän itsensä kokoamaa pientä, englanninkielistä tekstikorpusta, joka koostuu 28 tavallisesta tekstitiedostosta. Tekstikorpuksen koko on 248 kilotavua.

4.5 Tutkimusaineisto

Tämän tutkimuksen tutkimusaineistona ovat verkosta ladattavat ilmaiset, yksikieliset Windows-alustalle suunnatut konkordanssiohjelmat. Tässä tutkimuksessa käytetty aineisto on kerätty pääasiassa eri sovellusten www-sivustoilta. Rajausta tehtiin sillä perusteella, että sovelluksen www-sivulla yleensä esitellään itse konkordanssiohjelma ja sen toiminnot, ja lisäksi www-sivuston avulla ohjelmaa nähtävästi markkinoidaan halutulle kohdeyleisölle.

Tutkimuksessa rajoituttiin tarkastelemaan vain ilmaisia konkordanssiohjelmia, sillä kaupallisten konkordanssiohjelmien kartoitus ei ollut mahdollista rahallisista syistä. Vaikka useasta kaupallisesta konkordanssiohjelmasta on saatavilla ilmaisia kokeiluversioita, ne haluttiin jättää tämän tutkimuksen ulkopuolelle. Kokeiluversiot ovat usein käytössä vain rajatun ajan, ja jos käyttäjä haluaa jatkaa ohjelman käyttöä kokeiluajan umpeutumisen jälkeen, hänen on päivitettävä ohjelman maksulliseen versioon. Ilmaisista konkordanssiohjelmista valittiin lisäksi vain sellaiset, jotka voidaan ladata ja/tai asentaa kotikoneelle. Internetissä www-sivun kautta käytössä oleviin ilmaisiin konkordanssiohjelmiin ei siten tässä tutkimuksessa kiinnitetty huomiota. Lisäksi ilmaisista konkordanssiohjelmista valittiin vain yksikieliset konkordanssiohjelmat, sillä kaksi- tai monikielisten konkordanssiohjelmien kehitys näyttäisi olevan vielä melko rajoittunutta: kyseisiä ohjelmia löytyi verkosta vain kaksi, *ParaConc* (Athelstan 2012) ja *CasualPConc* (Imao 2012). Näistä *ParaConc* on kaupallinen konkordanssiohjelma, ja vaikka *CasualPConc* on ilmainen konkordanssiohjelma, se toimii vain Mac OS -käyttöjärjestelmässä (Imao 2012), joten siksi molemmat ohjelmat jätettiin tässä tutkimuksessa huomioimatta.

Ilmaisista konkordanssiohjelmista valittiin vain Windows-käyttöjärjestelmään suunnatut ohjelmat. Windows-käyttöjärjestelmä valittiin testausalustaksi kolmesta syystä. Ensinnäkin, useimmat ilmaiset konkordanssiohjelmat ovat saatavissa vain Windows-alustalle. Toiseksi useimmissa käännöstoimistoissa sekä kääntämistä opettavissa yliopistoissa Windows-käyttöjärjestelmä on yleensä ensisijainen käyttöympäristö. Kolmas syy

Windows-käyttöympäristön valintaan oli se, että tutkimuksen tekijän kotikoneella oli käytössään vain Windows Vista -käyttöjärjestelmä. Näin ollen sellaiset konkordanssiohjelmat, jotka toimivat vain Linux tai Mac OS -käyttöjärjestelmissä jätettiin kokonaan huomioimatta tässä tutkimuksessa. Ilmaisista konkordanssiohjelmissa karsittiin lopuksi pois sellaiset ohjelmat, joiden käyttöliittymää ei ollut saatavilla kokonaan englanniksi. Tästä syystä esimerkiksi *KHCoder* (Higuchi 2012) jätettiin tämän tutkimuksen ulkopuolelle, sillä kyseisen ohjelman englanninkielinen käyttöliittymä sisälsi suurimmaksi osaksi japaninkielistä tekstiä.

Tutkimuksessa tarkastellaan seuraavien ilmaisten, yksikielisten konkordanssiohjelmien toimintaa Windows-ympäristössä: *aConcorde*, *AdTAT*, *AntConc*, *KWIC Concordance for Windows*, *Multilingual Corpus Toolkit*, *Simple Concordance Program*, *TextSTAT* ja *WConcord*. Nämä konkordanssiohjelmat esitellään lyhyesti seuraavaksi.

4.5.1 aConCorde 0.4.2

aConcorde on Andrew Robertsin vapaa-ajallaan kehittämä avoimen lähdekoodin konkordanssiohjelma, joka on tehty erityisesti arabiankielisten tekstien prosessointia varten (Roberts 2012). Ohjelmassa on tuki usealle eri merkistölle (Unicode, ASCII, erilaiset arabiankielen koodausstandardit), ja lisäksi ohjelman käyttöliittymä on saatavilla englannin- ja arabiankielisinä. Ohjelma tarjoaa työkalut konkordanssien ja sanatiheyksien tutkimiseen teksteistä, jotka ovat doc-, rtf-, xml-, html- tai txt-muodossa. Lisäksi konkordanssien tulokset voidaan tallentaa ohjelman avulla txt- tai html-muotoon. (Roberts et al. 2005: 6–9.)

aConcorde on saatavilla ohjelman kotisivuilta zip-pakettina. *aConcorde* on kirjoitettu Javalla, joten se on ajettavissa missä tahansa käyttöjärjestelmässä, jossa Java Runtime Environment on asennettuna (Roberts 2012).

