

**Suomen ja tšekin sanaparin YSTÄVÄ – PŘÍTEL merkitykset ja
kollokaatit**

Maisterintutkielma

Markéta Sovová

Suomen kieli

Kieli- ja viestintätieteiden laitos

Jyväskylän yliopisto

Syksy 2021

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistis-yhteiskuntatieteellinen tiedekunta	Laitos – Department Kieli- ja viestintätieteiden laitos
Tekijä – Author Markéta Sovová	
Työn nimi – Title Suomen ja tšekin sanaparin YSTÄVÄ – PŘÍTEL merkitykset ja kollokaatit	
Oppiaine – Subject Suomen kieli	Työn laji – Level Pro gradu -tutkielma
Aika – Month and year Joulukuu 2021	Sivumäärä – Number of pages 82 sivua + liitteet
<p>Tiivistelmä – Abstract</p> <p>Maisterintutkielmassani tutkin suomen ja tšekin lekseemien YSTÄVÄ/PŘÍTEL kontekstuaalista vastavuutta. Ensin tarkastelen käännöskorpuksen avulla missä määrin lekseemit ovat toistensa käännösvastineita. Seuraavaksi käsittelen sanojen eri merkityksiä ja niiden yleisyyttä kummassakin kielessä. Analyysin kolmannessa vaiheessa havainnoin lemموjen kollokaatteja muun muassa semanttisesta näkökulmasta. Kollokaatit ovat ilmauksia, jotka tyypillisesti esiintyvät toisen läheisyydessä, ja kollokaation ominaisuudeksi yleensä määritellään niiden toistuvuus, tilastollinen merkitsevyys ja ulottuvuus (Jantunen 2004: 16). Kollokaatiotutkimus on tärkeää esimerkiksi kielenoppimisessa, sillä edistyneilläkin kielenopijoilla on vaikeuksia tuottaa sopivia kollokaatteja (ks. Nesselhauf 2003: 237).</p> <p>Tutkielmani pohjautuu etenkin kontekstuaalisen semantiikan ja fraseologian teoriaviitekehyksiin, jotka korostavat sanojen kontekstisidonnaisuutta. Käyttämäni aineisto ja tutkimusmenetelmät perustuvat korpustutkimukseen. Sekä tšekin että suomen puolesta on tarjolla laajoja korpusaineistoja ja käännösanalyysissäni hyödynnän kaksisuuntaista suomi–tšekki InterCorp 11 -korpusta. Merkitys- ja kollokaatioanalyysiä varten laadin sanomalehtitekstejä sisältävän verrannollisen korpuksen koostuen Suomen kielen tekstikokoelma- ja tšekin SYN2006PUB-korpuksista. Aineistosta haen merkitsevimmät kollokaatit tilastollisin menetelmin.</p> <p>Lekseemien YSTÄVÄ/PŘÍTEL keskinäisen käännösvastavuuden lasketaan olevan 85 %. 'Läheinen ihminen'-perusmerkityksessä sanat esiintyvät lähes yhtä usein. Toiseksi yleisin merkitys on YSTÄVÄ-lekseemillä 'harrastaja, kannattaja' ja PŘÍTEL-lekseemillä toisaalta 'puoliso'. Molemmat sanat jakavat n. 40 % yhteisiä kollokaatteja. Kielten yhteiset kollokaattimerkitysryhmät ovat 'ystävyyssuhde', 'tapaaminen', 'tapahtuma' ja 'ihmiset'. Suomen YSTÄVÄ ilmenee huomattavasti useammin 'harrastus-' ja 'tunne' semanttisissa ympäristöissä. PŘÍTEL-sanon kollokaatit puolestaan kuuluvat useammin 'parisuhde-', 'ammatti'- ja 'rikos' -merkitysryhmiin.</p> <p>Tutkimustuloksia voi hyödyntää sekä vieraan kielen oppimisessa että käännöstieteessä. Tutkielmani lisäksi edistää melko harvinaista suomen ja tšekin vertailevaa fraseologista tutkimusta.</p>	
Asiasanat – Keywords: korpuslingvistiikka, kontrastiivinen lingvistiikka, fraseologia, kollokaatio	
Säilytyspaikka – Depository: Jyväskylän yliopisto	
Muita tietoja – Additional information	

SISÄLLYS

1	JOHDANTO.....	1
1.1	Tutkimuksen aihe ja taustoitus.....	1
1.2	Tutkimusasetelma ja tutkimuskysymykset.....	2
1.3	Aiempi tutkimus.....	4
2	TEORIATAUSTA.....	6
2.1	Kontekstuaalinen semantiikka.....	7
2.1.1	Periaatteet ja lähtökohdat.....	7
2.1.2	Kontekstuaalinen valinta.....	9
2.2	Fraseologia.....	12
2.2.1	Fraseologinen yksikkö.....	12
2.2.2	Kollokaatio.....	14
2.3	Sanojen merkityssuhteet.....	17
3	AINEISTO JA MENETELMÄT.....	21
3.1	Aineisto.....	24
3.1.1	Lekseemien valinta ja perustelut.....	26
3.1.2	Suomen ja tšekin sanakirjat.....	29
3.1.3	Korpusaineisto.....	30
3.1.3.1	Rinnakkaiskorpus InterCorp.....	30
3.1.3.2	Verrannollinen korpus.....	32
3.2	Korpuslingvistiset menetelmät.....	33
3.2.1	Tilastolliset testit ja kollokaatioväli.....	33
3.2.2	Aineiston muodostaminen.....	36
3.2.3	Käsittelyohjelmiin liittyvät ongelmat.....	37
4	ANALYYSI.....	39
4.1	Käännösten vastaavuus.....	39
4.2	Merkitysjako ja polyseemisyys.....	45
4.2.1	Merkitysjako sanakirjoissa.....	46
4.2.2	Merkitysjako verrannollisessa korpuksessa.....	48
4.3	Kollokaatioanalyysi.....	56
4.3.1	MI-testin kollokaattien luokittelu.....	57
4.3.2	Tilastollisesti merkitsevimmät kollokaatit.....	61
4.3.3	Semanttinen analyysi.....	66
5	PÄÄTÄNTÖ.....	79

LÄHTEET.....	83
LIITTEET	89

1 JOHDANTO

1.1 Tutkimuksen aihe ja taustoitus

Vierasta kieltä opiskeltaessa kielenoppija todella usein joutuu tilanteeseen, jossa hän käyttää opittua sanaa väärin, vaikka sana vastaisikin oman äidinkiелensä käännösvastinetta. Syynä on usein se, että vaikka sanat vastaavat toisiaan ja ovat esimerkiksi sanakirjassa ekvivalentit, niitä voi kuitenkin eri kielissä käyttää eri konteksteissa, yhdistää erilaisten lekseemien kanssa ja niillä voi olla eri sävyjä sekä lisämerkityksiä. Tällaisia merkitysvivahteita ei ole helppoa oppia, tunnistaa tai opettaa, eikä sanakirja yleensä auta paljoa. Pienet merkitys- ja konnotaatioerot koskevat myös lähisynonyynejä yhden kielen puitteissa, puhumattakaan sitten sanoista useamman kielen välillä (Altenberg & Granger 2002: 21).

Maisterintutkielmassani tutkin suomen ja tšekin sanaparin YSTÄVÄ – PŘÍTEL merkityksiä ja kollokaatteja. Kollokaateilla tarkoitetaan sanan läheisessä tekstiympäristössä, eli kontekstissa, toistuvasti esiintyviä sanoja. Termi *kollokaatio* liittyy fraseologiaan, jonka yhtenä perusajatuksena on, että sanat muodostavat merkityksiä vasta ollessaan toistensa kanssa yhteydessä (Sinclair 2004: 28–29). Tutkielmassani perehdyn lekseemien YSTÄVÄ ja PŘÍTEL eri merkityksiin ja tutkin, millaiset kollokaatit kummankin ympäristössä tyypillisesti esiintyvät. Samalla tarkastelen valitsemani lekseemien yhtäläisyyksiä ja eroja sekä selvitän, missä määrin ne vastaavat toisiaan.

Tšekin ja suomen kieltä vertailevaa fraseologista tutkimusta on tehty melko vähän, joten tutkimukseni voisi avata kahden eri kieliryhmään kuuluvan kielen suhdetta mikrotasolla. Maisterintutkielmani on myös relevantti vieraan kielen oppimisen kannalta, sillä se korostaa sanojen kontekstisidonnaisuutta merkitysten selittämisessä. Kielitieteessä on, osittain korpuslingvistiikan ansiosta, jo kymmeniä vuosia vallinnut näkemys, että fraseologiset yksiköt, eli vakiintuneet monisanaiset myötäesiintymät, edustavat huomattavaa osaa kielestä sen sijaan, että kieli koostuisi yksittäisistä lekseemeistä, joita kielenkäyttäjät aina uudestaan liimaa yhteen (Nesselhauf 2005: 1–2). Kollokaatit ovat kuitenkin vaikeita oppia edistyneemmillekin oppijoille, ja näitä saatetaan käyttää epäluontevasti, joten kontekstisidonnaisuus ja fraseologia ovat olleet kielenopetuksessa korostettuja viimeaikaisten tutkimusten seurauksena (Kennedy 2008: 36–37). Vaikka englannin kielen opetuksessa on nykyään tavanomaista kohdata korpusten ja korpuslähteisten opetusmateriaalien hyödyntämistä, tilanne on erilainen, kun kyseessä on pienempi ja harvemmin opetettu kieli kuten suomi (Jantunen 2016: 98).

Tutkielmassani päätin vertailla suomen kielen sanaa YSTÄVÄ ja sen tšekin vastinetta PŘÍTEL, sillä ne kuuluvat kielten perussanastoon. Jo ensimmäisillä oppitunneilla kielenoppija kohtaa perheeseen ja ihmisen lähipiiriin kuuluvia sanoja. Kyseiset sanat ovat siis oleellisia, ja ne ovat taajuussanastojen mukaan varsin frekventtejä kummassakin kielessä. Tämä pätee myös sanomalehtikieleen, josta tutkielmani aineisto on muodostettu.

Lekseemien YSTÄVÄ ja PŘÍTEL väliltä löytyy selkeitä yhteyksiä jo ensisilmäyksellä. Koska sanojen perusmerkityksessä viitataan läheiseen ja luotettavaan ihmiseen, herättävät sanat molemmassa kielessä myönteistä kaikua. Syntaktisestikin molempia sanoja käytetään samankaltaisella tavalla: yleensä lauseen subjektina eli tekijänä tai lauseen objektina eli kohteena, jonka tapauksessa YSTÄVÄ usein esiintyy adpositiorakenteessa (*ystävän kanssa/s přítelem*). Tämän tutkimuksen aiheena on siten sanojen YSTÄVÄ/PŘÍTEL kollokaattien ja merkitysten yhtäläisyyksien ja erojen vertailu.

Tässä luvussa esitän tutkimuskysymykset ja aiheeseen liittyvää aiempaa tutkimusta. Luvussa 2 keskityn tutkielmani kannalta keskeisimpiin kontekstuaalisen semantiikan piirteisiin ja sanojen merkityssuhteisiin. Luvussa 3 kuvailen tarkemmin käyttämiäni korpuksia ja soveltamiani korpuslingvistisiä menetelmiä. Pehdyn myös fraseologisen ja kieliä vertailevan tutkimuksen periaatteisiin. Neljäs luku koostuu sanojen käänkösvastaavuuden testaamisesta, niiden polyseemisuuden tutkimisesta sekä kollokaatioanalyysistä. Viimeisessä luvussa pohdin analyysin tuloksia ja arvioin tutkimuksen onnistumista kokonaisuutena.

1.2 Tutkimusasetelma ja tutkimuskysymykset

Tutkielmani on korpuslingvistinen tutkimus, sillä korpus tarjoaa hyvän pohjan tutkia kieltä sen luonnollisessa ympäristössä, ja saatavilla on laajoja aineistoja sekä suomen että tšekin kielestä. Hyödynnän ensin käänköskorpusta InterCorp 11 saadakseni näkemyksen siihen, kuinka usein YSTÄVÄ ja PŘÍTEL vastaavat toisiaan käänköksessä. Merkitys- ja kollokaatioanalyysiä varten käytän kokoamaani verrannollista korpusta, joka koostuu suomen ja tšekin alkuperäisistä teksteistä, jotka ovat verrattavissa toisiinsa tekstilajin, ajan ja rekisterin osalta. Suomenkieliseksi aineistoksi valitsin kooltaan lähes 150 miljoonan sanan laajuisen SKTP-korpuksen (Suomen kielen tekstikokoelma), joka sisältää pääosin sanomalehtitekstejä 1990-luvulta. Tšekin puolelta käytän SYN2006PUB-korpusta, joka vastaavasti koostuu journalistisista teksteistä vuosilta 1989–2004 ja sisältää n. 300 miljoonaa sanaa.

Jotta voisi mielekkäästi vertailla sanojen kollokaatiota, kannattaa ensin pohtia niiden keskinäistä vastaavuutta ja niiden toisistaan eroavia merkityksiä. Näin vahvistetaan, että verrataan ylipäänsä verrattavissa olevia asioita. YSTÄVÄ- ja PŘÍTEL-lekseemien merkitysten tarkastelun jälkeen analysoin niiden kollokaatteja ja vertailen niitä keskenään. Olen muotoillut tutkimuskysymykseni seuraavasti:

- 1 Kuinka usein suomen ja tšekin sanat YSTÄVÄ/PŘÍTEL ovat käänkösvastineita rinnakkaiskorpuksessa?
- 2 Millaisia eri merkityksiä sanoilla on, ja kuinka yleisiä nämä merkitykset ovat?
- 3 Millaisia kollokaatteja kunkin sanan kontekstissa esiintyy ja kuinka ne vastaavat toisiaan sanomalehtiteksteissä?

Kahta sanaa, joilla on sama merkitys (esimerkiksi sanat AALTO ja LAINE) kutsutaan synonyymeiksi. Täydellisestä synonyymiasta voisi puhua siinä tapauksessa, että kahden sanan kaikki esiintymisympäristöt olisivat täysin samat, eli sanat vastaisivat toisiaan sekä denotatiiviselta että konnotatiiviselta merkitykseltään. (Kuiri 2012: 30.) Koska sanat eivät oletettavasti koskaan ole täysin keskenään vaihdettavia, eikä absoluuttista synonymiaa kielestä löydy, käytetään useammin termiä lähisyronyymit (Jantunen 2004: 1, 54; Inkpen & Hirst 2006: 223).

Kun otetaan vertailuun toisen kielen vastine, on tilanne vielä monimutkaisempi. Altenberg ja Granger perustelevat, että historiallisista, sosiaalisista ja kulttuurisista syistä konseptuaalinen maailma kehittyä eri kielissä eri tavalla (Altenberg & Granger 2002: 21). Sen seurauksena kielten merkitysjärjestelmät on jaoteltu toisistaan eroavasti. Joihinkin eroihin voi löytyä kulttuurisia syitä, esimerkiksi useat eri nimet kamelille arabian kielessä, siinä missä suomeksi riittää yksi. Toiset erot eivät kuitenkaan selity kulttuurien erilaisuudella, kuten esimerkiksi suomessa käytetään samaa pronominia HÄN sekä naisista että miehistä, kun taas monessa muussa kielessä sukupuolilla on omat pronominit. (Kuiri 2012: 22.) Useamman kielen vertailussa ei siis puhuta synonyymeistä, vaan eri kielen sanojen tai rakenteiden vastaavuudesta (*equivalence*) ja käänkösvastineista. Kuten absoluuttista synonymiaakin on vaikeaa löytää, niin myös eri kielten sanojen välinen täysi vastaavuus on hyvin harvinaista. Näin ollen sanoille, jotka ovat kaksikielisissä sanakirjoissa listattuna toistensa vastineiksi, löytyy usein erilaisia merkitysalueita. (Altenberg & Granger 2002: 21.) Tšekin- ja suomen kieli ovat hyvin kaukana toisistaan genealogisesti, kieliopillisesti sekä leksikaalisesti. Sen takia voi olettaa, että sanojen YSTÄVÄ ja PŘÍTEL väliltä löytyy eroja sekä niiden kollokaatio- että merkitysprofiilissa. Toisaalta valitsemani aineistosanat ovat toistensa yleisimmät käänkösvastineet ja mahdollisesti myös suomen

ja tšekin kielen oppijoiden ykkösvalinta sanaa kääntäessä, joten YSTÄVÄ/PRÍTEL -lekseemien välillä on luonnollisesti myös paljon yhteistä. Kiinnostavaa tutkielman tavoitteen kannalta on kuitenkin selvittää, missä ja miten sanat eroavat, ja missä ne jakavat yhteistä pohjaa.

1.3 Aiempi tutkimus

Kahden kielen ilmiöitä vertailevia kontrastiivisia tutkimuksia löytyy runsaasti. Kuitenkin tällaisia, joissa vertaillaan kahden kielen lekseemien kontekstuaalista valintaa, on huomattavasti vähemmän. Yhtenä esimerkkinä on Tognini-Bonellin (2001; 2002) tutkimus, jossa hän käsittelee englannin klusteria, eli monisanaista ilmaisua *in (the) case (of)* ja sen italian vastinetta. Hänen metodiikkansa koostuu kolmesta selkeästä osasta: ensin hän käyttää englannin yksikielistä korpusta ja lähdesanan kontekstin ja kontekstuaalisen valinnan perusteella tunnistaa lekseemin eri funktiot. Sen jälkeen hän löytää jokaiseen funktioon italian vastineen rinnakkaiskorpuksen avulla. Lopuksi Tognini-Bonelli tarkastelee vastineiden kontekstia italian yksikielisestä korpuksesta. (Tognini-Bonelli 2002: 81–82.) Omassa tutkimuksessani käytetty metodiikka eroaa siinä mielessä, että alusta asti pidän lekseemiä YSTÄVÄ ja PRÍTEL toistensa vastineina, minkä kuitenkin vahvistan hyödyntäen rinnakkais- eli käännöskorpusta. Lekseemien kontekstuaalisia valintoja tutkin yksikielisten korpusten kautta.

Xiao ja McEnery (2006) keskittyvät tutkimuksessaan käännösvastineisiin englannin ja kiinan kielissä, kiinnittäen huomionsa etenkin kollokaatteihin ja semanttiseen prosodiaan. Tutkimuksessaan he myös pohtivat tämäntyyppisen kontrastiivisen tutkimuksen hyötyä kielen oppimisessa (mts. 126). Mainitsemisen arvoinen on myös Salmisen (2008) maisterintutkielma, missä tutkitaan ympäristöön ja olosuhteisiin liittyvää neljää englannin sanaa ja niiden ruotsin vastineita merkityksen ja käytön osalta. Aineistona hän käyttää pääosin sanakirjoja, mutta hypoteeseja testataan yksikielisillä- sekä rinnakkaiskorpuksilla. Hyödynnän tutkielmassani myöskin suomen ja tšekin sanakirjoja valittujen lekseemien eri merkitysten määrittämiseksi, jonka jälkeen vertaan niitä korpusaineistoa vasten. Mielenkiintoa herättää myös Součekin maisterintutkielma, jossa kontrastiivisesti vertaillaan ajan substantiiveja (kuten AIKA, VUOSI, VIIKKO, PÄIVÄ jne.) suomen ja tšekin kielen välillä. Hänen tutkielmassaan käsitellään jokaisen sanan taajuutta, kieliopillisia piirteitä, kollokaatteja sekä fraseologisia ilmaisuja. Työn tavoitteena oli hakusanojen luominen uutta suomi–tšekki–suomi sanakirjaa varten lekseemien käytötyhteydet huomioon ottaen. (Souček 2008.)