4.5.2 AdTAT 1.5

AdTAT on Adelaiden yliopiston kehittämä ja rahoittama ilmainen konkordanssiohjelma. Ohjelman avulla voidaan hakea konkordansseja ja kollokaatteja sekä luoda frevenssilistoja.

Lisäksi konkordanssi- ja kollokaattiotteet voidaan tulostaa tai tallentaa tekstitiedostoon. (University of Adelaide 2012.)

Ohjelma on ladattavissa zip-pakettina ohjelman kotisivuilta, mutta käyttöohjeita tai dokumentaatiota sivustolta ei löydy. Sen sijaan ohjelmasta löytyy helppokäyttöinen avustus. *AdTAT* toimii sekä Windows- että Mac OS -käyttöjärjestelmissä, edellyttäen että Java Runtime Environment on asennettu (University of Adelaide 2012).

4.5.3 AntConc 3.2.4

AntConc on japanilaisessa Wasedan yliopistossa työskentelevän Laurence Anthonyn kehittämä konkordanssiohjelma, jota voi käyttää henkilökohtaiseen, ei-kaupalliseen tutkimustarkoitukseen. Ohjelmaa voisi oikeastaan kuvata erittäin monipuoliseksi korpustyökaluksi, sillä se sisältää seitsemän erilaista työkalua: konkordanssityökalu, konkordanssiintymien työkalu, tiedostotyökalu, klusterityökalu, kollokaattityökalu, sanalistatyökalu sekä avain-sanalistatyökalu. Kukin työkalu on käytettävissä käyttöliittymässä omalta välilehdeltään. (Anthony 2012.)

AntConc on kehitetty Perl-ohjelmointikieltä käyttäen, ja ohjelma toimii Windows-, Linux- ja Mac OS -käyttöjärjestelmissä (Anthony 2012). *AntConc* ja siihen liittyvä dokumentaatio sekä käyttöohje ovat saatavilla ohjelman kotisivuilta.

4.5.4 KWIC Concordance for Windows 5.0

KWIC Concordance for Windows on japanilaisen Satoru Tsukamoton kehittämä yksinkertainen konkordanssiohjelma. Ohjelma tarjoaa seuraavat kolme toimintoa: sanalistat, konkordanssit ja kollokaatit. (Tsukamoto 2012.) Toimintoihin pääsee käsiksi ohjelman päänavigointipalkista tai klikkaamalla työkalupalkissa näkyviä ikoneita. Haetut konkordanssit voidaan lisäksi tallentaa ASCII-tekstinä, puolipiste-eroteltuna CSV-muodossa, tabulointieroteltuna TSV-muodossa tai merkityssä muodossa. Perustoimintojen lisäksi ohjelma tukee erilaisia merkintäkieliä, joilla pyritään erottamaan tekstin looginen rakenne tekstin varsinaisesta sisällöstä. Ohjelmassa on tuki yleisille SGML ja COCOA-merkintäkielille,

sekä Penn-Helsinki Parsed Corpus of Middle English -korpukseen sidotulle PPCME1-merkintätavalle. (Tsukamoto 2002: 327–328.)

Tsukamoton (2002: 328) mukaan *KWIC Concordance for Windows* -konkordanssiohjelman suunnittelussa on huomioitu erityisesti käytettävyys ja käytön helppous. Lisäksi ohjelma on räätälöity käytettäväksi 1990-luvun alussa kootun Helsingin korpuksen (The Helsinki Corpus of English Texts) kanssa (Rissanen 2004: 22). *KWIC Concordance for Windows* on saatavilla suoritettavana exe-tiedostona ohjelman kotisivuilta.

4.5.5 Multilingual Corpus Toolkit

Multilingual Corpus Toolkit (MLCT) on Scott Piao kehittämä konkordanssiohjelma, joka on ilmainen tutkimuskäytössä. Ohjelma tarjoaa useita työkaluja tekstikorpusten käsittelyyn, sillä ohjelmalla voidaan hakea sanojen konkordansseja ja kollokaatteja sekä tarkastella sanojen samankaltaisuuksia erilaisten tilastollisten kielimallien avulla. Lisäksi ohjelmalla voidaan tarkastella www-sivulta poimittua tekstiä. (Piao 2009.) Ohjelman nimi on hie- man harhaanjohtava, sillä se ei toimi kaksi- tai monikielisen konkordanssiohjelman tavoin siten, että käyttäjä pystyisi tarkastelemaan alkuperäistekstiä ja sen käännöstä esimerkiksi vertailutarkoituksessa.

MLCT on saatavilla ohjelman kotisivuilta zip-pakettina. *MLCT* on kirjoitettu Javalla, joten ohjelman voi ajaa missä tahansa käyttöjärjestelmässä, jossa Java Runtime Environment on asennettuna. Ohjelma tukee myös Unicode-merkistöä. (Piao 2009.)

4.5.6 Simple Concordance Program 4.09

Simple Concordance Program (SCP) on Alan Reedin kehittämä konkordanssiohjelma, jonka avulla voi hakea konkordansseja, luoda frekvenssilistoja ja tarkastella статистиikkoja. Konkordansseja ja sanalistoja voi myös järjestää eri tavoin, ja lisäksi molemmat voidaan tulostaa tai tallentaa tiedostoksi. (Reed 2012).

SCP sekä sen ohjetiedosto ja pikaopas ovat saatavilla ohjelman kotisivuilta zip-pakettina tai valmiina asennustiedostona. Kotisivuilta löytyy myös zip-pakettina oleva videotutoriaali ohjelman käytöstä. Ohjelma toimii sekä Windows- että Mac OS -käyttöjärjestelmissä (Reed 2012).