Yhden kielen puitteissa lähisynonyymien kollokaatteja ja kontekstuaalista valintaa on tutkittu monesti. Vaikka tutkimukset keskittyvät etenkin Englantiin, suomen kielenkin osalta on nykyään tehty useita fraseologisia tutkimuksia. Jantunen (2004) tarkastelee kolmen astemäärätteen HYVIN, KOVIN ja OIKEIN kontekstuaalista erilaisuutta alkuperäis- ja käännösuomessa. Arppe (2008) tutkii synonyymisiä verbejä AJATELLA, MIETTIÄ, POHTIA, ja HARKITA sekä sitä, millaiset tekijät vaikuttavat niiden keskinäiseen valintaan. Viime vuosilta löytyy myös muutamia maisterintutkielmia, joissa tarkastellaan muun muassa seuraavia lähisynonyymejä: EHKÄ ja MAHDOLLISESTI (Pirkola 2016), KEHO, VARTALO ja RUUMIS (Laakkonen 2017) sekä TÄYSIN ja KOKONAAN (Kangas 2018).

2 TEORIATAUSTA

Vaikka korpuslingvistiikan juuret ulottuvat pitkälle aikaan ennen digitaalista vallankumousta, sen läpimurron on mahdollistanut vasta tekninen kehitys sekä kielitieteilijöiden kasvava kiinnostus todelliseen kielenkäyttöön pikemmin kuin kieleen abstraktina järjestelmänä (Johansson 2008: 33). Korpustutkimuksen tärkeimpiä etuja on mahdollisuus tarkastella kieltä sen autentisessa ja luonnollisessa ympäristössä, samalla empiirisesti tutkien kielen ilmiöitä (Tognini-Bonelli 2001: 2; Stubbs 2007: 130; Gries 2011: 89). Korpus on yleisesti sanoen laaja sähköinen tekstikokoelma, joka on rakennettu tiettyä tarkoitusta varten ja jossa kielidataa on yleensä annotoitu, eli varustettu metatiedolla (Biber, Conrad & Reppen 1998: 3–4; Hundt 2008: 170). Monet korpuksia on koottu ottamalla huomioon muuttujat kuten tekstilaji, kielen variaatio, rekisteri, ajanjakso ja muut, joten korpustutkimusta käytetään hyväksi monella soveltavalla kielitieteen alalla, muun muassa fraseologisessa tutkimuksessa, käännöstieteessä, leksikografiassa ja kielen opettamisen tutkimuksessa (Tognini-Bonelli 2001: 1–6).

Kielitieteilijöiden keskuudessa ei ole yksimielisyyttä siitä, edustaako korpuslingvistiikka omaa teoriakehystään vai pitäisikö sitä mieltä pelkkänä tutkimusmenetelmänä. Teoksessa *Perspectives on Corpus Linguistics* (Viana, Zyngier & Barnbrook 2011) haastatellaan useita korpustutkijoita kyseisestä ongelmasta ja moni toteaa, ettei kysymykseen löydy yksinkertaista vastausta. Esimerkiksi Leech (2011: 158–159) sekä McEnery, Xiao ja Tono (2006: 7) ovat sitä mieltä, että korpuslingvistiikka on pikemminkin metodologia, koska se ei edusta kielitieteen itsenäistä haaraa samassa mielessä kuin sociolingvistiikka tai semantiikka, vaan sitä voi soveltaa melkein mihin tahansa kielentutkimukseen. Toisaalta Tognini-Bonelli (2001: 1) väittää, että vaikka korpuslingvistiikka kuuluukin soveltavan kielentutkimuksen runkoon, sitä ei rajoiteta ennalta määriteltyjen sääntöjen ja asetelmien mukaan, vaan tutkijat asettavat säännöt ja käytännöt tutkimustarpeidensa mukaisesti. Monet tutkijat ovat tämän lisäksi sitä mieltä, että korpuslingvistiikkaa voi mieltä sekä menetelmänä että kielitieteen osa-alueena riippuen näkökulmasta ja tietyn tutkimuksen lähestymistavasta (ks. Aston 2011: 2–3; Conrad 2011: 49). Johanssonin (2011: 117) mukaan korpuspohjaisessa tutkimuksessa, jolla etsitään esimerkkejä ja tukea jo olemassa olevalle teorialle, painotus on usein korpuksessa menetelmänä. Korpusvetoiselle lähestymistavalle korpusaineisto on toisaalta ensisijainen, ja sitä tarkastelemalla voi luoda uusia hypoteeseja kielestä samalla hyläten vanhat. Tästä näkökulmasta korpuslingvistiikalla voi mieltä olevan teorian status. (Tognini-Bonelli 2001: 84.)

Tutkielmani sijoittuu korpuspohjaisten tutkimusten joukkoon, sillä lähestyn aineistoa tietyllä tavoitteella ja aiempien tutkimusten pohjalta muodostetuilla oletuksilla siitä, minkälaisiin tuloksiin voisi päästä. Tässä luvussa käsittelen yksitellen kontekstuaalisen semantiikan, fraseologian ja sanojen merkityssuhteiden teoreettiset lähtökohdat. Näistä jokaista lähestyn korpuslingvistiikan näkökulmasta, joka toimii tutkielmani punaisena lankana. Ensimmäisessä aluvussa esittelen kontekstuaaliseen semantiikkaan liittyvät tärkeimmät käsitteet ja ajattelutavat. Tämä luo tilaa seuraavalle, fraseologista tutkimusta koskevalle luvulle, jonka keskeinen käsite on kollokaatio. Sanojen merkityssuhteiden puolesta ydinkäsitteenä on polysemia ja siihen yhdistetty merkityksen muutos.

2.1 Kontekstuaalinen semantiikka

2.1.1 Periaatteet ja lähtökohdat

Lingvistinen semantiikka on kielitieteen osa-alue, joka tutkii kielen viestimää sisältöä, tarkkaan ottaen sanojen ja lauseiden merkitystä. Semantiikan tutkimusalue onkin hyvin laaja ja sisältää eri tutkijoiden vaihtelevia määritelmiä, käsitteitä ja käytäntöjä, mikä tekee siitä yhden kielitieteen monipuolisimmista osa-alueista. (Saeed 2011: 3–4.) Semantiikka koostuu yleisesti ottaen useammasta, osittain päällekkäisestä haarasta, kuten leksikaalisesta semantiikasta eli sanasemantiikasta, kieliopillisesta semantiikasta, joka tutkii kieliopillisten rakenteiden merkitystä ja pragmatiikasta, jossa keskeisenä kiinnostuksen kohteena on tilannekohtainen kielenkäyttö (Cruse 2004: 13–14). Semantiikan ja pragmatiikan välinen suhde on kuitenkin monimutkaisempi. Jos siirrytään vielä askel yleisemmälle tasolle, semantiikan ja pragmatiikan nähdään perinteisesti edustavan syntaksin kanssa omia suuntauksiaan semiotiikan eli merkkiopin alalla (Morris 1938; Vanderveken 2004: 710). Semantiikan ja pragmatiikan tarkka erottelu on kuitenkin hankalaa ja kiistanalaista, sillä molemmilla tutkimusaloilla on monta yhteistä aihetta (Kangasniemi 1997: 25; Saeed 2011: 17). Kuiri (2012) jaottelee sanasemantiikan vielä kahdeksi suuntaukseksi. Näistä synkroninen käsittelee merkitystä, millaisena se on ”tässä ja nyt”, toisin kuin diakroninen semantiikka, jonka tutkimuskohteena on sanojen merkitysten muutos (ks. mts. 30, 43) (tarkemmin luvussa 2.3).

Kontekstuaalisessa semantiikassa on ensisijaista selvittää merkityksen muodostumista kontekstilähtöisesti. Konteksti voi tarkoittaa ilmauksen lähintä tekstiympäristöä, eli kotekstia, tai laajempaa sosiaalista ja kulttuurista ympäristöä. (Jantunen 2004: 7.) Tutkielmani kuuluu

luonteeltaan pääosin leksikaalisen ja kontekstuaalisen semantiikan piiriin, sillä tutkimukseni keskiössä on suomen sana YSTÄVÄ ja sen tšekkin vastine PŘÍTEL. Näiden kahden sanan osalta minua kiinnostaa etenkin niiden merkityslaajentumat, semanttinen tekstiympäristö ja kontekstuaalinen valinta, jota käsittelen seuraavassa luvussa.

Brittiläinen kielitieteilijä J. R. Firth (1957: 179) on jo 1950-luvulla korostanut, ettei merkitys ole sanoissa yksikköinä, vaan että merkitys muodostuu muiden sanojen yhteydessä (*”You shall know a word by the company it keeps!”*). Kontekstin huomioon ottaminen on siis hänen mukaansa keskeinen perusta merkitystä käsiteltäessä. Firthin kontekstuaalisessa merkitysteoriassa (*contextual theory of meaning*) on olennaista, että jokainen puheteko ja laajemmin jokainen ihmisen toiminta on merkityksellinen ja merkitystä ei voi käsitellä irrallaan kontekstistaan. Firthin mukaan merkitykseen voidaan päästä käsiksi nimenomaan kielen kautta. (Firth 1957: 175–178; Tognini-Bonelli 2001: 157–159.) Tämä käsitys eroaa toisista kieliteorioista, kuten esimerkiksi konseptuaalisesta semantiikasta, joka pitää merkitystä kielenkäyttäjän tajunnassa olevana mielikuvana (Kangasniemi 1997: 10; Jantunen 2004: 7). Firthin teorian pääkäsite on tilannekonteksti, joka on läsnä jokaisen tekstin ja ilmaisun ympärillä, ja jonka puitteissa merkitys on tulkittava. Tilannekonteksti sisältää vuorovaikutukseen osallistujia, heidän verbaalit sekä ei-verbaalit toimintansa ja näiden väliset vaikutukset. Tilannekonteksti on Firthin käsityksessä abstraktio, joka koostuu toistuvista ja kielenkäyttötilanteelle tyypillisistä tapahtumista. (Firth 1957: 175–178; Tognini-Bonelli 2001: 157–159.) Toistuvuuden käsite on tämän tutkielman näkökulmasta tärkeä etenkin kollokaatiota tarkastellessa, koska kollokaattien yhtenä peruspiirteenä on juuri niiden toistuvuus.

Firthin ansiosta sekä laajan kielidatan käsittelyyn kykenevien ohjelmien kehityksen seurauksena syntyi ns. firthiläinen koulukunta, jonka kieliteoriaa ovat kehittäneet mm. John Sinclair ja M.A.K. Halliday. Tähän teoriarunkoon kuuluu ajatuksia ja periaatteita kuten tarve tutkia kieltä sen luonnollisessa ja autenttisessa ympäristössä, muodon ja merkityksen erottamaton luonne, sekä leksikon ja syntaksin kietoutunut suhde. Sen lisäksi painotetaan, että kielentutkimuksen yksikkönä on koko teksti, jota pitäisi tarkastella korpuksen avulla. (Stubbs 1993: 1–2.)

Kielitieteessä oli 1990-luvulle asti laajalle levinnyt perinne, joka nojasi intuitiivisesti keksittyjen, kontekstista irrotettujen kieliesimerkkien käsittelyyn, joista sitten tehtiin yleispäteviä johtopäätöksiä. Sinclair vastustaa tätä lähestymistapaa ja väittää, että kieltä pitäisi analysoida sen luonnollisessa kontekstissa, sillä tutkijoiden introspektiota ja kielitajua ei voi pitää objektiivisena, eivätkä ne yleensä riitä kielellisten ilmiöiden selittämiseksi. Sen lisäksi intuitiiviset esimerkkilauseet kattavat todellisesta kielenkäytöstä vain murto-osan. Suurinta osaa kielitiedosta on aiemmin kuitenkin perusteltu niillä. (Stubbs 1993: 8–10; Carter 2003: 2.) Tämän

vuoksi Sinclair (1991: 17) painottaa korpusaineiston käyttöä ja analysointia, sillä korpuksen avulla voi tunnistaa, mikä kielessä on keskeistä ja tyypillistä.

Muodon ja merkityksen sekä syntaksin ja leksikon yhteenkuuluvuus tarkoittaa, ettei ole mahdollista tutkia kieliopillisia rakenteita ikään kuin abstrakteina ja merkityksestä riippumattomina käsitteinä, kuten on ollut tapana chomskylaisessa kielentutkimuksen perinteessä (Carter 2003: 2–3). Päinvastoin merkitys vaikuttaa kielen rakenteisiin ja jokaisen merkityksen voi liittää omanlaisiinsa kielellisiin malleihin (Sinclair 1991: 496). Toisin sanoen yhden sanan eri taivutusmuodoilla on tekstissä taipumus esiintyä tietynlaisessa kielellisessä ympäristössä, johon kuuluu kieliopilliset rakenteet sekä muut lekseemit. Samalla tavalla yhden sanan eri muodot voivat liittyä kontekstissaan eri merkityksiin. Laajan korpuksen avulla saavutetun kielidatan keinoin on ollut mahdollista tunnistaa ja todistaa uudet suhteet lemmän¹, sananmuodon ja kieliopin välillä. (Stubbs 1993: 14–16.)

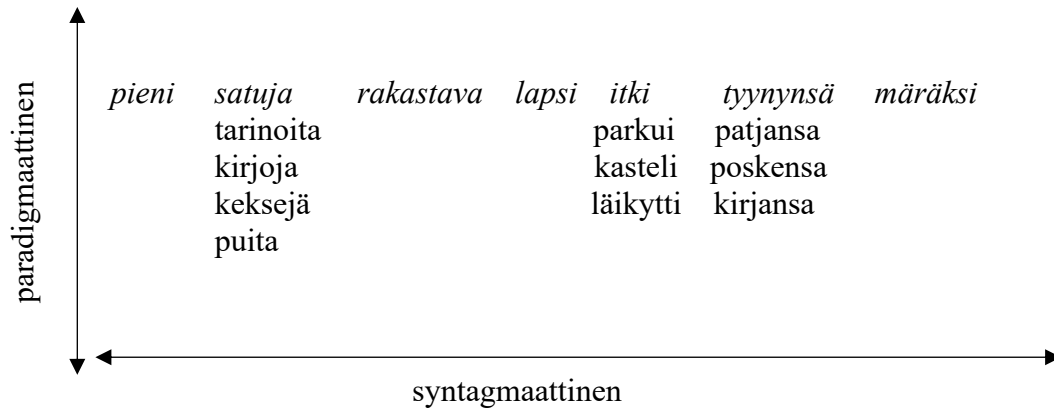
Esimerkiksi Jantusen lähisyonymien TÄRKEÄ ja KESKEINEN tutkimuksesta käy ilmi, että adjektiivi TÄRKEÄ esiintyy huomattavasti useammin partitiivissa sekä komparaatiomuodoissa kuin sana KESKEINEN. Adjektiivin TÄRKEÄ kontekstista löytyy myös runsaasti kvantiteettia ilmoittavia astemääritteitä (VÄHEMMÄN, PALJON, HYVIN), kun taas KESKEINEN-sanan ympäristössä ilmenee pikemminkin lukusanakvanttoreita (YKSI, TOINEN, ERÄS). (Jantunen 2001.)

2.1.2 Kontekstuaalinen valinta

Jotta kielen avulla voisi viestiä merkitystä, sanoja yhdistetään fraaseihin, ketjuihin ja lauseisiin. Jo sata vuotta sitten Saussure määritteli kaksi suhdetta sanan ja sen yhteydessä olevien sanojen välillä. Nämä ovat syntagmaattinen ja assosiatiivinen [nykyään paradigmaattinen] suhde. (Saussure 1959 [1916]: 124–125; Saeed 2011: 84.) Paradigmaattinen suhde heijastaa semanttista valintaa, joka tehdään, kun valikoidaan sopiva sana lauseen tiettyyn paikkaan muiden mahdollisten sanojen tai ilmausten joukosta. Paradigmaattinen suhde kattaa sanojen merkityssuhteita, kuten synonyymisyyttä ja hyponyymisyyttä. Tässä suhteessa olevat ilmaukset yleensä kuuluvat samaan syntaktiseen luokkaan, kuten substantiivit tai adjektiivit. Syntagmaattinen suhde toisaalta kuvaa yhden ketjun tai lauseen lekseemien lineaarista semanttista yhteyttä. (Cruse 2004: 145–146.) Sanojen paradigmaattisten ja syntagmaattisten suhteiden rinnastusta

¹ Korpustutkimuksessa lemmalla tarkoitetaan yhteen kuuluvien sananmuotojen abstraktiota, joka on usein haku-sanan perusmuoto. Sananmuoto on siten tietyssä taivutusmuodossa oleva lekseemi. (Karlsson 2004: 188.) Näin YSTÄVÄ-lemman eri sananmuodot ovat muun muassa *ystävä*, *ystävien*, *ystävämme*, *ystävänäkin*, jne.

ehkä parhaiten kuvaillaan esimerkeillä. Jantunen kuvaa ilmiötä ilmauksella ”*pieni satuja rakastava lapsi itki tyyneysä märäksi*” (kuvio 1). Hän selittää, että tässä tapauksessa voi syntagmaattisessa suhteessa oleviksi ilmauksiksi arvella esimerkiksi sanaparit *pieni – lapsi*, *lapsi – satu*, ja *lapsi – itkeä*. Paradigmaattisessa suhteessa olevia ilmauksia voi puolestaan olla monta erilaista. (Jantunen 2004: 8.)



Kuvio 1. Syntagmaattinen ja paradigmaattinen ulottuvuus (Jantunen 2004: 9).

Saussure väitti, että syntagmaattinen suhde toteutuu *in praesentia*, eli tässä ja nyt, yhdessä esiintyvien ilmausten välillä. Paradigmaattinen suhde on kuitenkin sanojen potentiaalinen yhteys ja tapahtuu kognitiivisen assosiaation kautta *in absentia*, eli se ei ole havaittavissa. (Saussure 1959 [1916]: 123–124.) Stubbs (2001: 240–241) toteaa, että korpuksen menetelmät ovat mahdollistaneet molempien suhteiden havainnoinnin ja tarkastelun. Kuviossa 2 on esitelty, miten korpuksesta haettaessa tulokset järjestyvät konkordansseihin, eli yksittäisiin riveihin, joiden keskellä on hakusana tai noodi ja sen ympärillä oikean- ja vasemmanpuoleinen konteksti (McEnery & Hardie 2012: 35).