4.5.7 TextSTAT 2.9

TextSTAT on Matthias Hüningin kehittämä yksinkertainen tekstien analyysiohjelma, joka lukee paitsi tavallisia tekstitiedostoja, myös doc-, odt- ja html-muotoisia tekstejä. Lisäksi ohjelma tukee Unicode-merkistää, joten se voi lukea erikielisiä dokumentteja. (Freie Universität Berlin 2012.) *TextSTAT* tarjoaa seuraavat toiminnot: sanalistat, konkordanssit ja sitaattit. Kukin toiminto löytyy käyttöliittymästä omalta välilehdeltään. Lisäksi sanalistoja ja konkordansseja voi järjestää usealla eri tavalla.

TextSTAT on saatavilla ohjelman kotisivuilta zip-pakettina. Ohjelman kotisivulla löytyy myös linkit pdf-muodossa oleviin käyttöohjeeseen sekä pikaoppaaseen.

4.5.8 WConcord 3.0

WConcord (Concordancer for Windows 3.0) on yksinomaan Windows-käyttöjärjestelmälle suunnattu konkordanssiohjelma, jonka ovat kehittäneet Zdenek Martinek ja Leslie Siegrist vuonna 1998 (Technische Universität Darmstadt 1998). Ohjelman avulla voidaan hakea yksinkertaisen tai hieman monimutkaisemman hakuehdon mukaisia konkordansseja ja kollokaatteja sekä muodostaa frekvenssien mukaisia sanalistoja sekä konkordansseista että kollokaateista. Konkordanssien ja kollokaattien haut sekä frekvenssilistat voidaan tallentaa esimerkiksi laskentataulukko-ohjelman ymmärtämässä muodossa. (Bartsch 2010: 3–4.)

Bartschin (2010: 3) mukaan *WConcord* on suunniteltu mahdollisimman helppokäyttöiseksi huomioiden erityisesti sellaiset käyttäjät, joilla on vähän tai ei yhtään kokemusta tietokoneista. Ohjelmaa voi käyttää vapaasti kuka tahansa, kunhan mainitsee sen käytön mahdollisissa julkaisuissa. Ohjelma on saatavilla ladattavana zip-pakettina, ja paketin purettuaan käyttäjä voi suoraan ajaa ohjelman.

5 Havaintoja ilmaisista konkordanssiohjelmissa

Tässä tutkimuksessa kartoitettiin Windows-ympäristöön suunnattuja ilmaisia, yksikielisiä konkordanssiohjelmiä. Aineiston rajaamisen jälkeen tutkimuksen kohteeksi valittiin seuraavat kahdeksan verkosta löytynyttä ilmaista, Windows-käyttöjärjestelmässä toimivaa yksikielistä konkordanssiohjelmaa: *aConcorde 0.4.2*, *AdTAT 1.5*, *AntConc 3.2.4*, *KWIC Concordance for Windows 5.0*, *Multilingual Corpus Toolkit*, *Simple Concordance Program 4.09*, *TextSTAT 2.9* ja *WConcord 3.0*. Verkosta löydetty konkordanssiohjelmat ladattiin tai asennettiin Hewlett-Packard -merkkiselle kannettavalle tietokoneelle, johon oli asennettuna Windows Vista Home Premium -käyttöjärjestelmä sekä Service Pack 2 -käyttöjärjestelmäpäivitys. Ohjelmia testattiin yksi kerrallaan pienen, itse tehdyn tavallista tekstitiedostoista kootun englanninkielisen tekstikorpuksen kanssa, jotta saatiin selkeä kuva ohjelmien sisältämistä toiminnoista. Seuraavassa esitetään yhteenveto yleisimmistä ilmaisten konkordanssiohjelmien sisältämistä toiminnallisuuksista. Ilmaisten konkordanssiohjelmien versiotiedot ja niiden sisältämät toiminnot esitetään kootusti Liitteen 1 taulukossa.

Kuten voitiin olettaa, konkordanssitoiminto löytyi jokaisesta testatusta konkordanssiohjelmasta. Jokainen ohjelma myös mahdollisti konkordanssihaun järjestämisen yleensä joko avainsanan tai vasemman tai oikean kontekstin mukaan. Lisäksi jokainen konkordanssiohjelma mahdollisti konkordanssiotteen tallentamisen tavalla tai toisella. Konkordanssiohjelmissa *adTAT*, *AntConc*, *MLCT*, *SCP*, *TextSTAT* ja *WConcord* tarjosivat mahdollisuuden konkordanssihaun tallentamiseen tekstitiedostoksi. *aConcorde* tallensi konkordanssiotteen html-muodossa, jos käyttäjä hoksasi tallentaa tiedoston html-päätteiseksi. Lisäksi *KWIC for Windows* -konkordanssiohjelmassa konkordanssiotteen pystyi kopioimaan leikepöydälle ja sitä kautta tallentamaan tekstitiedostoksi.

Myös kollokaattien hakeminen onnistui lähes kaikilla tässä tutkimuksessa testatuilla konkordanssiohjelmissa, sillä *aConcorde* oli ainoa ohjelma, josta kyseinen toiminto puuttui. Toimintojen löytäminen ohjelmista oli kuitenkin haastavaa. Kahdessa ohjelmista (*SCP* ja *TextSTAT*) kollokaattitoiminto oli osa konkordanssitoimintoa. Muissa ohjelmissa kollokaattitoiminto oli toteutettu omana toimintonaan esimerkiksi omalla välilehdellä. Myös

kollokaattien esitystapa vaihteli ohjelmittain. *KWIC Concordance for Windows* ja *WConcord* esittivät haetun sanan kollokaatit taulukkomuodossa, *MLCT* esitti haun xml-muodossa ja *SCP*, *adTAT*, *TextTAT* ja *AntConc* esittivät kollokaattihaun tulokset luettelomuodossa konkordanssien tapaan. Kaikki ohjelmat, joista kollokaattitoiminto löytyi, mahdollistivat kollokaattien järjestämisen vasemman tai oikean kontekstin, kollokoivan sanan frekvenssin, avainsanan tai sen frekvenssin mukaan. Lisäksi ohjelmat mahdollistivat kollokaattihaun tallentamisen.