Konkordanssilista antaa puitteet havaita ilmausten toistuvia esiintymiä: horisontaalisesti näkyy yksittäisten rivien syntagmaattisia myötäesiintymiä ja vertikaalisesti taas paradigmaattisia valintoja (Stubbs 2001: 241). Tutkielmassani keskityn pääosin valitsemani sanaparin syntagmaattisiin suhteisiin, koska kiinnostuksen kohteenani on se, millaiset ilmaukset esiintyvät toistuvasti sanojen leksikaalisessa ympäristössä, ja miten ne muodostavat merkityksiä. Keskeistä on siis aineistosanojen lineaaristen myötäesiintymien, etenkin kollokaattien tarkastelu (ks. luku 2.2.2).

vasen konteksti

lauantaina 2.9. klo 17 Karkun Tyrisevän seuraintalolla
 Yhtä mutkattomasti Rauno Mattila vieraili
 Haiderilla on
 ilahduttanee myös monia ikäihmisiä, joilla on koira
 viestii suomalaisten kasvavasta halusta muistaa
 on viiden, kuuden vuoden aikana neulonut ne käsin
 Erityisen sijaan saa runojen ja riimien laatiminen
 Hän kiitti erityisesti edesmennyttä
 Nicholas Paynen korviin, ja kun vielä Salosen
 on nerokas. Kun nainen päästetään sisälle
 joukossa ovat mm. Rakas sydän, Lacombe Lucien --

noodi

ystävien
ystäviensä
ystäviä
ystävänään
ystäviä
ystävillään
ystävälle
ystäväänsä
ystävä
ystävänä
ystävät

oikea konteksti

ja sukulaisten perunasoppajuhlana. Kaikki rahtarit
 ja sukulaistensa luona.
 ja tukijoita Yhdysvaltain republikaanien lisäksi
 ja turvanaan, mutta jonka kanssa ei nyt kyläilemään
 ja tuttavilla henkilökohtaisella tavalla. Lisäksi Posti
 ja tuttavilleen, viime aikoina myös pojanpojalleen
 ja työtovereille sopiviin lauluihin ympättyinä.
 ja työtoveriaan Leena Viholaa.
 ja työtoveri Peter Sellars halusi ohjata sen, palapelin
 ja vieraana, hänet voidaan sulkea pysyvämmiin ulos
 ja viholliset, Pretty Baby, Atlantic City, Alamo Bay,

Kuvio 2. YSTÄVÄ-hakusanan konkordanssilista järjestettynä oikeanpuolisen kontekstin mukaan.

Cruse huomauttaa, että jotkin sanojen ja ilmausten kombinaatiot lauseessa menevät yhteen luontevasti ja vaikuttavat sopivilta yhdessä. Toisaalta on yhdistelmiä, jotka näyttävät jotenkin oudoilta ja kuulostavat kielen puhujalle häiritseviltä, vaikka ne olisivatkin kieliopillisesti oikein ja semanttisesti ymmärrettäviä. Toisin sanoen lekseemien kontekstuaalinen valinta muodostuu joistakin preferensseistä ja rajoituksista, joiden noudattaminen edesauttaa kielen luonnollista käyttöä. (Cruse 2004: 217–219.) Sinclair käsittelee kahta eri periaatetta, joiden mukaan sanoja liitetään yhteen. Nämä ovat vapaan valinnan (*open-choice principle*) ja idiomaattisen valinnan (*idiom principle*) periaatteet. Vapaan valinnan perusteella lauseiden tai fraasien muodostamisen nähdään koostuvan isosta määrästä valintoja, joiden ainoa rajoitus on näiden kieliopillisuus. Tämä tulkinta siis asettaa vain minimaalisen rajoituksen kontekstuaaliseen valintaan: kunnes lause on syntaktisesti kunnossa, kielenkäyttäjät täyttävät vapaat paikat millä lekseemillä tahansa. Sinclair väittää tämän olleen hallitseva kielen näkemys, joka heijastuu myös kielioppikirjoihin. (Sinclair 1991: 109–110.)

Vapaaseen valintaan verrattuna idiomaattisen valinnan periaatteen mukaan ilmauksia ei aseteta tekstiin vain kieliopillisia sääntöjä seuraten, vaan sanojen kontekstuaalinen valinta jäljittelee sitäkin, miten ympäröivä fyysinen maailma on järjestynyt. Esimerkiksi asiat, jotka usein esiintyvät toistensa seurassa, mainitaan tekstissä myös tavallisesti yhdessä. Tärkeässä roolissa on myös se, että kielenkäyttäjällä on hallussaan jo edeltä osittain valmiiksi muodostettuja fraaseja, joista hän kieltä tuottaessaan valitsee sopivan. Tavalliset tekstit ovat nimenomaan idiomaattisen valinnan perusteella rakennettuja, vaikka onkin tekstejä, jotka noudattavat

pääosin vapaan valinnan periaatetta. (Sinclair 1991: 110–114.) Esimerkiksi tekniset- ja lakitekstit koostuvat monista termeistä, joiden merkitys on ennalta kiinteästi asetettu ja kielenkäyttäjällä ei ole paljoa valinnanvaraa, kun haluaa viitata samoihin tarkoitteisiin (mts. 114; Sinclair 1996: 83). Moon (1998: 30) kannattaa tätä näkökulmaa sillä, että kielen omaksuminen ja säilyttäminen sekä käyttäminen toteutuvat fraasien ja muiden monisanaisten yhdistelmien, eivät yksittäisten sanojen avulla. Wray (2002: 100–101) huomauttaa, että kieltä tuottaessaan kielenkäyttäjä pääsee käsiksi isoon määrään valmiita konstruktioita, jotka ovat dynaamisia eivätkä staattisia, ja joita jatkuvasti muokataan omiin tarpeisiin. Kaiken kaikkiaan tässä linjassa kollokaatio eli sanojen toistuva myötäesiintyminen nähdään idiomaattisen valinnan periaatteen ilmentymänä (Sinclair 1991: 115; Barnbrook 2009: 31). Kielikorpusten ja käsittelyohjelmien ansiosta on mahdollista tutkia konkordansseja tarkastellakseen lekseemien sekä syntagmaattisia että paradigmaattisia suhteita. Sen lisäksi on mahdollista laskennallisilla menetelmin tuottaa todisteita sanojen kontekstuaalisista valintasuhteista (ks. luku 3.2.1).

2.2 Fraseologia

2.2.1 Fraseologinen yksikkö

Fraseologiaa ei ole helppoa määritellä kielitieteen tutkimusalana yksiselitteisesti. Yleisesti ottaen fraseologia tutkii sanojen yhdistelmien rakennetta, merkitystä ja käyttöä. Fraseologia oli pitkään laiminlyöty alue, ja varhaisessa fraseologisessa tutkimuksessa tutkijat keskittyivät etenkin läpinäkymättömiin idiomaattisiin ilmaisuihin kuten *raining cats and dogs* (”sataa kaatamalla”). (Ebeling & Ebeling 2013: 2–3.) Viime vuosisadan lopusta asti fraseologia on kuitenkin ulottunut jokaiseen kielitieteen alaan, ja tutkijat ovat olleet yhä enemmän kiinnostuneita sanojen syntagmaattisista suhteista sekä sanojen ympäristössä toistuvasti esiintyvistä leksikaalisista yksiköistä tai kieliopillisista rakenteista. (Granger & Meunier 2008: xix–xx; Gries 2008: 5.)

Nykyaikainen fraseologian nousu on tapahtunut nimenomaan korpustutkimuksen ansiosta, koska korpuslingvistiikka on vahvistanut sen oletuksen, että luonnollinen kieli koostuu merkittävien osien sanojen ja rakenteiden toistuvista malleista, ja että suurin osa tekstistä on tulkittava idiomaattisen valinnan periaatteen mukaisesti (Ellis 2008: 4–5; Sinclair 2008: xvi). Sen lisäksi korpusmenetelmin on mahdollista saada sanojen kollokaattien tarkkoja ilmenemisfrekvenssejä ja tilastollista tietoa (Gries 2008: 15–16). Ellis luettelee fraseologian elpymisen

toiseksi syyksi sen, että fraseologia nitoo leksikon, kieliopin, semantiikan sekä sosiaalisen tutkimuksen yhteen, joten nämäkin alat tukevat fraseologista tutkimusta. Fraseologia saa kannatusta myös kognitiivisesta lingvistiikasta, sillä se korostaa kielen ja ulkoisen maailman yhteyttä. (Ellis 2008: 5.)

Fraseologisen teorian osalta löytyy monenlaisia näkökulmia siitä, mistä fraseologinen yksikkö koostuu, sekä eri tutkijoiden asenteita siihen, miten pitäisi analysoida yksikön fraseologista ympäristöä, ja minkä voi sellaiseksi edes laskea. Tämä tietenkin riippuu myös siitä, mitä koulukuntaa tai teoriaviitekehystä kyseinen tutkija edustaa, ja mitä on hänen mielestään olennaista tutkia. Tämä monimuotoisuus heijastuu alan vahvasti vaihtelevaan terminologiaankin. (Ebeling & Ebeling 2013: 49.) Esittelen seuraavaksi lähestymistapaa, jonka koen olevan relevantein tutkimukseni kannalta, eli Sinclairin merkitysteorian kontekstuaalista merkitysyksikköä² (*extended unit of meaning*), joka sisältää tämän työn pääkäsitteen – kollokaation.

Sinclair korostaa tarvetta tutkia sanan merkitystä ottaen huomioon sen tekstuaalisen ympäristön, sillä pelkkänä lekseemina sanalla on monta eri merkitystä, ja ilman kontekstia on mahdotonta päätellä, mikä sen merkityksistä on voimassa. Hän selittää, että jos tarkastellaan lekseemin kahta tai kolmea ympäröivää sanaa merkityksen selventämiseksi, lekseemin monitulkintaisuus pääosin häviää. (Sinclair 1998: 6–9.) Sinclair ehdottaa mallia, jonka mukaan merkitysyksiköllä on viisi myötäesiintymän osatekijää, joista kaksi on pakollisia ja kolme täydentäviä. Pakollisia ovat ydin (*core*), eli leksikaalinen yksikkö itse, ja semanttinen prosodia, joka esittää syytä miksi jokin tietty sana oli valittu, ja joka usein ilmaisee pragmaattista merkitystä. Prosodia on kuitenkin yleensä niin hento, että sitä on vaikea huomata. Kolme täydentävää luokkaa ovat Sinclairin mukaan kollokaatio, kolligaatio ja semanttinen preferenssi. Kollokaatit ovat ilmauksia, jotka esiintyvät toistuvasti yhdessä, kolligaatio vastaa lekseemin ja kieliopillisten rakenteiden keskinäistä riippuvuutta ja semanttinen preferenssi on sanan myötäesiintymien yhteinen semanttinen teema. (Mts. 14–16, 20.)

Stubbs kommentoi tätä teoriaa ja lisää, että merkitysteoria yhdistää muodon, merkityksen ja kommunikatiivisen funktion. Hän selittää, että kollokaatio ja kolligaatio tarkoittavat sitä, *miten* jokin on ilmaistu (muoto), semanttinen preferenssi kertoo, *mitä* on ilmaistu (teema) ja prosodia selittää *miksi* (puhujan motivaatio). (Stubbs 2013: 24.) Semanttista preferenssiä käsitellään tutkimuksissa harvemmin kuin kollokaatiota, ja joskus se sekoitetaan semanttiseen prosodiaan. Toisin kuin kollokaatio, semanttinen preferenssi on abstraktimpi ja vaatii tutkijan subjektiivista analyysiä. (Jantunen 2004: 24.) Stubbs (2001: 65) selittää käsitteen suhteeksi, joka

² suomen termi ks. Jantunen 2004:9

on semanttisesti toisiaan lähellä olevan sanajoukon ja tutkittavan sanan välillä. Hyvin tunnetun esimerkin esittää Sinclair ilmauksella *naked eye*. Tämä vahvasti kollokoi verbin SEE, adjektiivin VISIBLE sekä muiden semanttisesti läheisten sanojen kanssa, joten sen semanttiseksi preferenssiksi paljastuu *näkyvyys (visibility)*. (Sinclair 2004: 32–33.) Toisaalta semanttinen prosodia liittyy Stubbsin (2013: 24) mukaan puhujan arviointiin sekä asenteeseen, ja erilaisissa tutkimuksissa se on tyypillisesti havaittu negatiiviseksi, positiiviseksi tai neutraaliksi. Bednarek katsoo semanttisen preferenssin sekä prosodian olevan hyvin lähellä toisiaan, sillä erolla, että prosodia on käsitteistä yleisempi ja abstraktimpi. Lisäksi semanttinen prosodia on ongelmallinen käsite siinä mielessä, että sen arviointi negatiiviseksi tai positiiviseksi riippuu hyvinkin paljon tutkijan omasta subjektiivisesta asenteesta. (Bednarek 2008: 121–122.) Esimerkkinä voi olla sanan INTELLECTUALS kollokaatit kuten ACTIVIST, LEFTWING ja LIBERAL, joita sosiaalisesta tai poliittisesta piiristä riippuen tulkitaan sekä positiiviseksi että negatiiviseksi (Stubbs 1996: 188).

Tutkielmani rajallisuuden takia ei ole mahdollista perehtyä jokaiseen sanojen YSTÄVÄ/PŘÍTEL merkitysten ulottuvuuteen, vaan keskityn pääosin kollokaatioon, jota käsittelem yksityiskohtaisemmin seuraavassa alaluvussa. Vaikka semanttisen preferenssin ja prosodian varsinainen analyysi jää tämän tutkielman ulkopuolelle, tarkastelen kollokaatteja myös semanttisesta näkökulmasta.

2.2.2 Kollokaatio

Kollokaatiokäsitteen on ensimmäisenä määritellyt Firth (1957: 180–181) tyypillisesti toistensa seurassa esiintyvien sanojen suhteeksi. Käsitettä on kuitenkin kehitelty vasta myöhemmin, sillä kollokaation kokonaisvaltaista luonnetta ei ollut Firthin aikaan mahdollista havaita (Tognini-Bonelli 2001: 162). Sinclair selittää idiomaattisen periaatteen mukaisesti, että sanat ovat joskus valittu pareina tai ryhminä ja ne esiintyvät tekstissä toistensa läheisyydessä, mutteivat välttämättä juuri vierekkäisinä. Tätä kahden tai useamman sanan myötäesiintymää kutsutaan kollokaatioksi. (Sinclair 1991: 115, 170.) Kyseisen tekstuaalisen määritelmän lisäksi kollokaatiota voi käsitellä myös assosiatiivisena ja tilastollisena ilmiönä (Partington 1998: 15–16; Jantunen 2001: 173).

Assosiatiivisella tai psykologisella määritelmällä Partington (1998: 16) tarkoittaa sitä, että natiivipuhuja intuitiivisesti tunnistaa, mitkä kollokaatit kuulostavat tietyssä kontekstissa luonnollisilta ja mitkä taas ovat epätyypillisiä. Hän kuitenkin myöntää, että kielenpuhujien kielitajua ei voi nähdä yhteneväksi, ja että natiivipuhujienkin mielipiteet eroavat siitä, mitkä

kollokaatit ovat tyypillisiä ja mitkä eivät. (Mts. 16–19.) Karlsson (2004: 232) puolestaan esittää kollokaation assosiativista ulottuvuutta esimerkkisanoilla HEVONEN ja HIRNUA, joilla on taipumus esiintyä tekstissä toistensa lähellä, sillä kielenulkoisessa maailmassakin sanojen tarkoitteet liittyvät yhteen. Teoriaa kyseisestä ilmiöstä kehittää Hoey (2007: 8–9) termillä *leksikaalinen priming*, jota hän ei näe sanan, vaan ihmisen ominaisuutena. Toisin sanoen, kun kollokaatit tyypillisesti esiintyvät toistensa kanssa, se ei tapahdu sanojen synnynnäisen yhteenkuuluvuuden vuoksi, vaan sen takia, että kielenkäyttäjät liittävät ne yhteen assosiaation kautta. Hoey korostaa, että priming on jokaiselle puhujalle yksilöllinen ja sitä ei voi suoranaisesti saavuttaa korpuksen kautta. Hän kuitenkin tarkentaa, että korpus voi osoittaa, millaiset assosiaatiot sanoista todennäköisesti syntyvät useimmille kielenkäyttäjille. (Mp.; Hoey 2005: 8–15.)

Kollokaation tilastollisessa määritelmässä on ensisijaista kollokaation toistuvuus. Jotta kollokaatti olisi tilastollisesti merkitsevä, sen täytyy toistua korpuksessa sekä hakusanan, eli noodin, läheisyydessä riittävän usein. (Jantunen 2001: 173.) Käytännössä noodin ja sen kollokaatin yhdessä esiintymisen määrää vertaillaan molempien sanojen frekvenssiin korpuksessa ottaen huomioon korpuksen koko. Tilastollisten testien osalta, suosittuja ovat etenkin *MI*-testi (*Mutual Information*), *t*-testi ja *z*-testi, joista jokaisella on tiettyjä etuja sekä puutteita, joten erilaisia tilastollisia menetelmiä yleensä yhdistellään. (Barnbrook 1996: 92–100; Kenny 2001: 91–94.) Tutkielmassani nojaudun pääosin kollokaation tekstuaaliseen ja tilastolliseen määritelmään. Tarkastelen pelkästään niitä kollokaatteja, jotka esiintyvät etukäteen valitussa kollokaatiovälissä, eli noodin tiukassa läheisyydessä. Näistä kollokaateista sitten tutkin ainoastaan niitä, jotka tilastollisissa testeissä osoittautuivat tilastollisesti merkitseviksi. Käyttämiäni kollokatioanalyysin menetelmiä ja tilastollisia testejä käsittelem tarkemmin luvussa 3.2.1. Tilastollisesti merkitseviksi kollokaateiksi tietenkin voivat nousta myös ns. konseptuaaliset kollokaatit, jotka ilmestyvät taajaan yhdessä niiden merkityssuhteiden vuoksi, kuten esimerkkisanat HEVONEN ja HIRNUA (Jantunen 2004: 18). En kuitenkaan näe tätä ongelmana, koska tämän tyyppisetkin kollokaatit antavat nähdäkseni arvokasta tietoa tutkittavista sanoista ja niiden käyttäytymisestä eri kielissä.

Kaiken kaikkiaan ei ole olemassa yhtä yleismaailmallisesti hyväksyttyä kollokaation määritelmää (Mel'čuk 1998: 23). Martin (2008: 56) huomauttaa, että kielitieteilijät ovat kuitenkin yleisesti ottaen samaa mieltä kahdesta asiasta. Ensiksi kollokaatio koostuu kahdesta osasta: ytimeistä, eli noodista, ja sen kollokaatista. Toiseksi näillä kahdella on moninainen määrä kiinnittyneisyyttä sekä rajoittuneisuutta toisiinsa. Tällä tavoin kollokaatioita voi sijoittaa kuvitteelliselle viivalle, jonka yhdessä päässä olisi kiinteä idiomi ja toisessa vapaa kombinaatio. (Mp.)

Jantunen mainitsee toistuvuuden ja tilastollisen merkitsevyyden lisäksi myös ulottuvuuden yhtenä kollokaation keskeisenä ominaisuutena. Sillä tarkoitetaan kollokaationaalisessa suhteessa olevien sanojen lukumäärää, noodin ja kollokaatin etäisyyttä, näiden kieliopillista suhdetta sekä kollokaatioväliä. (Jantunen 2004: 18.) Kollokaatio voi siis periaatteessa muodostua useammasta kuin kahdesta sanasta, eli noodiin voi liittyä monia kollokaatteja yhden konkordanssin puitteissa. Yleensä kuitenkin tällaiset yhdistelmät, joissa sanat on ketjutettu peräkkäin, erotellaan, ja niitä kutsutaan muun muassa klustereiksi. Koska nämä ovat usein rakenteeltaan sekä merkitykseltään poikkeavia tavallisista kollokaateista, ne yleensä käsitellään erikseen. (Mts. 18–19.) Tässä tutkimuksessa otan tarkasteltavaksi ainoastaan kahden sanan myötäesiintymiä.

Kun analysoidaan noodin kollokaatteja, on olennaista päätellä, kuinka isosta ns. kollokaatiovälistä mahdollisia myötäesiintymiä tutkitaan. Stubbsin mukaan merkittäviä kollokaatteja yleensä ilmenee välillä L4:R4, joka tarkoittaa neljää lekseemiä ennen noodia ja neljää sen jälkeen. Näkemykset tähän kokoon kuitenkin vaihtelevat tutkimuksittain. (Stubbs 2001: 29.) Tarkasteltujen sanojen väli riippuu myöskin tutkimuksen päämääristä ja valittujen sanojen piirteistä sekä sovelletuista menetelmistä (Jantunen 2004: 83–84).

Kollokaatioanalyysiä hyödynnetään muun muassa lähisynonyymien tutkimuksessa. Lähisynonyymillä on samanlainen denotaatio, mutta ne eroavat toisistaan esimerkiksi konnotatiivisessa merkityksessä (RUOKA vs. SAPUSKA) sekä niiden esiintymisympäristöissä. (Jantunen 2004: 53–55.) Jantunen HYVIN-, KOVIN- ja OIKEIN-lähisynonyymien tutkimuksesta selvisi, että KOVIN saa kollokaatiksi kieltoverbin EI, eli sitä käytetään useammin negatiivisuutta ilmaisevassa kontekstissa, toisin kuin adverbis HYVIN ja OIKEIN (mts. 86–92). Simon-Vanderbengenin (2013) englannin kielen adverbien BASICALLY, ESSENTIALLY ja FUNDAMENTALLY tutkimuksessa paljastui, että sanat BASICALLY ja ESSENTIALLY kollokoivat useammin samanlaisuutta esittävien sanojen kanssa (*basically similar, essentially identical*), kun taas sanaan FUNDAMENTALLY liittyy negatiivisuutta osoittavia kollokaatteja (*fundamentally wrong*) (Simon-Vanderbengen 2013).

Tutkielmani on lähisynonyymien tutkimuksen kaltainen, vaikka kiinnostuksen kohteena onkin kahden eri kielen sanat. Sanapari YSTÄVÄ ja PRÍTEL viittaavat samaan denotatiiviseen kohteeseen, mutta kaikki näiden merkitykset, käyttötavat sekä esiintymisympäristö eivät ole välttämättä samoja, jolloin niiden konnotaatioissa voi olla eroja. Tätä tarkastelen kollokaatioanalyysin keinoin, jossa nojaudun tilastollisiin testeihin. Tutkimukseni ei kuitenkaan ole pelkästään kvantitatiivinen, sillä korpusaineistosta saatuja kollokaatteja käsittelen myös laadullisesti semanttisessa analyysissä.

2.3 Sanojen merkityssuhteet

Lingvistisessä semantiikassa kiinnostuksen kohteena on merkitys sanan, kieliopillisen rakenteen, lauseen sekä laajemman tekstin ja diskurssin tasolla (ks. Lyons 1995; Karlsson 2004). Aiemmissa luvuissa tarkastelin kontekstuaalisen lingvistiikan periaatteita ja sitä, miten konteksti vaikuttaa sanan merkitykseen. Silloin keskiössä oli etenkin sanojen syntagmaattiset suhteet, eli lekseemien lineaarinen semanttinen yhteys. Tässä alaluvussa keskityn sanojen paradigmaattisiin suhteisiin ja käsittelen tarkemmin leksikaalisen, eli sanasemantiikan käsitteitä kuten synonymia, polysemia ja semanttinen muutos.

Sanojen paradigmaattista ulottuvuutta luonnehtii lekseemien tai rakenteiden keskinäinen vaihdettavuus. Näin elementit, joiden välillä on vaihdettavuussuhde yleensä kuuluvat samaan kieliopilliseen luokkaan. (Lyons 1995: 124.) Crusen (1986: 16) mukaan sanojen paradigmaattinen kiintymys on suurempi, kun myös niiden syntagmaattiset suhteet ovat vastaavanlaisia. Esimerkiksi sanojen KISSA ja KOIRA välillä on vahvempi paradigmaattinen suhde, kuin niiden ja substantiivin LAMPPU välillä, koska lauseet *Arttu ruokki kissaa/koiraa* ovat tavallisia, kun taas lause *Arttu ruokki lamppua* on poikkeava (mp.; Kangasniemi 1997: 67).

Sanojen merkityssuhteiksi perinteisesti lasketaan synonymia, hyponymia, antonymia sekä homonymia ja polysemia (Kuiri 2012: 30–36). Synonymia viittaa sanojen merkitysten samuuteen, ja kuten jo mainittiin tutkielman johdannossa, on kyseenalaista, onko täydellistä synonymiaa edes olemassa. Se vaatisi, että synonyymit ovat täysin keskenään vaihdettavissa kaikissa konteksteissa ja merkityksissä (ks. Lyons 1995: 61). Kun puhutaan merkityksestä yleisemmällä tasolla, yleensä erotellaan päämerkitys, eli denotatiivinen merkitys sivumerkityksistä, eli assosiatiiivisista merkityksistä (Kuiri 2012: 28). Leech (1981: 9–23) käyttää denotatiivisesta merkityksestä termiä konseptuaalinen merkitys ja jaottelee sivumerkitykset edelleen viiteen alaryhmään: konnotatiivinen, sosiaalinen, affektiivinen, reflektiivinen ja kollokatiivinen merkitys. Näin ollen synonyymeilla on usein sama denotatiivinen merkitys, mutta ne enemmän tai vähemmän eroavat muissa sivumerkityksissään. Kangasniemi selittää, että jos kielessä esiintyy täydellinen synonymia, on todennäköistä, että toinen sana korvautuu toisella, tai että toisessa merkitys muuttuu. (Kangasniemi 1997: 42.) Kun puhutaan synonymiasta, tarkoitetaan yleensä yhden kielen semanttisesti lähellä olevia sanoja. Eri kielten käännösvastineiden välinen suhde kuitenkin muistuttaa synonymiaa siinä mielessä, että molemmat sanat oletettavasti

viittaavat ainakin osittain samaan denotatiiviseen kohteeseen, mutta niiden lisämerkitykset ja merkitysvivahteet eivät todennäköisesti ole samoja.

Synonymialla kuvataan eri sanojen merkitysten suhdetta, kun taas polysemian puolesta kyseessä on lekseemin sisäinen ominaisuus. Yleisesti ottaen sana on polyseeminen, kun sille löytyy enemmän kuin yksi denotaatio, eli kun sanalla on useampia alamerkityksiä. (Karlsson 2006: 213.) Vaikka sanojen monimerkityksisyys ei yleensä aiheuta ongelmia ihmisten viestinnässä, polysemiaa ei ole yksiselitteistä soveltaa esimerkiksi käänntieteessä tai leksikografiassa. Eri sanakirjat määrittelevät lekseemien eri merkityksiä, eli niiden polyseemisuuden asetta eri tavalla ja yksittäisten alamerkitysten sisällöt usein eroavat. (Ravin & Leacock 2002: 1.)

Ensiksi on tärkeää erotella toisistaan polysemia ja homonymia. Homonymia perinteisesti määritellään tilanteeksi, jossa kahdella lekseemillä on eri merkitys, mutta niillä on ikään kuin sattumalta sama muoto, kuten puulaji KUUSI ja luku KUUSI. Keskeinen ero polysemian ja homonymian välillä liittyy siihen, että homonyymeillä ei ole yhteistä lähtökohtaa, eli ne ovat aina muodostaneet eri lekseemejä. (Kuiri 2012: 32–33.) Homonyymien erilaisen etymologisen alkuperän lisäksi toisena määritelmän vaatimuksena on, että niiden merkitysten välillä ei ole yhteyttä (Lyons 1995: 58; Kangasniemi 1997: 53). Sen sijaan polyseemisen sanan alamerkitykset yleensä liittyvät semanttisesti yhteen. Yhteenliittymistä päätellään puhujien intuition ja sanojen historiallisen kehittymisen perusteella. (Saeed 2011: 64.) Esimerkiksi sanan KIELI Kielitoimiston sanakirjan määrittelemät alamerkitykset 1. 'suuontelon elin', 2. 'kieleke, läppä', 3. 'soittimen osa', ja 4. 'puheen järjestelmä' eivät ole satunnaisia, vaan ne liittyvät toisiinsa metaforan ja metonymian kautta (Karlsson 2006: 213). Täten polyseeminen sana sanakirjoissa edustaa yhtä lekseemiä, jonka alla luetellaan sen eri merkitykset, kun taas homonyymejä on sanakirjoissa tapana erotella kahdeksi itsenäiseksi sanaksi. Cruse korostaa, että vaikka on monta tapusta, joissa on selvää, onko kyseessä polysemia vai homonymia, molempien ilmiöiden välillä ei ole selkeää rajaa ja ne muodostavat pikemminkin jatkumoa. (Cruse 2000: 109.)

Toinen sanan polyseemisuuden määrittelemisen ongelma liittyy siihen, milloin voi lekseemin merkityksiä tai käyttötapoja pitää tarpeeksi toisistaan poikkeavina, että ne edustaisivat yksittäisiä alamerkityksiä (Ravin & Leacock 2002: 3; Saeed 2011: 61). Kangasniemi selittää, että sanalla KIRJA voi viitata sekä yksittäiseen fyysiseen nidokseen että teoksen abstraktiin sisältöön, mutta pelkästään tämän perusteella ei KIRJA-lekseemiä voi pitää kaksimerkityksenä. Sanojen polyseemisuuden aste ei kuitenkaan aina ole kiistaton tosiasia ja usein riippuu siitä, kuinka hienosti merkityksiä jaotellaan, jonka takia eri sanakirjoissa lekseemien merkityskirjot eivät ole yksikäsitteisiä. (Kangasniemi 1997: 45–46.) Sanan monimerkityksisyyden

määrittämistä voi helpottaa esimerkiksi rinnastustestien soveltaminen (ks. Cruse 2000: 106–109; Kangasniemi 1997: 45–46). Ottaen huomioon tutkielmani tutkimuskysymykset, on tärkeää vertailla lekseemien PRÍTEL ja YSTÄVÄ merkityslaajentumaa ja korpusaineston avulla tarkastella eri merkitysten yleisyyttä, eikä siis tarvitse tarttua merkityksen määritelmän yksityiskohtiin. Sen takia aineistosanojen merkityskirjon selvittämiseksi nojaudun nykysanakirjojen antamaan selitykseen (tarkemmin luvussa 3.1.2).

Jokainen vieraan kielen oppija jossain vaiheessa huomaa, että yhden kielen vastineella ei ole välttämättä samaa merkityslaajentumaa kuin toisessa kielessä. Niin esimerkiksi suomen verbiä PAISTAA voi käyttää sekä ilmaisussa *aurinko paistaa* että *paistaa kalaa*, mutta englannin kielessä on kumpikin merkitys toteutettu eri vastineella, jotka ovat SHINE ja FRY (Kangasniemi 1997: 47). Kielissä on siis jaoteltu semanttinen avaruus toisistaan poikkeavin tavoin, mikä näkyy myös käännösvastineiden eri merkitysjärjestelmissä kaksikielisissä sanakirjoissa (Altenberg & Granger 2002: 21). Näin ollen kahden kielen käännösvastineet voivat jakaa samoja alamerkityksiä, mutta näistä yksi tai molemmat sanat voivat myös esittää merkityksiä, joita ei löydy toisen kielen sanasta (ks. François 2008: 166–167). Altenberg ja Granger (2002: 22) erottelevat kahden kielen lekseemien vertailussa päällekkäistä ja eroavaa polysemiaa (*overlapping/diverging polysemy*). Päällekkäinen polysemia, eli tilanne, jossa kaksi lekseemiä ilmaisevat suurin piirtein samoja merkityksiä, on suhteellisen harvinainen. Toisaalta eroavaa polysemiaa, joka tarkoittaa, että kahden kielen käännösvastineilla on erilaisia alamerkityksiä, kohdataan vertailevassa tutkimuksessa erittäin usein. Tämä pitää paikkaansa etenkin, kun vertailukohteena ovat kielen suurtaajuiset sanat, ennen kaikkea verbit, koska ne ovat usein hyvin polyseemisiä. (Mts. 21–23.)

Sanojen merkitykset eivät ole staattisia vakioita, vaan merkitykset ajan myötä muuttuvat laajenemalla, supistumalla tai kokonaan siirtymällä (Kangasniemi 1997: 75). Semanttisen muutoksen tarkempiin periaatteisiin perehtyminen on tutkielmani näkökulmasta tarpeetonta, mutta mainittakoon ainakin metafora, joka on yleisin tapa merkityksen laajentamiseen (ks. Cruse 2000: 202–211). Merkityksen muutos metaforan kautta perustuu kahden tarkoitteen yhteneväiseen piirteeseen, joten yhteen tarkoitteeseen viittaavaa sanaa käytetään myös toisesta tarkoitteesta. Alkuperäinen merkitys yleensä viittaa konkreettiseen ominaisuuteen tai toimintaan, josta metaforisessa siirrossa tulee abstraktimpi. (Robert 2008: 61–62.) Jossain vaiheessa sanalla, jota käytetään metaforisesti, on siis sekä vanhaan että uuteen tarkoitteeseen viittaava merkitys, mikä johtaa polysemiaan. Joskus voi käydä niin, että uudesta merkityksestä tulee sanan päämerkitys ja vanha merkitys poistuu kokonaan käytöstä. (Sweetser 1990: 9; Kuiri 2012: 45.) Monet yleiset kielimetaphorat liittyvät luontoon ja eläimiin: ihmistä voi nimetä

käärmeeksi, siaksi tai leijonaksi, mikä tarkoittaa, että puhuja ilmaisee ihmisen ja tietyn eläimen jakavan yhteisen ominaisuuden (Kangasniemi 1997: 79). Joihinkin tiettyihin käsitteisiin liittyy runsaasti metaforia. Kyseessä ovat usein tabuaiheet, kuten sukupuolisuus tai kuolema, joita syystä tai toisesta ei haluta kutsua tarkalla nimellä. Näin esimerkiksi kuolemasta käytetään metaforaa *nukkua pois* tai *nukahtaa ikuiseen uneen*. (Kuiri 2012: 46.)

Sanan polyseemisyys harvoin tuottaa ongelmia puheen tai tekstin ymmärtämisessä, koska kielenpuhuja tunnistaa, mitä merkitystä ilmaistaan sanan leksikaalisesta ja kieliopillisesta ympäristöstä. Sillä merkitys ja kielenrakenne ovat kietoutuneet yhteen, lekseemin erilaiset merkitykset toteutuvat myös syntagmaattisten suhteiden tasolla. Sanan eri merkityksiin siis liittyvät tietyt kollokaatiot ja kieliopilliset rakenteet. (ks. Sinclair 1991: 104–105; Yarowsky 1993.) Ennen kollokaatioanalyysiä koen oleelliseksi tunnistaa valittujen lekseemien eri merkitykset. Kun käytetään miljoonista sanoista koostuvaa aineistoa, ei ole mahdollista erotella yksittäisiä konkordansseja merkitysten mukaisesti ja analysoida niitä erikseen. Kollokaattien käsittelyssä kuitenkin otan huomioon lekseemien eri alamerkitykset ja käyttötavat.

3 AINEISTO JA MENETELMÄT

Tämän luvun alussa esitän vertailevaan fraseologiseen tutkimukseen liittyvät ongelmat ja lähestymistavat sekä monikielisten korpusten typologiaa. Sen jälkeen keskityn käsiteltävien leksiemien valinnan perusteluun, aineistojen kuvailuun ja käyttämäni tutkimusmenetelmien esittämiseen. Aineistojen osalta tutkielmani pohjautuu suomen ja tšekin sanomalehtiteksteistä koostuviin korpuksiin ja apuaineistona käytän suomi–tšekki rinnakkaiskorpusta sekä sanakirjoja. Aineiston analyysimenetelmänä hyödynnän pääosin tilastollisia testejä kollokaatioanalyysiä varten.

Pian sen jälkeen, kun fraseologinen tutkimus sai alkunsa Euroopassa viime vuosisadan puolivälissä, nousi kiinnostus vertailla fraseologisia ilmiöitä myös eri kielten välillä (Colson 2008: 192). Varhaisen vertailevan fraseologisen tutkimuksen keskiössä olivat ennen kaikkea idiomit ja sananlaskut (Ebeling & Ebeling 2013: 6). Näiden figuratiivisten ilmausten tutkimusta on motivoinut muun muassa oletettu yhteys idiomien ja ihmiskulttuurin välillä. Piirainen kuitenkin huomauttaa, ettei kulttuuripiirteiden selittäminen fraseologisten ilmaisujen kautta ole yksiselitteistä. (Piirainen 2008: 219–221.) Korpustutkimuksen havainto, että idiomit ovat kvantitatiivisesta näkökulmasta itse asiassa hyvinkin marginaalinen osa kieltä, sytytti asennemuutoksen muita fraseologisia piirteitä kohti. Fraseologisen tutkimuksen ytimeen on muun muassa Firthin ja Sinclairin ajatusten ansiosta noussut leksikaaliset myötäesiintymät, etenkin kollokaatio. (Colson 2008: 197.) Kielten vertailussa käytetään monikielisiä korpuksia, joita hyödynnetään kontrastiivisen tutkimuksen lisäksi myös esimerkiksi käännöstieteessä (Johansson 2007: 4).

Vertaileva fraseologinen tutkimus on osa kontrastiivista lingvistiikkaa (jatkossa KL) (Altenberg & Granger 2002: 5). Varhaisen KL:n tarkoitus oli vertailla kielten eri piirteitä ja näin ollen selittää kielenmuutosta ja kielten typologiaa, eli kuulumista samaan- tai eri kielikuntaan (Lyons 1968: 21–22). KL koki kukoistuskautensa 1960-luvulla, kun sitä pidettiin soveltavan kielitieteen haarana, jonka avulla yritettiin ratkaista ongelmia vieraan kielen oppimisessa ja opettamisessa kielten yhtäläisyyksien ja eroavuuksien perusteella. 2000-luvun alussa Altenberg ja Granger toteavat KL:n olleen uudestaan kielitieteen vallitsevassa asemassa ja luettelevat tähän kehitykseen kolme pääsyitä. Ensiksi Euroopan asteittainen kansainvälistyminen oli vaatinut monikielistä pätevyyttä käännös- ja kielenopetustarpeisiin, mikä oli luonut kysyntää monikieliselle tutkimukselle. Toiseksi yleisen kielitieteen suuntauksen muutos on alkanut suosia luonnollisen kielenkäytön tutkimusta sekä kvantitatiivista dataa havainnoinnin todisteena

kielen abstraktin ja staattisemman käsityksen sijaan. Tähän liittyykin sanaston ja kontekstuaalisen valinnan nousu kielitieteellisen tutkimuksen keskiöön (ks. tarkemmin luku 2.1.2). Kolmantena syynä oli digitaalinen vallankumous, joka heijastui lingvistiikkaan etenkin mahdollisuutena tutkia autenttisen kielen laajaa aineistoa korpusten avulla. (Altenberg & Granger 2002: 5–7.)

KL ei ole teoreettinen koulukunta sen yleisessä merkityksessä, eikä se keskity tiettyyn kielen osa-alueen tutkimiseen. J. Ebeling ja S. Ebeling pitävät KL:a pikemminkin tutkimusmenetelmänä korostaen, että usein käytetään myös termiä kontrastiivinen analyysi. Sen lisäksi KL:n tutkimusala jakautuu sekä soveltavaan että teoreettiseen suuntaukseen. (Ebeling & Ebeling 2013: 13.) Nykyinen fraseologinen ja kontrastiivinen tutkimus on luonnollisesti sidoksissa korpuslingvistiikkaan.