Jokainen tässä tutkimuksessa testattu konkordanssiohjelma sisälsi myös frekvenssilistojen luontitoiminnon. Lisäksi käyttäjä pystyi järjestämään frekvenssilistoja jokaisen ohjelman avulla yleensä aakkosjärjestykseen tai sanan frekvenssin mukaan. Testatuista ohjelmista kaikki muut paitsi *aConcorde* ja *adTAT* mahdollistivat frekvenssilistan tallentamisen: *TextSTAT* salli frekvenssilistan tallentamisen CVS- tai Excel-muodossa; *AntConc*, *MLCT*, *SCP* ja *WConcord* tallensivat frekvenssilistan tekstimuodossa, ja *KWIC for Windows* mahdollisti frekvenssilistan kopioinnin leikepöydälle. Ainoastaan *AntConc* ja *SCP* mahdollistivat avainsanalistan luomisen sekä estolistan käyttämisen frekvenssilistoja luotaessa.

Yllättävää oli, että vain yksi tässä tutkimuksessa testatuista konkordanssiohjelmista (*SCP*) tarjosi varsinaista tilastollista tietoa tai siihen liittyviä toimintoja. Tosin *KWIC for Windows* laski frekvenssilistan luomisen yhteydessä automaattisesti sanastollisen rikkauden indeksin, mutta ohjelma ei kuitenkaan tarjonnut muunlaista tilastotietoa. *SCP* laski myös kyseisen indeksin, mutta tarjosi tämän lisäksi koko korpuksen, korpuksen sanastoprofiiliin sekä korpuksessa esiintyvien kirjainten ja merkkien esiintymiseen liittyvää tilastotietoa. Lisäksi tilastotiedon pystyi muodostamaan erikseen joko kaikista korpuksen sanoista tai vain avainsanoista.

Yhdestä konkordanssiohjelmasta löytyi myös toimintoja, joita ei esiintynyt muista konkordanssiohjelmista. *AntConc* sisälsi ns. klusteritoiminnon, jonka avulla korpuksesta pystyi hakemaan klustereita eli vähintään kolmen toistuvasti yhdessä esiintyvän sanan muodostavia jonoja ja niiden frekvenssejä. Lisäksi samasta ohjelmasta löytyi ns. konkordanssiesiintymien tarkastelutoiminto, jolla haetun konkordanssin esiintymät esitettiin graafisessa muodossa tiedostokohtaisesti useista tiedostoista koostuvan korpuksen tapauksessa.

Ilmaisten konkordanssiohjelmien testaamisessa huomattiin, että ohjelmien välillä oli suuria eroja. Heikoin tässä tutkimuksessa kartoitetuista konkordanssiohjelmissa oli *aConCorde*, sillä ohjelma tarjosi vain konkordanssi- ja frekvenssilistatoiminnot, ja näistäkin vain konkordanssien tallennus onnistui jatkokäsittelyä varten. *aConCorde* ei kuitenkaan tarjonnut minkäänlaisia kollokaattien käsittelyyn tai statistiikkaan liittyviä toimintoja, ja lisäksi ohjelman käyttöliittymä vaikutti hyvin keskeneräiseltä.

Tässä tutkimuksessa kartoitetuista konkordanssiohjelmissa monipuolisimpia olivat *AntConc* sekä *Simple Concordance Program*. Ensinnäkin, molemmat konkordanssiohjelmat sisälsivät luvussa 3.2 esiteltyt kolme konkordanssiohjelmien yleisintä toimintoa. Lisäksi molemmat konkordanssiohjelmat täyttivät kaikki luvussa 3.3 kuvatut Barnbrookin (1996) ja Wilkinsonin (2012) esittämät hyvän konkordanssiohjelman tunnusmerkit. Lisäksi sekä *AntConc* että *SCP* huomioivat käyttäjänsä mahdolliset erityistarpeet, sillä molemmat ohjelmat mahdollistivat konkordanssien, frekvenssilistojen ja kollokaattien tallentamisen tekstitiedostoon jatkokäsittelyä varten. Lisäksi molemmat ohjelmat tarjosivat muista konkordanssiohjelmissa poiketen avainsanalistojen ja estolistojen käyttömahdollisuuden. Muun ohella *SCP* tarjosi monipuolisia statistiikkaan liittyviä toimintoja ja laski käsiteltävästä aineistosta automaattisesti sanastollisen rikkauden indeksin. Sekä *AntConc* että *SCP* olivat helppokäyttöisiä konkordanssiohjelmissa, sillä ne sisälsivät selkeän käyttöliittymän.

6 Yhteenveto

Tämän tutkimuksen tarkoituksena oli kartoittaa verkosta löytyviä, Windows-käyttöjärjestelmälle suunnattuja ilmaisia, yksikielisiä konkordanssiohjelmiä sekä niissä esiintyviä toimintoja, sekä selvittää konkordanssiohjelmien hyödyntämismahdollisuuksia kääntämisessä. Ensin lähdekirjallisuuteen perehtymällä esiteltiin korpuslingvistisen käännöstutkimuksen ja korpuksen käsitteet, sekä tarkasteltiin korpuksen elektronisen prosessin etuja ja haittoja sekä korpuksen hyödyntämismahdollisuuksia kääntämisessä. Lisäksi perehdyttiin konkordanssiohjelmien jaotteluun ja niiden yleisiin toimintoihin ja pohdittiin toimintojen hyödyntämistä kääntämisessä. Lähdekirjallisuuteen perehtymisen jälkeen konkordanssiohjelmiä etsittiin verkosta, ja löydetty ohjelmat asennettiin kannettavalle tietokoneelle. Tämän jälkeen konkordanssiohjelmien toimintaa testattiin Windows Home Premium -käyttöjärjestelmässä, ja samalla kirjattiin ylös ohjelmissa esiintyneet toiminnot. Lopuksi esitettiin tutkimuksen johtopäätökset.

Tässä tutkimuksessa tehdyn kartoitustyön tuloksena verkosta löytyi kahdeksan ilmaista, yksikielistä konkordanssiohjelmaa jotka toimivat Windows-ympäristössä. Lisäksi tutkimuksessa selvisi, että useimmat tässä tutkimuksessa testatut ilmaiset konkordanssiohjelmat mahdollistivat konkordanssien ja kollokaattien hakemisen, frekvenssilistojen muodostamisen sekä kaikkien edellä mainittujen tulosten tallentamisen. Sen sijaan tilastotietoon liittyviä toimintoja ilmaiset konkordanssiohjelmat tarjosivat vähän, sillä vain yksi konkordanssiohjelma (*SCP*) tarjosi varsinaista tilastotietoa käsiteltävänä olevasta korpuksesta. Yhdestä ilmaisesta konkordanssiohjelmasta (*AntConc*) löytyi myös sellaisia toimintoja, joita ei esiintynyt muissa testatuissa ohjelmissa. Tässä tutkimuksessa kartoitetuista konkordanssiohjelmissa heikoin ohjelma oli *aConCorde*, sillä se sisälsi vain muutamia hyvältä konkordanssiohjelmalta edellytettäviä ominaisuuksia. *AntConc* ja *Simple Concordance Program* sen sijaan erottuivat testattujen konkordanssiohjelmien joukosta edukseen niiden tarjoamien monipuolisten ominaisuuksiensa ansiosta.

Tämän tutkimuksen kirjallisuuskatsauksen perusteella näyttää siltä, että konkordanssiohjelmiä voidaan hyödyntää monin erin tavoin vieraan kielen kääntämisessä. Konkordansseja voidaan käyttää sanakirjan rinnalla, kun halutaan tarkastella sanan tai ilmauksen käyttöä

kontekstissaan tai saada kokonaiskuva sanan erilaisista käyttötavoista. Kollokaatteja voidaan hyödyntää tutkittaessa lähdetekstissä esiintyviä merkityksiä. Lisäksi konkordansseja ja kollokaatteja voidaan käyttää käännösvastineiden etsimiseen. Frekvenssilistojen avulla saadaan hyödyllistä tietoa käsiteltävänä olevan tekstin pääpiirteistä ja tekstissä yleisimmin esiintyvistä sanoista. Avainsanalista puolestaan auttaa kääntäjää selvittämään, mistä teksti kertoo ja mitä käsitteitä siinä käytetään, mihin erikoisalaan teksti sijoittuu, miten tekstissä liitetään sisältöjä toisiinsa ja millaisia metaforia tai epätyypillisiä sanoja tekstin kirjoittaja käyttää. Sanastollisen rikkauden indeksin avulla voidaan analysoida tekstin leksikaalista vaihtelevuutta vertailtaessa eri tekstityyppejä tai lähtö- ja kohdetekstejä keskenään.

Tätä tutkimusta tehdessä ilmeni myös monenlaisia jatkotutkimusideoita. Ilmaisten konkordanssiohjelmien testaamisessa huomattiin, että ohjelmien välillä on suuria eroja. Konkordanssiohjelmat eroavat suuresti paitsi käyttöliittymältään, myös esimerkiksi sen suhteen, millaisia korpuslingvistiikan käsitteitä ohjelmassa on käytetty, mitä toimintoja ne tarjoavat, miten ne käsittelevät useista tiedostoista koostuvaa korpusta, miten ohjelma asennetaan ja millaista dokumentaatiota on saatavilla. Näin ollen yksi jatkotutkimusidea olisi vertailla tässä tutkimuksessa kartoitettuja ilmaisia konkordanssiohjelmiä keskenään ja selvittää, mikä ohjelmista olisi paras kääntäjien näkökulmasta. Lisäksi tässä tutkimuksessa kartoitettuja ilmaisia, yksikielisiä konkordanssiohjelmiä voisi vertailla kaupallisiin, monikielisiin konkordanssiohjelmiin ja esimerkiksi selvittää, millaisia eroja ohjelmien välillä on ja kumpi ohjelmatyyppi sopii paremmin kääntäjien käytännön työhön.

Yksi tämän tutkimuksen heikkous on se, että tässä tutkimuksessa keskityttiin kartoittamaan vain Windows-ympäristöön suunnattuja ilmaisia, yksikielisiä konkordanssiohjelmiä. Eräs jatkotutkimusidea olisikin kartoittaa myös Linux- tai Mac OS -ympäristöön suunnattuja ilmaisia konkordanssiohjelmiä ja niiden toimintoja. Tämän kandidaatintutkielman kirjoitushetkellä Mac OS -käyttöjärjestelmälle on saatavilla ainakin kolme yksikielistä konkordanssiohjelmaa: *CasualConc*, *Conc* ja *Concorder Pro*, tosin kahta jälkimmäistä ei enää ilmeisesti päivitetä. Tässä tutkimuksessa testatuista konkordanssiohjelmissa *AntConc* ja *TextSTAT* toimivat kehittäjiensä mukaan Linuxissa, ja *adTAT* toimii Mac OS -ympäristössä. Eri ympäristöön sopivien konkordanssiohjelmien testaaminen ja arviointi olisi hyödyllistä, sillä niiden käyttäjillä voi olla käytössään hyvinkin erilaiset tietokoneet ja

käyttöjärjestelmät, eivätkä tässä tutkimuksessa kartoitetut Windows-pohjaiset konkordanssiohjelmat siten sovi välttämättä kaikille.