Edellä mainitussa todettiin, että korpus on autenttisen kielen lähde, joka antaa mahdollisuuden tarkastella mittavaa kieliaineistoa ja kvantitatiivisesti mitata eri ilmiöiden taajuutta. Vertailevassa korpustutkimuksessa erotellaan kaksi korpuksen päätyyppiä, joita hyödynnetään tutkimusaiheen ja päämäärän perusteella. Nämä ovat verrannollinen korpus (*comparable corpus*) ja käännös- eli rinnakkaiskorpus (*parallel corpus*). McEnery ja Xiao (2008: 19) toteavat monikielisten korpusten soveltamisen olevan suhteellisen uusi ilmiö, jonka vuoksi korpustyyppien terminologia on osittain hämmentävää ja vakiintumatonta. Näin ollen tutkimuksissa on aiemmin käytetty termiä rinnakkaiskorpus sekä verrannollisesta että käännöskorpuksesta (mp.; Johansson 2007: 9). On siis syytä tarkasti määritellä millä termillä mitään korpustyyppiä kutsutaan. Jäsenän tutkielmani esim. Bakerin (1998), McEneryn ja Xiaon (2008) sekä Jantusen (2011) käyttämien käsitteiden ja termien avulla.

Käännöskorpuksella tarkoitetaan korpusta, joka koostuu lähdekielen alkuperäisistä teksteistä ja niiden käännöksistä toiseen tai useampaan kieleen. Käännöskorpus voi olla yksisuuntainen sisältäen käännöksiä vain yhdestä kielestä toiseen, tai kaksisuuntainen, josta nimensä mukaisesti löytyvät myös käännökset toisesta kielestä ensimmäiseen. (McEnery & Xiao 2008: 20.) Tässä mielessä on käyttämäni korpus InterCorp kaksisuuntainen, sillä se sisältää sekä tšekiksi käännettyjä suomenkielisiä alkutekstejä, että alun perin tšekinkielisiä tekstejä ja niiden suomennoksia.³ Pelkästään tämäntyyppiseen korpukseen viitataan tässä tutkimuksessa termillä rinnakkaiskorpus, ja termejä käännöskorpus sekä rinnakkaiskorpus käytän synonyymisesti. Rinnakkaiskorpuksen suurin hyöty on juuri käännöstieteessä, sillä sen avulla on

³ Tarkkaan ottaen InterCorp-korpus on monisuuntainen, sillä se sisältää tekstejä, jotka on käännetty useampaan eri kieleen. Tutkimuksessani kuitenkin käytän ainoastaan suomen- ja tšekinkielisiä tekstejä.

mahdollista tutkia, miten samaa sisältöä ilmaistaan toisessa kielessä (mts. 21). Käännöskieli kuitenkin monella tavalla eroaa alkuperäiskielestä. Yhtenä ongelmana on se, että käännös aina vaikuttaa kohdekieleen, joten käännetystä tekstistä löytyy lähdekielen jälkiä. (ks. Baker 1993; Mauranen & Kujamäki 2004.). Pelkkä rinnakkaiskorpus ei siis yleensä ole hyvä pohja kahden kielen kontrastiiviseen tutkimukseen, mutta sitä voi hyödyntää analyysin alkuvaiheessa vahvistamaan käännösvastineiden vastaavuutta (Johansson 2007: 9).

Verrannollinen korpus toisaalta koostuu alkuperäisistä teksteistä kahdella tai useammalla kielellä, jotka vastaavat toisiaan tekstilajin, aiheen tai rekisterin osalta (Altenberg & Granger 2002: 8; McEnery & Hardie 2012: 20). Verrannollisella korpuksella voidaan viitata myös aineistoon, joka sisältää yhden kielen useampia variantteja, esimerkiksi alkuperäis- ja käännösuomen (Jantunen 2011: 90). Tutkielmani verrannollinen korpus ei kuitenkaan koostu käännöksistä ollenkaan, vaan suomen ja tšekin toisistaan riippumattomista teksteistä, eli kahdesta yksikielisestä korpuksesta. Tämän tyyppisen korpuksen etu on luonnollisen kielen esiintyminen, johon ei ole vaikuttanut käännös. Sitä vastoin on kuitenkin kyseenalaista, missä määrin kielidata on oikeasti vertailevaa sen funktion ja stilistiikan osalta. Yhtenä mahdollisuutena riskien pienentämiseen on käyttää hyvin laajaa tasapainotettua korpusta, missä voi hallita muutujia kuten aihe ja rekisteri. (Johansson 2007: 10; McEnery & Hardie 2012: 20–21.)

Molemmasta monikielisen korpuksen tyypistä löytyy selkeitä etuja sekä puutteita, joten niiden käyttöä kannattaa pohtia tutkimustarpeiden mukaisesti. Monessa tutkimuksessa hyödynnetään molempia korpuksia tai näiden sekoitusta, millä pyritään tasapainottamaan niiden vahvuuksia ja heikkouksia. Näin muun muassa Mauranen (2002: 182) toteaa, että yhtenä käännöskorpuksen hyötynä on nimenomaan se, että se tuottaa uusia hypoteeseja sekä lähde- että kohdekielestä, joita voi tutkiskella syvemmin yksikielisten korpusten kautta. Tognini-Bonelli (2001: 6, 133) myös suosii sekä vertailevan- että käännöskorpuksen käyttöä kontrastiivisessa tutkimuksessa, sillä ne täydentävät toisiaan.

Kontrastiivisen tutkimuksen ja analyysin olennainen ongelma juontuu vertailtavien ilmausten vastaavuuden (*equivalence*) vahvistamisesta. Kun vertaillaan kahta sanaa, pitää varmistaa, että ne ilmaisevat ja tarkoittavat suurin piirtein samaa asiaa, eli että niitä on edes mielekästä verrata. Vasta niiden yhteisen pohjan löydyttyä voi tarkastella ilmausten samanlaisuutta ja eroavuutta. Tätä yhteistä pohjaa kutsutaan termillä *tertium comparationis*. (Ebeling & Ebeling 2013: 17.) Kun sanoja vertaillaan verrannollisen korpuksen kautta, ei ole aina selkeää, peilataanko samaa asiaa. Rinnakkaiskorpus toisaalta tarjoaa mahdollisia käännösvastineita, mutta niitäkään ei voi aina pitää luotettavina. Käännöspätevyys onkin intuitiivinen, eikä kääntäessä yleensä pyritä ilmaisun kirjaimelliseen käännökseen. (Altenberg & Granger 2002: 15–

17.) Yhtenä mahdollisuutena siihen, miten tunnistetaan sanan ja sen käännösvastineen vastaavuus, on käyttää kaksisuuntaista rinnakkaiscorpusta ja mitata kuinka usein käännösvastine esiintyy lähdekielestä kohdekieleen sekä toisinpäin (mts. 17–18; Johansson 2007: 23).

Tutkielmassani käytän pääosin kokoamaani suomen ja tšekin verrannollista corpusta, jolla tutkin, missä merkityksissä lekseemit YSTÄVÄ sekä PŘÍTEL esiintyvät ja tunnistan mitkä kollokaatit ilmenevät niiden autenttisessa ympäristössä. Tämän jälkeen tutkin kollokaattien semanttista ja kontekstuaalista vastaavuutta. Hyvän ja edustavan verrannollisen korpuksen koaamista pidän tutkimukseni onnistumisen perustana. Apuvälineenä käytän suomi–tšekki-rinnakkaiscorpusta, jonka avulla vahvistan sanojen käännöksien vastaavuuden (korpuksista tarkemmin ks. luku 3.1.3).

3.1 Aineisto

Seuraavassa alaluvussa esittelen tarkemmin tutkittavat lekseemit YSTÄVÄ ja PŘÍTEL, perustelen näiden valintaa tutkimukseni keskiöksi sekä käsittelen näistä johdettuja feminiinijohdoksia (YSTÄVÄTÄR, PŘÍTELKYNĚ), jotka sisällytän analyysiin. Tämän jälkeen esittelen analyysissä käyttämäni aineiston, jonka olen valinnut niin, että se mahdollisimman hyvin vastaisi tutkielman tutkimustarpeita. Analyysin ensimmäisessä vaiheessa tarkastelen, kuinka vahvasti sanat ovat toistensa käännösekvivalentteja, sillä kahden kielen sanojen kontrastiivisessa vertailussa on olennaista vahvistaa, että sanoille löytyy jokin yhteinen pohja. Ensin selvitän, miltä tilanne näyttää kaksisuuntaisissa suomi–tšekki- ja tšekki–suomi-sanakirjoissa. Nämä kahden kielen sanakirjat ovat kuitenkin hyvin suppeita, joten käännösekvivalenssia selventäessä nojaudun pääosin käännöskorpukseen, jonka avulla tarkastelen, missä määrin tutkittavia sanoja oikeasti käytetään toistensa vastineina. Sen lisäksi olisi kiinnostavaa nähdä millaista vaihtelua käännöksissä esiintyy. Kuten edellisestä alaluvusta kävi ilmi, käännöskieli useimmiten eroaa alkuperäisestä kielestä siinä mielessä, että käännettyyn tekstiin jää lähdekielen jälkiä, joten pidän käännöskorpuksesta saatua tietoa pikemminkin suuntaa antavana.

Analyysin toisessa vaiheessa hyödynnän sekä verrannollista corpusta että yksikielisiä sanakirjoja. Edellisessä luvussa käsittelin polysemiaa, eli sanojen monimerkityksisyyttä, ja miksi sitä kannattaa pohtia ennen kollokaatioanalyysiä. Sanakirjojen avulla siis ensin määrittelen lekseemien erilaisia merkityksiä, ja katson, miten ne vastaavat tai eroavat toisistaan. Sanakirjat auttavat luomaan sanojen peruskuvan, ja voivat antaa vihjeitä mihin kannattaa kiinnittää huomiota kollokaatioanalyysissä. Sanakirjoissa listatut sanojen selitykset ja kuvailut eivät

kuitenkaan aina vastaa kielen todellista käyttöä (Sinclair 1991: 7). Sen vuoksi otan tässä vaiheessa käyttöön verrannollisen korpuksen, jolla tarkastelen missä merkityksissä lekseemit YSTÄVÄ sekä PŘÍTEL esiintyvät ja miten korpusaineiston sisältämä luonnollinen kielenkäyttö vastaa sanakirjojen määritelmiä. Koska sanakirjat eivät aina voi tarjota ajan tasalla olevaa tietoa kielestä, näen tärkeänä soveltaa korpusta päätelläkseni lekseemien todellisen merkitysjaon.

Verrannolliseen korpukseen tukeudun myös kollokaatioanalyysissä, eli analyysin kolmannessa vaiheessa, joka on tutkielmani ytimessä. Tätä alkuperäisistä teksteistä koostuvaa korpusta olen laatinut suomen yksikielisestä ja tšekin yksikielisestä korpuksesta. Vertailevan korpuksen laatimisessa on kuitenkin olennaista varmistaa sen osien samankaltainen edustavuus, tasapaino ja otoskehys, toisin sanoen se, että korpukset koostuvat samasta tekstilajista, ajankohdasta ja aihepiiristä (McEnery & Xiao 2008: 20). Vaikka kummastakin kielestä löytyy melko laaja monenlaisten korpusten tarjonta, ei ole yksinkertaista valita korpuksia, jotka täysin vastaisivat toisiansa kaikissa vaadituissa aspekteissa niin, ettei tarvitsisi muokata jo valmiiksi tehtyjä korpuksia. Suomen kielen tekstikokoelma -korpus sekä tšekin SYNPUB2006-korpus koostuvat molemmat journalistisista teksteistä, jotka ovat peräisin 1990-luvulta, mutta tšekin tekstit lisäksi 2000-luvun alusta. Tšekin aineisto on myös laajuudeltaan kaksinkertainen suomen korpukseen verrattuna, jonka en kuitenkaan koe haittaavan, sillä kollokaatioanalyysin tilastollisissa testeissä sovelletaan suhteellisia mittakaavoja. Sen lisäksi molemmat korpukset ovat melko kattavat: suomen korpus koostuu 160 miljoonasta ja tšekin 300 miljoonasta sanasta.⁴ Altenberg ja Granger (2002: 8) toteavatkin, että laajat vertailevat korpukset ovat hyödyllisiä juuri kollokaatiotutkimuksessa, ja korpuksen koko sekä suuret datamäärät ovat erittäin merkittäviä kontrastiivisessa sanastotutkimuksessa.

Sanomalehtiaineiston valitsin siksi, että lehtikieli on mahdollisimman neutraalia ja noudattaa standardikielen normeja tekstien kohdeyleisön laajuuden vuoksi (ks. Koivusalo 1992). Kirjoittajien idiosynkraattisia valintoja on vältelty esimerkiksi kaunokirjallisiin teksteihin verrattuna. En siis oletta kummankaan kielen kielenkäytön poikkeavan standardista, mikä helpottaa niiden keskinäistä vertailua. Vaikka aineistoni onkin jo kaksikymmentä vuotta vanha, ei tämän pitäisi vähentää tulosten tulkinnan pätevyyttä verrattuna nykyhetkeen, sillä lehtikielen muuttuminen on melko hidas prosessi (ks. Mäkinen 1992: 214).

Korpustutkimuksessa keskeiseksi ongelmaksi nousee esille korpuksen edustavuus. Korpusaineiston valinnassa on oleellista pohtia vastaako valittu korpus tarpeeksi hyvin tutkimustarpeita, ja voiko korpuksen, joka on poikkeuksetta aina jollakin tavoin rajattu kieliaineisto,

⁴ Sane (*word-token*) viittaa korpuslingvistiikan puitteissa yksittäiseen sananmuodon esiintymään.

perusteella tehdä yleistäviä päätelmiä kielestä. (ks. Biber, Conrad & Reppen 1998: 246.) McEnery ja Hardie (2012: 10) toteavat, että korpuksen täydellistä edustavuutta ja sen sisältämien tekstien tasapainoa pystyy harvoin saavuttamaan korpusta kootessa. Tutkimuksessani valitsin pelkästään sanomalehtikielestä koostuvan verrannollisen korpuksen, koska näin pystyy paremmin varmistamaan, että osakorpuksat vastaavat toisiaan ajanjakson ja laajuuden osalta. Näen siis tärkeämpänä, että suomen ja tšekin korpuksat ovat hyvin verrattavissa, kuin että tavoittelisin kattamaan useampia kielivariantteja, mikä ei muutenkaan olisi tutkielman rajallisen laajuuden vuoksi mahdollista. Tulosten tulkinnassa kuitenkin pidän mielessä, että tekstilaji väistämättä vaikuttaa fraseologisiin ilmiöihin, ja että kollokaatteja on selitettävä kyseisen tekstilajin puitteissa.

3.1.1 Lekseemien valinta ja perustelut

Tutkittavien lekseemien YSTÄVÄ ja PŘÍTEL valinnan yhtenä periaatteena on niiden suhteellinen yleisyys molemmissa kielissä. Suomen 9996 yleisintä lemmaa sisältävän vuoden 2004 Suomen sanomalehtikielen taajuussanaston (CSCb) mukaan sana YSTÄVÄ löytyy sijalta 817. Tšekin vuoden 2010 taajuussanakirjan mukaan lemma PŘÍTEL sijaitsee 619. paikalla (Čermák 2010). Niiden hyvin korkea taajuus teksteissä mahdollistaa kattavan aineiston keräämisen ja näin ollen antaa suuren määrän kielidataa analyysiä varten. Kyseessä on siis usein käytetty substantiivi, joka kuuluu kielen perussanastoon. Toisena perusteena havaintoyksikköjen valintaan on oletus siitä, että ne ovat toistensa lähivastineita. Tätä toki käsitellään tarkemmin seuraavissa luvuissa. Ensikatsauksella kuitenkin vaikuttaa siltä, että sanat ovat hyvin lähellä toisiaan perusmerkityksissään ja molemmat edustavat yleiskielen rekisteriä.

Tarkasteltavat lekseemit siis viittaavat konkreettiseen fyysiseen olioon, toista lähellä olevaan ihmiseen (tai muuhun olentoon). Molemmilta löytyy myös lähisynonyymi, joka on luonteeltaan puhekielisempi ja kuvaa merkitykseltään väljempää ihmissuhdetta (suom. KAVERI (1193)⁵, tšek. KAMARÁD (1938)). Synonyymisanakirjoissa⁶ on sanoille listattu muitakin merkitykseltään lähempänä tai kauempana olevia sanoja: suomenkieliset: KUMPPANI (2908), TOVERI (6643), KUOMA, SYDÄNYSTÄVÄ; tšekinkieliset: DRUH (443)⁷, SPOLEČNÍK (3324), BLIŽNÍ

⁵ Sulkeissa lemmaojen sijainti CSCb:n tai Čermákin (2004) taajuussanastoissa. Sanat ilman sijaintinumeroa eivät löydy taajuussanastoista.

⁶ Suomen synonyymisanasto – Leino & Leino 1990: Synonyymi-sanasto; Tšekin synonyymisanasto: Bečka 1979: Slovník synonym a frazeologismů.

⁷ Korkea frekvenssi johtuu sanan homonymiasta, tarkoittaen lisäksi LAJIA/TYYPPIÄ; sanan PŘÍTEL synonyymina se on vanhahtava, nykyään harvoin käytetty (SSJČ, s.v. *druh*)

(4688). Koska keskusteluissa usein syntyy tarve ilmaista ihmissuhteen luonnetta ja viitata läheiseen ihmiseen, löytyy sanoille myös runsaasti eri ajanjaksojen vaihtelevaa slangisanastoa (suom.: FRENDI, KAVERUS, KAMU, KAIFARI; tšek.: KÁMOŠ, KÁMO, KUMPÁN). Kuten sanojen taajuuden perusteella voi olettaa, kaikki lekseemien YSTÄVÄ ja PŘÍTEL lähisynonyymit ovat merkittävästi vähemmän yleisiä. Puhekielisempien sanojen, kuten KAVERI/KAMARÁD tapauksessa se taitaa johtua siitä, että taajuussanasto on koottu kirjoitettujen, muodollisempia ilmaisuja suosivien tekstien mukaisesti. Oletuksenani kuitenkin on, ettei mikään yllä listatuista synonyymeistä lähesty ihmisten välisen suhteen vakavuutta ja ainutkertaisuutta samassa mielessä kuin YSTÄVÄ/PŘÍTEL.

YSTÄVÄ/PŘÍTEL ovat siis konkreettisia substantiiveja, joilla on kielenulkoisen referentti. Niiden merkitys voi kuitenkin olla abstraktimpi, kuten seuraavassa tapauksessa.

1) *Hyvinvointivaltio naisen paras ystävä.*

Esimerkkilauseessa (1) sanaa YSTÄVÄ käytetään metaforisesti, kun hyvinvointivaltiolle tunnustetaan samankaltaisia ominaisuuksia kuten ystävällä oletetaan olevan. Kyseessä on dia-kroninen merkityssiirtymä, joka aiheuttaa sanan polysemiaa, jonka tyypillisin laji on juuri metaforinen polysemia (Blank 2003: 268). Sanat YSTÄVÄ ja PŘÍTEL siis esiintyvät useissa eri merkityksissä, eikä kyseessä ole aina konkreettinen henkilö (lekseemien eri alamerkityksiä käsitel- telen tarkemmin luvussa 4.2).