Toinen tämän tutkimuksen heikkous on se, ettei ilmaisten konkordanssiohjelmien sopivuutta voitu tämän tutkimuksen puitteissa testata käytännön käännoistyössä. Toinen jatkotutkimusidea olisikin testata ilmaisten, yksikielisten konkordanssiohjelmien toimintaa kääntämisessä tai kääntämisen opiskelussa. Lisäksi tässä tutkimuksessa kartoitettuja ilmaisia konkordanssiohjelmiä voisi arvioida niiden käytettävyyden ja pedagogisen käytettävyyden kannalta, koska tätä näkökulmaa ei ole otettu huomioon tässä tutkimuksessa eikä myöskään missään aiemmassa konkordanssiohjelmiä arvioivassa tai kartoittavassa tutkimuksessa. Opetuskäyttöön suunnatun sovelluksen pedagogisen käytettävyyden arviointi olisi tärkeää, jotta opettajat osaisivat valita sopivimmat työkalut ottaessaan korpuksen käytön osaksi opetustaan. Konkordanssiohjelmiä tai muita korpustyökaluja voisi myös arvioida kääntäjän näkökulmasta, jotta saataisiin tietoa siitä, miten konkordanssiohjelmaa voitaisiin hyödyntää esimerkiksi käännösongelmien ratkaisemisessa. Lisäksi voisi verrata esimerkiksi uusimpien konkordanssiohjelmien ja käännösmuistiohjelmien sopivuutta kääntämisessä, kuten ovat tehneet Bowker ja Barlow (2004).

Edellä mainittujen ohjelmien käyttöön, käytettävyyteen tai käyttötarkoituksen sopivuuteen liittyvien jatkotutkimusideoiden lisäksi voisi kartoittaa sellaisia tekijöitä, joita konkordanssiohjelmien kehittämisessä tulisi ottaa huomioon, ja joita nykyiset konkordanssiohjelmat eivät vielä tue. Tällaisten tekijöiden kartoittaminen on tärkeää, jotta konkordanssiohjelmiä voitaisiin edelleen kehittää sellaisiksi, että niitä voitaisiin paremmin hyödyntää kääntämisessä tai kääntämisen opettamisessa.

Lähteet

Konkordanssiohjelmat

Anthony, L. 2012. *AntConc*. Saatavilla osoitteesta

http://www.antlab.sci.waseda.ac.jp/antconc_index.html (viitattu 12.12.2012).

Freie Universität Berlin 2012. *TextSTAT - Simple Text Analysis Tool*. Saatavilla osoitteesta

<http://neon.niederlandistik.fu-berlin.de/en/textstat/> (viitattu 12.12.2012).

Piao, S. 2009. *Multilingual Corpus Toolkit*. Saatavilla osoitteesta

<https://sites.google.com/site/scottpiaosite/software/mlct> (viitattu 12.12.2012).

Reed, A. 2012. *Simple Concordance Program*. Saatavilla osoitteesta

<http://www.textworld.com/scp/index.html> (viitattu 12.12.2012).

Roberts, A. 2012. *aConCorde*. Saatavilla osoitteesta [http://www.andy-](http://www.andy-roberts.net/coding/aconcorde)

[roberts.net/coding/aconcorde](http://www.andy-roberts.net/coding/aconcorde) (viitattu 12.12.2012).

Technische Universität Darmstadt 1998. *WConcord 3.0 (Concordancer for Windows)*.

Saatavilla osoitteesta <http://www.linglit.tu-darmstadt.de/index.php?id=linguistics> (viitattu 12.12.2012).

Tsukamoto, S. 2012. *KWIC Concordance for Windows Ver.5*. Saatavilla osoitteesta

http://www.chs.nihon-u.ac.jp/eng_dpt/tukamoto/kwic_e.html (viitattu 12.12.2012).

University of Adelaide 2012. *AdTAT - the Adelaide Text Analysis Tool*. Saatavilla osoitteesta

<http://www.adelaide.edu.au/red/adtat/> (viitattu 12.12.2012).

Kirjallisuus

Anagnostou, N. K & G. R. S. Weir 2006. *Review of software applications for deriving collocations*. ICT in the Analysis, Teaching and Learning of Languages, Preprints of the IC-TATLL Workshop 2006, Glasgow. 91–100. Saatavilla PDF-muodossa

http://www.cis.strath.ac.uk/cis/research/publications/papers/strath_cis_publication_1541.pdf (viitattu 18.4.2010).