Suomi ja tšekki kuuluvat genealogisesti eri kielikuntiin: suomi on uralilaisten itämeren-suomalaisten kielten jäsen, ja tšekki on indoeurooppalaisen kielikunnan länsislaavilainen kieli. Niin kuin muilla kielillä, näiden kahdenkin välillä löytyy eroavuuksia ja jonkin verran samana-laisuksia. Kieliä perinteisesti luokitellaan myös typologisesti näiden eri rakennepiirteiden mu- kaan. Morfologian kannalta suomi edustaa agglutinoivaa kieltä, jolle on tyypillistä, että yksit- täisillä morfeemeilla on yksi funktio ja niitä liimataan peräkkäin sanavartaloon (Karlsson 2004: 116). Tšekistä taas puhutaan flekteeraavana tai fuusiovana kielenä, jossa päätteillä voi olla use- ampikin funktio, joten yksittäisten morfeemien erottelu on hankalaa. Korostetaan kuitenkin, ettei mikään kieli esitä puhtaasti tiettyä typologista ryhmää, vaan kielillä on enemmän tai vä- hemmän eri ryhmiin kuuluvia piirteitä. (Čermák 2007: 203–205.) Itkonen (2001: 54) seuraa toisenlaista jaottelua analyttisiin, synteettisiin ja polysynteettisiin kieliin, jolla luonnehditaan kielten morfologisia sekä syntaktisia ominaisuuksia. Synteettisten kielten ryhmälle, johon kuu- luu sekä suomi että tšekki, on ominaista nominien sekä verbien taivutus, niiden välinen kong- ruenssi ja rektiovaatimus (mts. 60–63). Näin ollen suomen ja tšekin nominien taivutukseen kuuluu luku ja eri määrä sijoja. Toisaalta tšekissä, kuten muissakin flekteeraavissa kielissä,

nomineilla on taivutuskategoriasta suku, joka on tšekissä tapauksessa joko maskuliini, feminiini tai neutri (Čermák 2007: 127). Kieliopillisen suvun ja biologisen sukupuolen välillä onkin osittainen yhteys, niin esimerkiksi sana MUŽ (mies) on maskuliini ja ŽENA (nainen) feminiini (mp.).

Karlsson (2004: 257) huomauttaa, että eri kielissä miessukupuolinen sana on useimmiten tunnusmerkitön ja naissukupuolinen tunnusmerkkinen, jota ilmaistaan mutkikkaammalla muodolla. Niin sekä suomessa että tšekissä naispuolista henkilöä edustavaan sanaan lisätään suffiksi (esim. KUNINGAS – KUNINGATAR, KRÁL – KRÁLOVNA). Tämän tutkimuksen kannalta olennainen ero suomen ja tšekin kielen välillä on siinä, että tšekissä on pakollista ilmaista nais- sukupuolta johtimella sen kieliopillisen suvun takia. Sen lisäksi suku ilmenee myös kongruenssissa verbien ja muiden nominien kanssa (Čermák 2007: 128). Korpuksimerkki (2) havainnollistaa, miten koko lauseen päätteet muuttuvat, kun vaihdetaan sukua:

fem. přítelkyně: 2) *Když jedna přítelkyně objevila její fotografii v novinách, vyděsila se.*

mask. přítel: *Když jeden přítel objevil jeho fotografii v novinách, vyděsil se.*

[Knj][Num][N-NOM][V-PRET⁸][Pron][N-AKK][P][N-LOK][V-PRET]

(Kun eräs ystävä huomasi hänen kuvansa sanomalehdessä, hän säikähti.)

Suomen kielessä toisaalta feminiinijohdos *-tAr* on ylipäätään melko harvinainen ja vanhahtava. Ison suomen kieliopin mukaan sitä kuitenkin joissakin sanoissa käytetään, kuten NÄYTTELIJÄTÄR, KAUNOTAR tai YSTÄVÄTÄR. (VISK § 193.) Vaikka sana YSTÄVÄTÄR toisinaan esiintyy, sanalla YSTÄVÄ viitataan suomeksi luonnollisesti sekä nais- että miespuolisiin henkilöihin. Koska miessukupuolinen sana on tšekissä useimmiten tunnusmerkitön ja geneerinen, voi joissakin tapauksissa (3) käyttää sanaa PŘÍTEL eikä PŘÍTELKYNĚ, vaikka kyseessä on feminiini:

3) *Žena je prý nejlepším přítelem člověka.*

[Nainen on kuulemma ihmisen paras ystävä.]

Tätä käyttötapaa voi kuitenkin pitää pikemminkin poikkeuksena ja useimmiten naispuoliseen henkilöön viittaaminen sanalla PŘÍTEL olisi kieliopillisesti väärin. Mainittujen suomen ja tšekin ominaisuuksien kannalta ei yllätä sanojen YSTÄVÄTÄR ja PŘÍTELKYNĚ esiintymien epätasainen määrä korpuksessa. Valitsemani korpusten eri koot huomioituna esiintyy aineistossa lemma PŘÍTELKYNĚ lähes seitsemän kertaa useammin kuin YSTÄVÄTÄR.

⁸ Mennyt aikamuoto preteriti on hyvin verrattavissa suomen imperfektiin.

Tässä luvussa kerrotun perusteella pidän asiaankuuluvana sisällyttää analyysiin myös lekseemeistä YSTÄVÄ/PŘÍTEL johdetut feminiinijohdokset YSTÄVÄTÄR/PŘÍTELKYNĚ. Jos nämä jätettäisiin analyysin ulkopuolelle, tšekin puolelta ei todennäköisesti päätyisi analyysiin niitä kollokaatteja, jotka ilmenevät sanan PŘÍTEL kotekstissa, kun sillä viitataan naishenkilöön. Jos tämäntyyppisiin kollokaatteihin pääsisi käsiksi ainoastaan suomen kielessä, syntyisi aineistossa epätasapaino, joka mahdollisesti vääristäisi tuloksia. Välttyäkseni toistolta käytän lekseemeistä kirjoitettaessa useimmiten vain merkintää YSTÄVÄ/PŘÍTEL, johon kuitenkin sisältyvät myös feminiinijohdokset YSTÄVÄTÄR/PŘÍTELKYNĚ.

3.1.2 Suomen ja tšekin sanakirjat

Sanakirjojen tarkastelun tarkoituksena on perehtyä sanojen YSTÄVÄ ja PŘÍTEL merkitysten ja käyttötapojen kuvailuun, jotta saatuja tietoja voisi sitten verrata korpusaineistoon. Korpus koostuu 90-luvun teksteistä, joten pyrin käyttämään senaikaista tai ajankohtaista sanakirjaa.

Suomen kielen puolesta valitsin Kielitoimiston sanakirjan ja Nykysuomen sanakirjan. Näistä kahdesta uudempi, vuoden 2021 Kielitoimiston sanakirjan sähköinen versio on Kotimaisten kielten keskuksen laatima suomen kielen sanakirja, joka esittää nyky-yleiskielen keskeisen sanaston ja sisältää yli 100 000 hakusanaa. Suomalaisen Kirjallisuuden Seuran vuosina 1951–1961 toimittama Nykysuomen sanakirja on KS:n edeltäjä ja suomen kielen laajin sanakirja. Se on luonteeltaan kattavampi, sisältäen yli 200 000 hakusanaa. Se kuvaa etenkin viime vuosisadan alkupuolen kielenkäyttöä ja sisältää suuren määrän murteellisia ilmaisuja ja nykyisin jo vanhahtavia sanoja (Haarala 1994). Vaikka Nykysuomen sanakirjan antama tieto on pikemminkin historiallista, kahden eri ajanjakson kuvaus sanasta voi kuitenkin tuottaa ymmärrystä sanan merkitysten kehittymisestä.

Tšekistä valitsin sanan PŘÍTEL merkitysten selventämiseen Tšekin Tiedeakatemian vuosina 1960–1971 julkaiseman sanakirjan Slovník spisovného jazyka českého, joka koostuu lähes 200 000 hakusanasta. Sanakirja on kooltaan tšekin toiseksi suurin, ja sen tarkoituksena oli kuvailla senaikaisia sanavaroja sekä kodifioida kieltä. Tutkielmani kirjoitushetkellä ei ole saatavilla tšekin kielitoimiston virallista nykyaikaista sanakirjaa lukuun ottamatta uutta Akateemista nykytšekin sanakirjaa, joka on ollut työn alla vuodesta 2012 lähtien, ja josta on julkaistu vasta ensimmäinen osa. Smejkalová (2014) pohdiskelee artikkelissaan sanan PŘÍTEL eri merkityksiä tähdentäen, ettei sanakirjoissa ilmene avopuolisoa tai kumppania, jotka ovat nykyään sanan tavallisia käyttötapoja. Koska Slovník spisovného jazyka českého ei siis heijasta sanan PŘÍTEL ajankohtaista käyttöä, tarkastelen myös Lingea-kustantamon kaupallista vuoden 2011

sanakirjaa Slovník současné češtiny, vaikka se on saanut joitakin negatiivisia arvosteluja sen korkean virhemäärän ja epäsystemaattisuuden takia (ks. Čermáková 2013; Bezdíčková 2014).

Käännösvastaavuutta tutkiessa otan rinnakkaiskorpuksen lisäksi käyttöön myös kaksisuuntaiset suomi–tšekki ja tšekki–suomi sanakirjat. Niitä on saatavilla kaksi, vuonna 1984 WSOY:n julkaisema Suomi–tšekki–suomi taskusanakirja ja vuoden 2016 Finsko-český, česko-finský šikovný slovník Lingea-kustantamosta. Koska kooltaan pienet kaksikieliset sanakirjat ovat ennen kaikkea tarkoitettu kielenoppijoille ja turisteille, ei voi odottaa laajempaa merkitysten selittämistä. Käännösvastaavuuden selvityksen painopiste on siis käännöskorpuksessa.

3.1.3 Korpusaineisto

3.1.3.1 Rinnakkaiskorpus InterCorp

Korpus InterCorp on Český národní korpusin⁹ hanke, jonka tavoitteena on rakentaa laaja synkroninen rinnakkaiskorpus, joka sisältäisi mahdollisimman paljon eri kieliä. Se koostuu alkupe-
räisistä teksteistä ja niiden käännöksistä, jotka on tasattu useammalla, esim. sanan, lauseen tai kappaleen tasolla. (Příručka ČNK.) Tässä tutkimuksessa käytän vuoden 2018 InterCorp version 11 suomi–tšekki- ja tšekki–suomi-aineistoja, joiden avulla tarkastelen, missä määrin valitut lekseemit ovat toistensa käännösekvivalentteja.

Korpus InterCorp koostuu erilaisista tekstilajeista ja alakorpuksista, mm. kaunokirjallisuuden käännöksistä, Euroopan parlamentin tekstien Europarl-korpuksesta, Raamatun käännöksistä ja elokuvien tekstityksistä. Suomen ja tšekin käännösten käsittelyssä hyödynnän kuitenkin ainoastaan tekstejä, joiden lähtökieli on suomi tai tšekki, ja joille löytyy käännös suomeksi tai tšekiksi. Käännökset jostakin kolmannelta kielestä, kuten englannista, eivät heijasta suomen ja tšekin lekseemien välistä suhdetta, joten ne ovat tämän tutkimuksen kannalta epäolennaisia. Suomen ja tšekin aineistona käytän siis korpuksen ns. ydinosaa (*core*), joka sisältää kaunokirjallisia tekstejä. Käyttämäni alakorpus (taulukko 1) koostuu eri määrästä kirjoja, etenkin romaaneista, ajankohdaltaan ulottuen aina 1900-luvun alkupuolelle. Suurin osa alakorpukseni kirjoista kuitenkin laadittiin 1900-luvun jälkipuoliskolla.

⁹ Tšekin kansalliskorpus

Taulukko 1. InterCorp 11 -korpuksen alkuperäinen aineisto käännoksineen.

	SUOMI → TŠEKKI	TŠEKKI → SUOMI
Koko	n. 1,7 miljoonaa sanetta	n. miljoona sanetta
Kirjat	26	15
Kirjailijat	15	13
Ajanjakso	1939–2011	1923–2010
Kääntäjät	5	6
Osumat - YSTÄVÄ/PŘÍTEL	531	216
Osumat - YSTÄVÄTÄR/PŘÍTELKYNĚ	37	19

Vaikka tšekki on InterCorp-korpuksen ns. pivot-kieli, eli jokaiselle tekstille löytyy tšekinkielinen versio (käännös tai alkuperäinen teksti), aineistossani on enemmän suomenkielisiä tekstejä, joilla on tšekin käännös kuin toisinpäin. Tämä toki johtuu korpuksen koostajien päätöksistä ja valinnoista, kun päätetään, millaisia tekstejä sisällytetään korpukseen ja miksi. Toisaalta se saattaa heijastaa tilannetta, jossa suomesta tšekiksi käännetään kaunokirjallisuutta ylipäättään enemmän kuin tšekistä suomeksi.¹⁰ Muiden muuttujien, kuten ajankohdan, eri kirjojen, kirjailijoiden sekä kääntäjien määrän perusteella aineisto on toisaalta melko monipuolinen, mutta sitä ei kuitenkaan voi pitää suomen ja tšekin välisiä käännoksiä edustavana. Aineiston sisällön yksityiskohtaisemmalla tarkastelulla selviää, että jotkin tekstit ovat tässä alakorpuksessa yliedustettuja. Esimerkiksi yli puolet suomenkielisistä alkuperäisistä teksteistä koostuu Mika Waltarin ja Väinö Linnan romaaneista ja tšekinkielisissäkin tilanne on samankaltainen. Tämä on ongelmallista sen vuoksi, että yksittäisten kirjoittajien idiosynkraattisia valintoja voisi tulkita kielen yleiseksi piirteeksi. Yksittäiset kääntäjät on kuitenkin tasaisemmin siroteltu halki aineiston, mikä tasoittaa tilannetta. Alakorpukseni on lisäksi ylipäänsä suhteellisen pieni, sillä voi olettaa, ettei suomesta tšekiksi ja tšekistä suomeksi käännetä kaunokirjallisuutta yhtä paljon kuin vaikka englannista tai saksasta. Käännoksien tarkastelu voi kuitenkin antaa tietoa siitä, mikä on käännoskorpuksessa yleisin suuntaus ja mikä taas on pikemminkin poikkeuksellista.

¹⁰ Kirjallisuuden vientikeskuksen FILI:n tilaston mukaan, tšekki on ollut viime vuosina ensimmäisen kymmenen kielen joukossa, joihin suomalaista kirjallisuutta käännetään eniten, <https://www.finlit.fi/fili/kirjallisuusvienti/ti-lastoja-ja-selvityksia/#millekielille>

3.1.3.2 Verrannollinen korpus

Korpusaineiston valinnassa oli tärkeintä sen edustavuus ja sopivuus tutkimukseni tarkoitukseen, joka on tutkia lekseemien YSTÄVÄ ja PRÍTEL merkityksiä ja kontekstuaalista valintaa. Verrannollista korpusta laatiessa olen keskittynyt siihen, että sen suomen ja tšekin osakorpukset olisivat mahdollisimman hyvin verrattavissa toisiinsa. Edustavan verrannollisen korpuksen oleellisin ominaisuus onkin sen osien toisiaan vastaava näytteenottomenetelmä, täten molemmat osakorpukset koostuvat samasta tekstilajista, ja tekstien mittasuhteet sekä alkuperän ajankohdat ovat toistensa kaltaisia (McEnery & Hardie 2012: 20).

Verrannollisen korpuksen suomen osakorpuksena käytän hyödyksi Kielipankki-palvelun tarjoamaa Suomen kielen tekstikokoelmaa (SKTP). Kielipankki on korpusaineistoja ylläpitävä palvelukokonaisuus, josta vastaa suomalaisten yliopistojen, Tieteen tietotekniikan keskuksen ja Kotimaisten kielten keskuksen muodostama konsortio FIN-CLARIN. Konsortion tehtävänä on välittää tutkijoille sähköistä kieliaineistoa. (Kielipankki.) SKTP sisältää useat suomen sanomalehdet ja aikakauslehdet vuosilta 1990–2000, koostuen yhteensä noin 150 miljoonasta saneesta. Sanomalehdistä mittavimmat ovat Karjalainen, Helsingin Sanomat ja Turun Sanomat. Korpuksesta löytyy myös pieni määrä kauno- ja tietokirjoja, jotka olen kuitenkin jättänyt analyysin ulkopuolelle, jotta suomen osakorpus sisältäisi tšekin korpuksen tavoin ainoastaan journalistisia tekstejä. Kaikista Suomen sanomalehtikorpuksista SKTP koostuu ajankohdaltaan uusimmista sanomalehdistä, joten se sopii tutkimustarpeisiini parhaiten.

Tšekin puolesta hyödynnän SYNPUB2006-korpusta, jota ylläpitää Český národní korpus (ČNK) kuten myös edellisessä luvussa kuvattua käännöskorpusta. ČNK on Kaarlen yliopistoon kuuluva laitos, joka hallinnoi montaa erityyppistä korpusta sekä julkaisee korpuslingvistiikkaan liittyviä kustanteita ja kehittää ohjelmistoa (ÚČNK). SYNPUB2006-korpuksen kuuluu viisi sanomalehteä sekä useita aikakauslehtiä vuosilta 1989–2004. Korpus edustaa yhteensä yli 300 miljoonaan saneen laajuista aineistoa. Kun tarkemmin tutkii korpuksen sisältöä, huomaa, että enemmistö teksteistä ajoittuu ajanjakson jälkimmäiselle puolelle ja yli 40 prosenttia aineistosta muodostuu tšekin suurilevikkisimmästä päivälehdessä MF Dnes. Korpuksen tekijät korostavatkin, ettei korpuksen suurin etu ole sen edustavuudessa, vaan sen laajuudessa. Toisin kuin suomen korpus, tšekin SYNPUB2006 ei sisällä mitään alueellisia sanomalehtiä, mikä voisi tosiaan johtua niiden matalammasta suosiosta Tšekissä. Suomen aineisto siis vaikuttaa olevan monipuolisempi kuin tšekin korpus, mikä tšekin puolella toisaalta tasapainottuu suurella koolla. ČNK tarjoaa uudempiakin sanomalehtikorpuksia, mutta ottaen huomioon SKTP:n painopiste 90-luvun teksteissä, näen SYNPUB2006:n olevan sen sopivin vastine.

Taulukko 2. Verrannollisen osakorpusten koot ja hakusanojen osumien määrät.

korpus	korpuksen koko	YSTÄVÄ/PŘÍTEL	YSTÄVÄTÄR/PŘÍTELKYNĚ	Yhteensä
SKTP	144 milj. sanetta	18 988	600	19 588
SYNPUB2006	305 milj. sanetta	44 971	9 750	54 721

Taulukossa 2 tarkastellaan verrannollisen korpuksen molempien osien kokoja ja hakusanojen osumien määriä. Tšekin korpus on kooltaan kaksinkertainen suomen korpuksen verrattuna, mitä vastaavat myös sanojen YSTÄVÄ ja PŘÍTEL osumien määrät. Eri tilanne on kuitenkin, kun katsoo feminiinijohdoksia YSTÄVÄTÄR ja PŘÍTELKYNĚ. Nämä ovat odotuksenmukaisesti suomessa melko harvinaisia, edustaen pelkästään kolmea prosenttia kaikista osumista. Tšekin puolelta feminiinijohdos muodostaa melkein 18 prosenttia osumista, ja suomeen verrattuna on oleellinen osa aineistoa. Korpusten eri koot eivät analyysin kannalta ole ongelma, sillä tämä epäsymmetrisyys on tasapainotettu aineiston huomattavalla laajuudella. Kollokaatioanalyysi pohjautuu tilastollisiin testeihin, jotka ottavat kunkin osakorpuksen koon huomioon. Samalla oletetaan, että vaikka suomen aineisto on pienempi, se on kuitenkin tarpeeksi laaja, ettei tutkittavien lekseemien yleiset piirteet jäisi huomaamatta. Lekseemin YSTÄVÄTÄR osumien esiintymismäärä onkin melko matala, mutta feminiinijohdoksia ei tässä tutkielmassa käsitellä erikseen, vaan yhtenä joukkona sanojen YSTÄVÄ/PŘÍTEL kanssa.

3.2 Korpuslingvistiset menetelmät

3.2.1 Tilastolliset testit ja kollokaatioväli

Yhtenä kollokaation ominaisuutena on sen toistuvuus noodin, eli hakusanan läheisyydessä (ks. luku 2.2.2). Kvantitatiivisessa kollokaatioanalyysissä on siis keskeistä verrata kollokaatin esiintymismäärää noodin kontekstissa odotettavaan esiintymismäärään, eli siihen kuinka tiheästi kollokaatti esiintyisi tekstissä, mikäli kyseistä kollokaatiota ei olisi (Barnbrook 1996: 92). Tämä odotuksenmukainen esiintymismäärä perustuu hypoteettiseen korpuksen, joka sisältää samoja sanoja, mutta satunnaisessa järjestyksessä. Nollahypoteesi tässä tapauksessa olisi, ettei löytyisi mitään eroa oikean ja hypoteettisen korpuksen välillä noodin ja kollokaatin myötä-esiintymien laskiessa (Stubbs 1995: 7). Tällaisen hypoteettisen korpuksen korvaamiseen sovelletaan satunnaisjakaumaa, sillä jos noodi ei olisi vahvassa kollokaatiosuhteessa muiden

sanojen kanssa, muut sanat jakautuisivat satunnaisesti samassa suhteessa kuten muualla tekstissä. Sanojen odotuksenmukaista esiintymismäärää yleensä lasketaan joko koko tekstin, josta on poimittu konkordanssit, tai tarkemmin laajan korpuksen perusteella. Vertailemalla kollokaatin odotuksenmukaista frekvenssiä todelliseen frekvenssiin voidaan kertoa, kuinka todennäköistä on, että kollokaatti esiintyy noodin ympäristössä pelkästään sattumalta. Mikäli tämän sattuman todennäköisyys on tarpeeksi matala, voidaan myötäesiintymää pitää tilastollisesti merkitseväenä. (Barnbrook 1996: 92–94.)

Barnbrook (1996: 94–99) kuvailee kolme yleisintä tilastollista menetelmää, joilla voi laskea mainittua todennäköisyyttä ja määrittää kollokaattien tilastollista merkitsevyyttä. Nämä ovat z -testi, t -testi ja MI (*Mutual Information*) -testi, joista kukin kuvaa hiukan erilaisia asioita ja painottaa tietynlaisia kollokaatteja (mp., Jantunen 2004: 79). Z - ja t -testit muistuttavat toisistaan, koska molempien kaavat perustuvat kollokaatin todelliseen ja odotuksenmukaiseen frekvenssiin, joita vertaillaan keskihajonnan pohjalta. Eroa löytyy vain keskihajonnan laskelmasta, mikä kuitenkin vaikuttaa tuloksiin merkittävästi. Näistä t -testi antaa tarkempaa tietoa myötäesiintymistä, joiden yleisfrekvenssi on korpuksessa matala. (Barnbrook 1996: 97.) MI -testin I -arvoa lasketaan myös hyödyntämällä kollokaatin todellista ja odotuksenmukaista frekvenssiä, mutta ei keskihajonnan perusteella, jolloin MI -testi poikkeaa edellisistä. Sillä mitataan noodin ja kollokaatin välisen suhteen vahvuutta eli niiden todennäköisyyttä esiintyä tekstissä yhdessä. (Barnbrook 1996: 98; Clear 1993: 278.)

Jäsennellessään näiden kolmen testin tuloksia Barnbrook (1996: 100–101) huomaa, että z - ja MI -testien nostamat tilastollisesti merkitsevät kollokaatit vastaavat melko hyvin toisiaan, kun taas t -testin kollokaatit eroavat selvemmin. Tämä johtuu z - ja MI -testien taipumuksesta kasvattaa korpuksessa harvemmin esiintyvien sanojen merkitsevyyttä (mp.). Clear (1993: 280) muistuttaa, että välttääkseen idiosynkraattisia kollokaatteja nousemasta liian merkitseviksi, kannattaa asettaa kollokaattien vähimmäisesiintymistäajuus tarpeeksi korkealle. T -testin tapauksessa suuntaus on melko vastakkainen siinä mielessä, että t -arvon kollokaattilistojen kärjessä sijaitsee usein hyvin yleiset kieliopilliset sanat. Ilmiö johtuu t -arvon luonteesta kasvaa sitä suuremmaksi, mitä yleisempi noodin ja kollokaatin absoluuttinen yhteisesiintymä on. Etenkin laajan korpuksen tapauksessa kieliopilliset kollokaatit helposti kohoavat kaikista merkitsevimmiksi. (Stubbs 1995: 10–12.)

Yllä mainitusta tulee ilmi, että jokaisesta tilastollisesta testistä löytyy sekä etuja että haittapuolia, eikä mitään niistä voi pitää kaikenkattavana ja täydellisenä keinona mitata kollokaattien tilastollista merkitsevyyttä. Tiettyä testiä soveltaen on aina mietittävä tutkimuksen vaatimuksia ja tarpeita. On olemassa muitakin tilastollisia lähestymistapoja ja testejä, joilla

tutkimuksissa määritellään kollokaattien tilastollista merkitsevyyttä (ks. esim. Jantunen 2004; McEnery, Xiao & Tono 2006), mutta perehdyin tässä vain niihin menetelmiin, joita on edes mahdollista soveltaa ottaen huomioon käyttämäni käsittelyohjelmat.

Tutkielmassani käytän *MI*-testiä saadakseni sanojen YSTÄVÄ/PŘÍTEL tärkeimmät kollokaatit. Tämän päätöksen päätekijänä on se, että korpukseni on hyvin laaja ja niin esimerkiksi *t*-testin sadasta merkitsevimmästä kollokaatista kieliopilliset sanat edustavat vajaata puolta. Ongelma ei ole siinä, etteivät kieliopilliset kollokaatit olisi itsessään mielenkiintoisia. Tutkimukseni ydin on kuitenkin kollokaattien semanttisessa näkökulmassa, mikä johtuu osittain kahden erityyppisen kielen hankalasta kieliopillisesta vertailusta. Näin ollen *MI*-testi sopii tutkimustarpeisiini parhaiten, koska sen painopiste on leksikaalisissa- eli sisältösanoissa (ks. Stubbs 1995: 15). Karsiakseni harvinaisemmat kollokaatit tuloksista pois, säädän vähimmäisesiintymisfrekvenssin 30:ksi, mikä vastaa korpuksen laajuutta asianmukaisesti.

I-arvo on kollokaatin havaitun frekvenssin (*O*) suhde odotuksenmukaiseen (*E*) frekvenssiin binäärisessä logaritmissa. Frekvenssien laskiessa otetaan huomioon myös korpuksen saaneiden määrä (*N*), kuten näkyy *I*-arvon laskentakaavassa kuviossa 3:

$$I = \log_2 \frac{O}{E} \quad \leftrightarrow \quad I = \log_2 \frac{f(x, y) \cdot N}{f(x)f(y)}$$

Kuvio 3. Tilastollisen *MI*-testin *I*-arvon laskentakaava (Stubbs 1995: 15).

Mikäli havaittu frekvenssi on esimerkiksi kolme kertaa suurempi kuin odotuksenmukainen, saadaan *I*-arvoksi 1,58. Mitä suurempi *I*-arvo on, sitä vahvempi kollokaatiosuhde on kyseessä. (Barnbrook 1996: 98-99.) Tutkielmassani valitsen *MI*-testin minimiarvorajaksi $I = 3,0$ useamman muun tutkimuksen esimerkin mukaisesti (ks. Clear 1993; Jantunen 2004; Xiao & McEnery 2006).

Tilastollisten testien tärkeänä tekijänä on kollokaatioväli, eli noodin ja kollokaatin tekstuaalinen etäisyys, sillä tarkastelualueen laajuus vaikuttaa sekä kollokaattien kokonaismäärään, että niiden merkitsevyyteen (Jantunen 2004: 19; McEnery & Hardie 2012: 127–128). Kollokaatiovälin määrittäminen vaihtelee tutkimuksittain, ja korpustutkijat eivät ole yhtä mieltä tarkastelualueen sopivasta rajasta (Stubbs 1995: 8). Kollokaatiotutkimuksissa on käytetty vaihtelevia kollokaatiovälejä toisesta aina viidenteen sanan asemaan asti (McEnery & Hardie 2012: 129). Tutkimukset osoittavat, että mitä lähempänä noodia sana on, sitä todennäköisemmin on kyseessä merkittävä kollokaatti. Toisaalta viidennen sanan asemasta eteenpäin kollokaattien merkitsevyys laskee huomattavasti. (Sinclair ym. 2004: 42–44.) Sinclairin (1991) mukaan

merkitsevimmät kollokaatit esiintyvät noodista neljä sanaa molempiin suuntiin. Stubbs (2001: 29) on samalla kannalla kuin Sinclair siitä, että kollokaatioväli L4:R4 on kohtuullinen, jotta voi tunnistaa tärkeimmät kollokaatit, samalla korostaen optimaalisen havaintoalueen riippuvuutta tutkimustarpeista. On myös suositeltavaa pitää mielessä tutkittavan sanan kieliopilliset piirteet tarkastusalueen määrittämistä varten (Jantunen 2004: 83; McEnery & Hardie 2012: 129). Jantunen (2004) esimerkiksi käyttää astemääritteiden kollokaatiotutkimuksessaan epäsymmetristä väliä 3:2 huomioiden syntaktisia rakenteita, joissa tarkasteltavat lekseemit esiintyvät.

Omassa tutkielmassani kiinnitän huomiota lekseemien YSTÄVÄ/PRĪTEL kieliopillisiin seikkoihin vain toissijaisesti. Pääsyyinä siihen on tšekin ja suomen kielen toisistaan eroava kieliopillinen luonne, mikä ei ole tutkimukseni varsinaisena kiinnostuksen kohteena. Valitsemani lekseemit ilmenevät luonnollisesti erilaisissa syntaktisissa ympäristöissä, ja toisin tutkimukseen tarpeetonta epäsymmetrisyyttä, mikäli sen perusteella valitsisin kummallekin erilaisen kollokaatiovälin. Mainituista syistä nojaudun tarkastelualueen rajaamisessa aikaisempiin tutkimuksiin ja lasken kollokaattien havaintoalueeksi neljä sananväliä sekä vasemmalle että oikealle noodista nähden. Uskon, että geneerinen tarkasteluväli L4:R4 soveltuu hyvin kollokaatiotutkimukseen, jonka kohteena ovat leksikaaliset sanat, ja jossa on tarvetta kunnioittaa kunkin kielen eroavaa luonnetta.

3.2.2 Aineiston muodostaminen

Tässä ja seuraavassa luvussa kuvailen, miten muotoilin suomen ja tšekin korpusaineistoista kollokaattilistat, sekä millaisia pulmia tähän liittyi. Suomen kielen tekstikokoelma -korpuksen käsittelyyn olen ensin käyttänyt Kielipankin konkordanssihakuohjelmaa Korpia ja sitten laatinut konkordansseista kollokaattilistat ohjelmalla WordSmith Tools v 6.0. Aluksi olen hakenut Korpin avulla SKTP-korpuksesta kaikki lemموjen YSTÄVÄ ja YSTÄVÄTÄR konkordanssit CQL-lausekkeella: [lemma="ystävä"] | [lemma="ystävätär"]. Konkordansseja löytyi lausekkeella yhteensä 19 588. Korpista on otettu mukaan laajempi konteksti, joka pääosin koostuu useammasta lauseesta hakusanan oikealta sekä vasemmalta puolelta. Laajempi konteksti ei kuitenkaan ollut aina saatavilla, sillä joskus hakusana esiintyi artikkelin otsikossa, kappaleen alussa tai sen lopussa. Konkordanssit on tämän jälkeen tallennettu Exceliin ja koko muodostettu konkordanssikorpus on käyty läpi sekä automaattisesti että käsin duplikaattien poistamiseksi. Duplikaatilla tarkoitetaan kahta täsmälleen samaa konkordanssia, joita aineistossa esiintyy melko paljon, ja jotka voivat johtua esimerkiksi eri lehtien käyttämästä samasta

uutistoimiston uutisesta. Näin on poistettu yhteensä 998 konkordanssia, ja suomen aineisto koostuu lopulta 18 590 konkordanssista.

Aineisto sisältää sekä lemma- että sananmuotoversioita, joista kollokaatioanalyysissä otetaan käyttöön pelkästään lemmatsoitu konkordanssiaineisto, koska sananmuotoon perustuvissa kollokaattilistoissa monet yleiset kollokaatit toistuvat eri sananmuodoissa. Tutkielman näkökulmasta on kiinnostavampaa se, millaiset kollokaatit lekseemien YSTÄVÄ/PŘÍTEL kontekstissa esiintyvät, eikä niinkään se, missä sijoissa ne ilmenevät. Seuraavaksi WordSmith Tools -ohjelmassa kaikista konkordansseista, eli 577 173 sanasta, muodostettiin sanalista, jota tarvitaan tilastollisten testien laskemista varten.

Aineiston käsittelyä varten WordSmith Tools -ohjelmaan määritellään asetukset, jotka perustuvat edelliseen lukuun. Kollokaatioväliksi asetetaan L4:R4, eli neljä sanaa hakusanan kontekstista sekä oikealta että vasemmalta. Minimifrekvenssi viittaa siihen, kuinka monta kertaa kollokaatin pitää vähintään esiintyä kollokaatiovälissä. Koska aineisto on hyvin laaja, säädetään minimifrekvenssi 30:ksi ja *MI*-testin minimiarvo rajataan 3,0:ksi. Lopuksi asetetaan, että hakusanan kollokaatit voivat muodostua ainoastaan saman lauseen puitteissa. Konkordanssi laajempine konteksteineen koostuu tyypillisesti useammasta lauseesta, mutta yksittäisiäkin lauseita löytyy runsaasti. Ilman kyseistä asetusta voisi valitusta kollokaatiovälistä laskea kollokaateiksi myös edellisen tai seuraavan lauseen sanat. Tämä on ongelmallista, koska konkordanssikorpus ei muodosta yhtenäistä tekstiä, ja seuraava lause voisi siis olla aivan toisesta kohdasta tai jopa eri artikkelista.

Tšekin SYN2006PUB-korpuksen on sisäänrakennettu Kontext-sovellus, jolla pystyy muun muassa hakemaan konkordansseja sekä tuottamaan kollokaattilistoja. Vastaavasti tšekin konkordansseja on haettu CQL-lausekkeella [lemma="přítel"] | [lemma="přítelkyně"], jolla on saatu 54 721 osumaa. Kontextissa on mahdollista hankkia kollokaatit erilaisten muuttujien perusteella suoraan, joten ei tarvitse käsitellä konkordansseja toisessa ohjelmassa. Kollokaattilistojen muodostamista varten Kontextiin on asetettu samanlaiset rajoitteet kuin suomen aineistoon, eli L4:R4 -kollokaatioväli, minimiesiintymisfrekvenssi 30 ja *MI*-testin minimiarvo 3,0. Varmistin, että molempien ohjelmien *MI*-testin kaavat vastaavat täysin toisiaan.

3.2.3 Käsittelyohjelmiin liittyvät ongelmat

Kuten edellisestä luvusta kävi ilmi, suomen ja tšekin aineistoja on käsitelty hiukan eri tavalla, mikä johtuu suomen ja tšekin konkordanssiohjelmien erilaisista ominaisuuksista. Oleellinen ero syntyy siitä, että tšekin kollokaattien *I*-arvot on laskettu koko korpuksen perusteella, kun

taas suomen aineistona toimivat pelkästään yksittäiset konkordanssit laajempine konteksteineen. Tämän eroavuuden seurauksena *I*-arvot kollokaattilistojen välillä ovat epätasaiset siinä mielessä, että tšekin kollokaateilla on yleisesti ottaen korkeammat arvot kuin suomen kollokaateilla.

Aineistokäsittelyn alkuvaiheessa yritin ensin syöttää koko SKTP-korpuksen WordSmith Tools -ohjelmaan, jotta saisin tasapainoisempia tuloksia tšekin korpukseen verrattuna. Tämä sinänsä melko työläs prosessi onnistui kuitenkin vain osittain. Pääongelma johtui siitä, että korpuksen ladattava versio ei ole lemmatisoitu, joten kollokaattilistan voi tuottaa vain yksittäisistä sananmuodoista. Tšekin aineiston vaihtoehtoinen käsittely WordSmith Tools -ohjelmassa suomen aineiston tavalla taas ei onnistunut sen takia, ettei Kontextista pysty ladata kokonaisia konkordansseja konteksteineen.

I-arvon suuruuteen vaikuttavat neljä muuttujaa: korpuksen saneiden määrä, noodin frekvenssi, kollokaatin frekvenssi sekä noodin ja kollokaatin yhteisesiintymän frekvenssi (ks. kuvio 3). Mikäli käsittelyohjelmaan syötetään koko korpuksen sijaan vain konkordanssikorpus, silloin noodin frekvenssi ja yhteisesiintymän frekvenssi pysyvät samana, mutta sekä korpuksen koko, että yksinäisen kollokaatin frekvenssi laskevat. *MI*-laskentakaavan osoittajassa sijaitseva korpuksen koko laskee paljon enemmän kuin nimittäjässä oleva pelkän kollokaatin frekvenssi, mistä johtuu hieman alhaisempi *I*-arvo.

Kuvattuun ongelmaan löytyi lopulta kuitenkin melko suoraviivainen ratkaisu, sillä *MI*-testin laskentakaavan kaikki neljä muuttujaa ovat minulla tiedossa. Korpuksen saneiden määrän ja kollokaatin sekä noodin itsenäisen esiintymisfrekvenssin olen hakenut Korpin avulla koko korpuksen puolesta. Kollokaatin ja noodin yhteisesiintymän frekvenssi oli saatavilla WordSmith Tools -ohjelman edellisestä aineiston käsittelystä. Olen näillä muuttujilla laskenut *I*-arvon uudestaan jokaiselle suomen 360 kollokaatille, jotka ovat alun perin saaneet suuremman *I*-arvon kuin 3,0. Ainoa ongelma tässä lähestymistavassa on se, että jonkin tilastollisesti merkitsevän kollokaatin *I*-arvo olisi alun perin voinut olla pienempi kuin 3,0, joten sille ei olisi laskettu *I*-arvoa uudelleen, eikä se pääsisi lopulliselle kollokaattilistalle. Kummastakin kielestä käsittelen analyysissä kuitenkin vain ensimmäistä 250 kollokaattia (ks. liitteet 1 ja 2), joten en pidä todennäköisenä, että jokin merkittävä kollokaatti jäisi analyysin ulkopuolelle. Uskon, että olen lopullisella aineiston käsittelyllä saavuttanut tarkemman tiedon kollokaattien tilastollisesta merkitsevyydestä.