- Anthony, L. 2012a. *AntConc (Windows, Macintosh OS X, and Linux) Build 3.2.4*. Saatavilla PDF-muodossa http://www.antlab.sci.waseda.ac.jp/software/README_AntConc3.2.4.pdf (viitattu 14.12.2012).
- Antworth, E. L. & J. R. Valentine 1998. Software for doing field linguistics. Teoksessa Lawler, J. & A. D. Helen (eds.). *Using computers in linguistics: a practical guide*. London: Routledge.
- Athelstan 2012. *Concordancers and Corpora*. Saatavilla HTML-muodossa <http://www.athel.com/mono.html> (viitattu 14.12.2012).
- Ari, O. 2006. Review of Three Software Programs Designed to Identify Lexical Bundles. *Language Learning & Technology*, 10(1), 30–37. Saatavilla PDF-muodossa <http://llt.msu.edu/vol10num1/pdf/review3.pdf> (viitattu 18.4.2010).
- Baker, M. 1993. Corpus linguistics and translation studies: implications and applications. Teoksessa Baker, M., Francis, G. & E. Tognini-Bonelli (eds.). *Text and Technology: in honour of John Sinclair*. Amsterdam: John Benjamins. 17–45.
- Baker, M. 1995. Corpora in Translation Studies: An overview and some suggestions for future research. *Target* 7(2), 223–243.
- Baker, M. (ed.) 1998. *Routledge encyclopedia of translation studies*. London: Routledge.
- Barnbrook, G. 1996. *Language and computers: a practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Bartsch, S. 2010. *Manual and Help File for Concordancer for Windows 3.0 aka WConcord*. Saatavilla PDF-muodossa <http://www.linglit.tu-darmstadt.de/fileadmin/linglit/bartsch/corptool/WCONCORD.pdf> (viitattu 19.4.2010).
- Bauer, C. 1992. Conc. *Literary and Linguistic Computing*, 7(2), 154–156.
- Bergenholtz, H & S. Tarp (eds.) 1995. *Manual of specialised lexicography: the preparation of specialised dictionaries*. Amsterdam: John Benjamins.

Biber, D., S. Conrad & R. Reppen 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Birmingham City University 2012. *WebCorp Live. Concordance the web in real-time*. Saatavilla HTML-muodossa <http://www.webcorp.org.uk/live/> (viitattu 14.12.2012).

Bloch, J. 2009. The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology*, 13(1), 59–78. Saatavilla PDF-muodossa <http://llt.msu.edu/vol13num1/bloch.pdf> (viitattu 18.4.2010).

Bowker, L & M. Barlow 2004. *Bilingual concordancers and translation memories: a comparative evaluation*. LRTWRT '04 Proceedings of the Second International Workshop on Language Resources for Translation Work, Research and Training, 70-79. Saatavilla PDF-muodossa http://dl.acm.org/ft_gateway.cfm?id=1708096&type=pdf&CFID=155702429&CFTOKEN=99959617 (viitattu 11.12.2012).

Bowker, L & J. Pearson 2002. *Working with specialized language: a practical guide to using corpora*. London: Routledge.

Bowker, L. 2002. *Computer-Aided Translation Technology: A Practical Introduction*. Canada: University of Ottawa Press.

Daelemans, W & V. Hoste 2008. Evaluation of translation technology. Teoksessa Vandepitte, S (ed.). *Looking for meaning: methodological issues in translation studies*. *Linguistica Antverpiensia - New Series*, 7. Antwerpen: Artesis. 9–17. Saatavilla PDF-muodossa <http://lt3.hogent.be/media/uploads/publications/2009/Daelemans2009.pdf> (viitattu 11.12.2012).

Diniz, L. 2005. Comparative Review: TextSTAT 2.5, AntConc 3.0, and Compleat Lexical Tutor 4.0. *Language Learning & Technology*, 9(3), 22–27, ISSN1094-3501. Saatavilla PDF-muodossa <http://llt.msu.edu/vol9num3/pdf/review2.pdf> (viitattu 18.4.2010).

Eskola, S. 2002. *Syntetisoivat rakenteet käännössuomessa: suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmällä*. Joensuun yliopiston humanistisia julkaisuja n:o 30. Joensuu: Joensuun yliopisto.

Evans, D. 2012. Unit 1: Introduction. *Introduction to Corpus investigative techniques: an on-line information pack about corpus investigation techniques for the Humanities*. University of Birmingham. Saatavilla PDF-muodossa

<http://www.birmingham.ac.uk/Documents/college-artslaw/corpus/Intro/Unit1.pdf> (viitattu 10.12.2012).

Forchini, P. 2005. Concorde Pro 1.0 a Text-analysis Tool for Mac OS X. *The Electronic Journal for English as a Second Language*, 9(1), 1–8. Saatavilla PDF-muodossa <http://tesl-ej.org/ej33/m1.pdf> (viitattu 18.4.2010).

Gawronska, B. 2008. *Corpora and Concordances in Human Translation, Machine Translation, and Language/Translator Education*. Luentokalvot. Norja: Dept. of Foreign Languages and Translation, University of Agder. Saatavilla PDF-muodossa www.sprakradet.se/5361 (viitattu 14.12.2012).

Hunston, S. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Higuchi, K. 2012. *KH Coder*. Saatavilla osoitteesta <http://khc.sourceforge.net/en/> (viitattu 15.12.2012).

Imao, Y. 2012. *CasualPConc*. Saatavilla HTML-muodossa

<https://sites.google.com/site/casualconc/utility-programs/casualpconc> (viitattu 14.12.2012).

Kemppanen, H. 2008. Keywords revisited: lähdetekstin korpuspohjainen analyysi kääntäjän apuvälineenä. Suomen kääntäjien ja tulkkien liitto. *MikaEL. Kääntämisen ja tulkkausten tutkimuksen symposiumin verkkojulkaisu*, Vol. 2. Saatavilla PDF-muodossa

<http://www.sktl.fi/@Bin/41020/Kemppanen.pdf> (viitattu 14.12.2012).

Kenny, D. 2001. *Lexis and creativity in translation: a corpus-based study*. Manchester: St. Jerome Publishing.

- Koskenniemi, K. 2004. *Terms and concepts of language technology*. Saatavilla HTML-muodossa <http://www.ling.helsinki.fi/kit/2004s/terms-en.shtml> (viitattu 14.4.2010).
- Kübler, N. 2011. Working with corpora for translation teaching in a French-speaking setting. Teoksessa Frankenberg-Garcia, A., Flowerdew, L. & G. Aston (eds.). *New trends in corpora and language learning*. London: Continuum. 62–80.
- Laviosa, S. 2002. *Corpus-based translation studies: Theory, findings, applications*. Amsterdam: Rodopi.
- McEnery, T., Xiao, R. and Y. Tono 2006. *Corpus-based language studies: an advanced resource book*. London: Routledge.
- O'Donnell, M. B. 2008. KWICgrouper - Designing a tool for corpus-driven concordance analysis. *International Journal of English Studies*, 8(1), 107–121. Saatavilla PDF-muodossa, <http://revistas.um.es/ijes/article/download/49121/46991> (viitattu 11.12.2012).
- Olohan, M. 2004. *Introducing corpora in translation studies*. London: Routledge.
- Reppen, R. 2001. Review of MonoConc Pro and WordSmith Tools. *Language Learning & Technology*, 5(3), 32–36. Saatavilla PDF-muodossa <http://llet.msu.edu/vol5num3/pdf/review4.pdf> (viitattu 18.4.2010).
- Rissanen, M. 2004. Englannin kielen tutkimusta tietokoneiden aikaan. *Tieteessä tapahtuu*, 5, 19–23. Saatavilla PDF-muodossa <http://www.tieteessatapahtuu.fi/0504/rissanen.pdf> (viitattu 19.4.2010).
- Roberts, A., L. Al-Sulaiti & E. Atwell 2005. *aConCorde: towards a proper concordance for Arabic*. Saatavilla PDF-muodossa http://www.andy-roberts.net/research/papers/cl05_aConCorde.pdf (viitattu 18.4.2010).
- Scott, M. 2008. Developing WordSmith. *International Journal of English Studies*, 8(1), 95–106. Saatavilla PDF-muodossa <http://dialnet.unirioja.es/servlet/articulo?codigo=2729155> (viitattu 14.4.2010).

- Smith, N., S. Hoffmann & B. Rayson 2008. Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations. *Literary and Linguistic Computing*, 23(2), 163–180. Saatavilla HTML-muodossa <http://llc.oxfordjournals.org/cgi/content/full/fqn004v1> (viitattu 18.4.2010).
- Stevens, V. 2002. Concordance. *CALICO Journal*, 19(3), 690–708. Saatavilla HTML-muodossa [https://www.calico.org/p-180-Concordance\(62001\).html](https://www.calico.org/p-180-Concordance(62001).html) (viitattu 18.4.2010).
- Tsukamoto, S. 2002. KWIC Concordance for Windows: Easy Access to Corpora. Teoksessa T. Saito, J. Nakamura & S. Yamazaki (eds.). *English Corpus Linguistics In Japan*. Language and Computers: Studies in Practical Linguistics No. 38. Amsterdam & New York: Rodopi, 327–340.
- University of Oxford 2010. *British National Corpus*. Saatavilla HTML-muodossa <http://www.natcorp.ox.ac.uk/> (viitattu 14.12.2012).
- Watt, R. J. C. 2010. *Concordance*. Saatavilla HTML-muodossa <http://www.concordancesoftware.co.uk/index.htm> (viitattu 14.4.2010).
- Wiechmann, D. & S. Fuhs 2006. Concordancing software. *Corpus Linguistics & Linguistic Theory*, 2(1), 107–127.
- Wilkinson, M. 2012. The best freeware corpus analysis program for translators? *Translation Journal*, 16(2). Saatavilla HTML-muodossa <http://www.bokorlang.com/journal/60corpus.htm> (viitattu 5.12.2012).
- Wilkinson, M. 2006. *The corpus analysis tool – an under-exploited translation aid*. Saatavilla DOC-muodossa http://www.lexically.net/wordsmith/corpus_linguistics_links/Wilkinson.doc (viitattu 14.12.2012).
- Williamson, G. 2009. *Type-Token Ratio*. Saatavilla HTML-muodossa <http://www.speech-therapy-information-and-resources.com/type-token-ratio.html> (viitattu 15.12.2012).

Liitteet

A Ilmaisten konkordanssiohjelmien tiedot ja toiminnot

Tieto / toiminto	aConcorde	adTAT	AntConc	KWIC for Win	MCLT	SCP	Text STAT	WConcord
Versio	0.4.2	1.5	3.2.4	5.0	?	4.09	2.9	3.0
Julkaisuvuosi	2008	2012	2012	2012	2009	2010	2012	1998
Konkordanssit	x	x	x	x	x	x	x	x
Konkordanssien järjestäminen	x	x	x	x	x	x	x	x
Konkordanssien tallennus	x	x	x	x	x	x	x	x
Frekvenssilistat	x	x	x	x	x	x	x	x
Frekvenssilisto- jen järjestäminen	x	x	x	x	x	x	x	x
Frekvenssilisto- jen tallennus	-	-	x	x	x	x	x	x
Avainsanalistat	-	-	x	-	-	x	-	-
Estolistat	-	-	x	-	-	x	-	-
Kollokaatit	-	x	x	x	x	x	x	x
Kollokaattien järjestäminen	-	x	x	x	x	x	x	x
Kollokaattien tallennus	-	x	x	x	x	x	x	x
Statistiikka	-	-	-	-	-	x	-	-
Sanastollisen rikkauden indeksi	-	-	-	x	-	x	-	-