4 ANALYYSI

Analyysi-pääluku koostuu kaiken kaikkiaan kolmesta osasta. Kahta ensimmäistä osaa voi miettiä ikään kuin valmistelevana osa-analyysinä varsinaista kollokaatioanalyysiä varten. Kun vertaillaan kahden eri kielen käännösvastineita, voisi olla harhaanjohtavaa loikata suoraan niiden kollokaattien rinnastamiseen ilman, että ensin miettii kyseisten sanojen keskinäistä vastaavuutta ja niiden eri merkityksiä kummassakin kielessä. Näin ollen, ensin luvussa 4.1 käsittelen, kuinka usein lekseemien YSTÄVÄ ja PŘÍTEL käännökset vastaavat toisiaan rinnakkaiskorpuksessa, sekä millaista vaihtelua käännöksissä esiintyy. Seuraavassa luvussa 4.2 tarkastelun kohteeksi tulevat lekseemien YSTÄVÄ ja PŘÍTEL eri merkitykset, eli niiden polyseemisyys. Kiinnostusta herättää myös se, heijastaako korpuksen kielidata sanakirjojen merkitysjakoa (ks. luku 2.3). Viimeisenä on luvun 4.3 kollokaatioanalyysi, jossa tutkin ensin tarkemmin kaikista merkitsevimpiä kollokaatteja ja sitten analysoin tšekin ja suomen kielen 233 kollokaattia semanttisesta näkökulmasta.

4.1 Käännösten vastaavuus

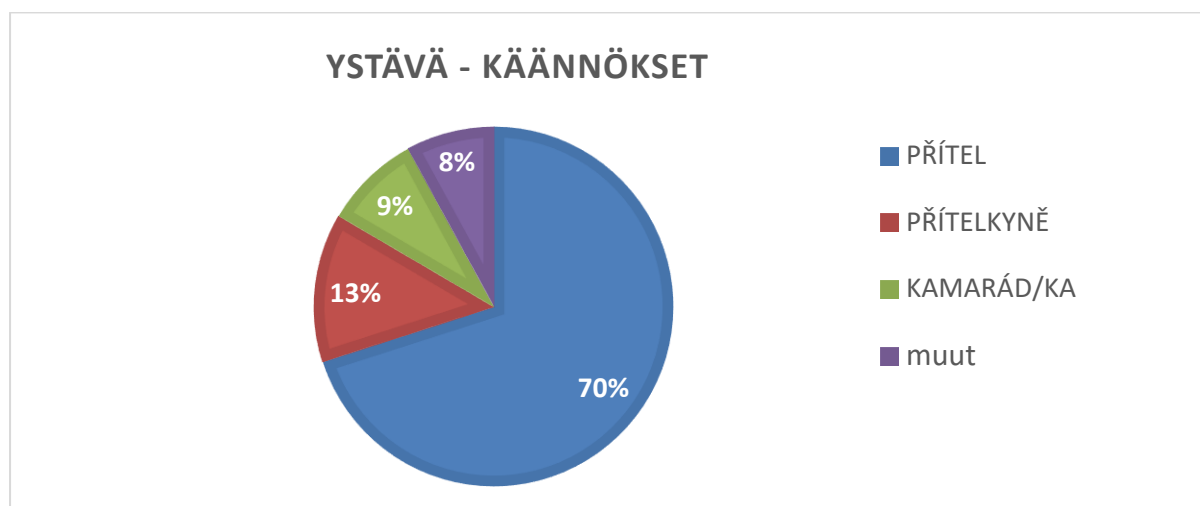
Teorialuvussa esitin, että käännösvastineet eivät oletettavasti ole koskaan täysin identtisiä, vaan ne useimmiten eroavat sävyissään, sivu- ja lisämerkityksissään, mikä heijastuu siihen, missä kontekstissa ja rekisterissä sanat ilmenevät. Tässä tutkielmassa vertaillaan kahden eri kielen lekseemejä, joten tulee pitää mielessä, että kielten rakenteelliset ja sanastolliset erotkin vaikuttavat analyysiin.

Tässä alaluvussa pyrin vastaamaan kysymykseen, löytyykö lekseemeille YSTÄVÄ ja PŘÍTEL tarpeeksi yhteistä pohjaa, että niiden kollokaatteja on mielekästä verrata keskenään. Toisin sanoen, on tarpeellista vahvistaa, että tutkittavat sanat ilmaisevat ja tarkoittavat suunnilleen samaa asiaa, paitsi sanakirjoissa myös laajoissa tekstiaineistoissa. Keskinäistä vastaavuutta voi tarkastaa laskemalla, kuinka usein toinen lekseemi on käännetty toiseksi (Altenberg 1999: 254; Johansson 2007: 26). YSTÄVÄ/PŘÍTEL -sanojen käännöksien vastaavuutta arvioin siis rinnakkais- eli käännöskorpuksen InterCorp 11 kautta (tarkemmin valitusta korpuksesta ks. luku 3.1.3.1). Otan ensin selvää, miten YSTÄVÄ on käännetty tšekiksi, ja sitten toisinpäin, miten PŘÍTEL on käännetty suomeksi. Käsittelen lyhyesti myös feminiinijohdoksia YSTÄVÄ-TÄR/PŘÍTELKYNĚ, joiden esiintymismäärä on korpuksessa kuitenkin varsin pieni. Alla olevasta taulukosta 3 näkyy kaikkien tarkasteltujen lekseemien esiintymismäärät.

Taulukko 3. InterCorp 11 -korpuksen suomen ja tšekin käännösten esiintymismäärät.

lemma	SUOMI → TŠEKKI	TŠEKKI → SUOMI
YSTÄVÄ/PŘÍTEL	531	216
YSTÄVÄTÄR/PŘÍTELKYNĚ	37	19

Jokaisesta lekseemistä hain korpukselta käännöksiä lemmana, eli kaikkine sananmuotoineen. Sanasta YSTÄVÄ löytyi eniten osumia, mikä johtuu siitä, että tšekiksi käännetty suomenkielinen aineisto oli rinnakkaiskorpuksessa lähes kaksikertainen verrattuna suomeksi käännettyyn tšekinkieliseen aineistoon. Ensimmäisenä tutkin lemman YSTÄVÄ käännöksiä tšekiksi (kuvio 4). Käännöksistä 70:tä prosenttia edustaa odotuksenmukaisesti sana PŘÍTEL. Jos nämä yhdistetään sanan PŘÍTELKYNĚ osuuden kanssa, saadaan 83 prosenttia koko näytteestä.



Kuvio 4: Lekseemin YSTÄVÄ (n = 531) tšekin käännökset InterCorpin 11 alakorpuksessa.

Vaikka tšekissä sana PŘÍTEL viittaa pääsääntöisesti mieshenkilöön (ks. luku 3.1.1), ei voi kuitenkaan päätellä, että sen jokaisessa esiintymisessä olisi kyse miespuolisesta henkilöstä. YSTÄVÄ-sanamuodoista, jotka olivat käännettyjä PŘÍTEL-sanamuodoiksi, 37 % tapauksista edustaa nimittäin monikkoa, jolla osoitetaan ryhmään. Ryhmä puolestaan voi koostua sekä miehistä että naisista:¹¹

- 4) *Onko Elinalla muita ystäviä tai sukulaisia, joiden luokse hän olisi voinut mennä?
Má Elina nějaké příbuzné nebo přátele, ke kterým by mohla jít?*

¹¹ Lemman PŘÍTELKYNĚ monikko toisaalta viittaa ainoastaan naispuolisista henkilöistä koostuvaan ryhmään.

Seuraava esimerkkilause (5) selventää, miksi käänöksien vastaavuutta määrittäessä kannattaa yhdistää sukupuolitetut ryhmät PŘÍTEL ja PŘÍTELKYNĚ:

5) Oi, *ystäväni pöllö, viisain kaikista linnuista.*

Ach, přítelkyně má, sovo, nejmoudřejší ze všech ptáků!

Ystävä pöllö on käännetty ilmaisulla *přítelkyně sova*, sillä sanan SOVA kieliopillinen sukupuoli sattuu olemaan feminiini. Jos mainittu ystävä olisi esimerkiksi PES (koira), joka on maskuliini, oikea määrite olisi PŘÍTEL. Näin voi tämänkaltaisissa tapauksissa lekseemien PŘÍTEL ja PŘÍTELKYNĚ välisen eron nähdä pikemminkin kieliopillisena kuin merkityksellisenä.

Huomionarvoinen on myös 9 prosentin ryhmä tapauksia, joissa YSTÄVÄ käännettiin sanalla KAMARÁD tai sen feminiinijohdoksella KAMARÁDKA. KAMARÁD on lekseemiä PŘÍTEL puhekielisempi ja osoittaa vähemmän vakavaa suhdetta, kuten myös suomenkielinen KAVERI. Sanat KAMARÁD ja KAVERI eivät kuitenkaan todennäköisesti ole samaa alkuperää.¹² Kahdeksassa prosentissa tapauksista käännettiin sana YSTÄVÄ jollakin toisella tavalla tai ilmaisulla. Yhteensä näitä on 42 kappaletta, ja ne edustavat melko kirjavaa valikoimaa erilaisia käänös-ratkaisuja. Tapauksista yleisimpiin kuuluu sanan korvaaminen jollakin toisella ekspressiivisellä ilmaisulla esim. *miláček, moje milá* (rakas, kultaseni), sanan kokonaan pois jättäminen ja muiden lekseemistä PŘÍTEL johdettujen sanojen hyödyntäminen, kuten PŘÁTELSKÝ, PŘÁTELSTVÍ (ystävällinen, ystävällisyys):

6) [...] *himoitsin yksinäisyydessäni nähdä ystävän kasvot ja kuulla ystävän ääneen.*

[...] *ve své osamělosti jsem si přál vidět přátelskou tvář a slyšet přátelský hlas.*

[kirjaimellisesti: ...*himoitsin yksinäisyydessäni nähdä ystävälliset kasvot ja kuulla ystävällisen äänen.*]

Käännösten analyysissä pitää muistaa, että käänösprosessissa yleisesti, etenkin kaunokirjallisuuden käänöksessä, tavoitteena ei ole kääntää tekstiä sanasta sanaan, vaan pyritään siirtämään samaa merkitystä ja vaikutusta lähdekielestä kohdekieleen. Sen takia sanoja ja lauseita saatetaan eri semanttisista ja pragmaattisista syistä kääntää moninaisilla tavoilla ja rakenteilla, jotka ovat lähempänä tai kauempana alkuperäistä ilmaisua. (Viaggio 2008: 174–175.)

Seuraavaksi tarkastelen lemman PŘÍTEL suomenkielisiä käänöksiä, joista YSTÄVÄ-lekseemin esiintymät muodostavat 84 % (kuvio 5). Vallitseva käänöstapa näyttää siis vielä suoriivaisemmalta kuin käänöksissä suomesta tšekiksi. Pienen viiden prosentin ryhmän

¹² *Kaveri* voi olla peräisin jiddiisistä tai yhdistelmä ruotsin lainasta *kamraatti* ja venäjän lainasta *toveri* (Itkonen 1992, s.v. *kaveri*). Tšekin sana *kamarád* on mahdollisesti tullut ranskan kielen sanasta *camarade*, tai saksankielisestä *kamerad* (Machek 1968, s.v. *kamarád*).

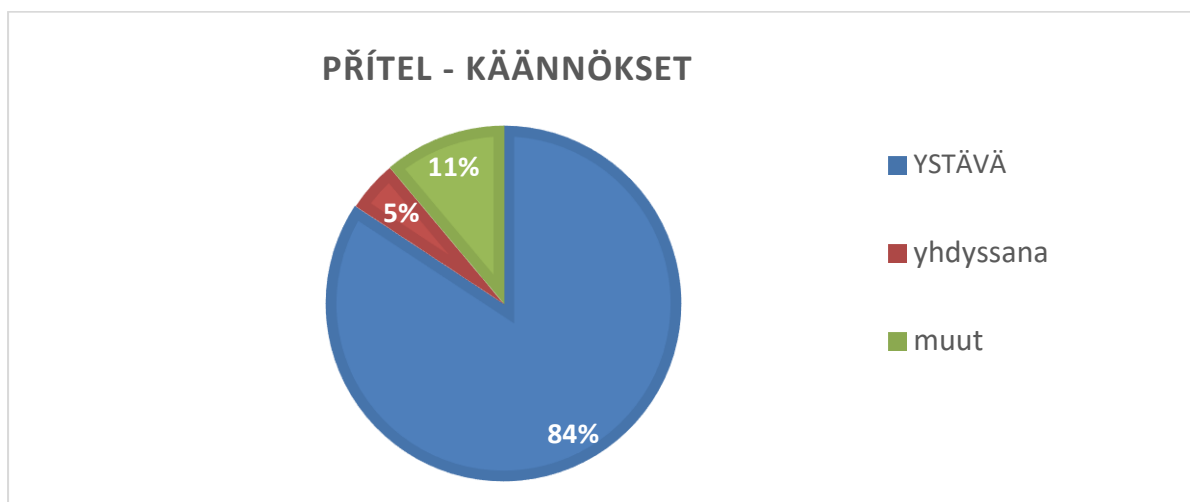
yhdysanoja voisi myös sisällyttää YSTÄVÄ-pääryhmään, sillä kaikissa näissä yhdysanoissa on sana YSTÄVÄ yhtenä yhdysosana kuten esimerkissä (7):

7) *Podle domluvy se zahraničnimi přáteli* [...]

[kirjaimellisesti: *ulkomaalaisten ystävien kanssa tehdyn sopimuksen mukaan*]

Ulkomaalaisystävien kanssa tehdyn sopimuksen mukaan [...]

Tšekin käännösten joukossa (kuvio 4) tämänkaltaisia käännöksiä ei esiinny ollenkaan, johtuen siitä, että tšekin kielessä yhdistäminen ei ole yhtä suosittu sananmuodostuskeino kuin suomessa. Loput PŘÍTEL-sanankäännöksistä erittelin 'muut'-ryhmäksi, sillä yksikään niistä ei nouse selvästi esille omaksi paradigmakseen. Sama kuin tšekin puolella, 'muut'-ryhmä koostuu erilaisista käännöstavoista, kuten sanan pois jättämisestä tai muiden merkitykseltään lähellä olevien sanojen avulla kääntämisestä (YSTÄVYKSET, TUTTU).



Kuvio 5: Lekseemin PŘÍTEL (n = 216) suomen käännökset InterCorpin 11 alakorpuksessa.

Vertaillen kuvioita 4 ja 5 on merkittävää, ettei käännöksissä tšekistä suomeen sanaa KAVERI löydy yhtäkään. Vaikka sanan PŘÍTEL käännöksiä onkin 2,5 kertaa vähemmän kuin YSTÄVÄ-käännöksiä, on silti yllättävää, ettei yhtäkään 216 tapauksesta ollut käännetty puhekielisemmäksi KAVERI-sanaksi, etenkin kun ottaa huomioon, että suomessa KAVERI on yksi YSTÄVÄ-lekseemin lähimmistä synonyymeistä. Tämä eroavuus herättää kysymyksen, miltä tilanne näyttäisi, kun tarkastelisi vastakkaisen suunnan käännöksiä. Käännetäänkö yhtä epätaisisesti lekseemien KAVERI/KAMARÁD vastineiksi PŘÍTEL/YSTÄVÄ? Hain samasta rinnakkais-korpuksesta lemmat KAVERI ja KAMARÁD, joita löytyi melkein sama kappalemäärä. Sitten tarkastelin, kuinka usein niitä on käännetty sanoiksi PŘÍTEL ja YSTÄVÄ. Kun lisätään sanojen YSTÄVÄ/PŘÍTEL käännökset, tilanne näyttää seuraavalta:

Taulukko 4. Lekseemien YSTÄVÄ/PŘÍTEL ja KAVERI/KAMARÁD keskinäisten käännösten määrät InterCorpin 11 alakorpuksessa.

lemma	esiintymismäärä	käännössuunta	prosenttiosuus
YSTÄVÄ	531	YSTÄVÄ → KAMARÁD	9 %
KAMARÁD	208	KAMARÁD → YSTÄVÄ	12 %
KAVERI	188	KAVERI → PŘÍTEL	5 %
PŘÍTEL	216	PŘÍTEL → KAVERI	0 %

Taulukkoa 4 tutkiessa käännöstävät vaikuttavat siltä, että suomen YSTÄVÄ ja tšekin KAMARÁD ovat toisiaan lähempänä kuin tšekin PŘÍTEL ja suomen KAVERI. Käännössuunnat YSTÄVÄ → KAMARÁD → YSTÄVÄ toteutuivat 9 ja 12 prosentissa tapauksista. Toisaalta KAVERI oli käännetty sanaksi PŘÍTEL harvemmin, viidessä prosentissa tapauksista, ja kuten edellä jo mainittiin, PŘÍTEL ei aineistossani kääntynyt sanaksi KAVERI kertaakaan. Taipumus siis osoittaa, että YSTÄVÄ olisi laajemmin käytettävä kuin sen tšekin vastine PŘÍTEL, ja että joissakin tilanteissa, joissa suomen kielessä käytetään sanaa YSTÄVÄ, tšekki suosisi sen ekvivalentiksi mieluummin sanan KAMARÁD kuin PŘÍTEL. Hypoteesina voisi miettiä, että tšekin PŘÍTEL sopii muodollisempaan käyttöön ja YSTÄVÄ-sanaa voi siihen verrattuna käyttää puhekielisemmissä tilanteissa. Tätä näkökulmaa tukee myöskin se, että ottaen huomioon suomen ja tšekin aineistojen eroavat koot, esiintyy PŘÍTEL aineistossa huomattavasti vähemmän kuin YSTÄVÄ, ja toisaalta KAMARÁD taas paljon useammin kuin KAVERI.

On syytä pohtia vähintään muutamaa seikkaa, jotka voivat vääristää rinnakkaiskorpuksen analyysiä. Ensiksi tarkastelemani aineisto muodostuu yksinomaan kaunokirjallisista teksteistä, mikä itse ei välitä tarpeeksi monipuolista näkemystä kielen yleisestä käytöstä. Toiseksi aineisto koostuu vain rajatusta määrästä kääntäjiä ja kirjailijoita, joista jotkut ovat yliedustettuja (tarkemmin ks. luku 3.1.3.1), joten käyttämäni alakorpus ei ole riittävän edustava. Käännöksiä läpikäydessä huomaa, että joissakin aineiston kirjoissa PŘÍTEL/YSTÄVÄ systemaattisesti käännetään vain tietyllä tavalla. Näin esimerkiksi Waltarin ja Linnan teoksissa YSTÄVÄ-sanasta tulee aina PŘÍTEL/KYNĚ, mutta Oksasen romaanissa YSTÄVÄ on taas useammin KAMARÁD/KA. Ilmiö voi johtua monestakin syystä, kuten kirjan julkaisuajasta, genrestä tai kääntäjän omasta valinnasta. Sen lisäksi eri lekseemien vaihteleva esiintymismääräkin voisi vaikuttaa tulosten tulkintaan. Mainituista syistä ei kannata tehdä yleistäviä päätelmiä vain tämän lyhyehkön käännösanalyysin perusteella. Lekseemien tarkempi ja kattavampi analysointi voisi tuoda

mielenkiintoisia havaintoja, mutta se kuitenkin jää tämän tutkielman rajallisen laajuuden takia käsittelemättä.

Perinpohjaisuuden vuoksi tarkastelin, miten aineistossa oli käännetty feminiinijohdokset YSTÄVÄTÄR ja PŘÍTELKYNĚ. Näiden lukumäärät ovat alakorpuksessani varsin pienet (37 ja 19 kappaletta), joten mitään syvempää analyysia näistä ei kannata tehdä. Voi kuitenkin todeta, että käännökset heijastavat melko hyvin sanojen YSTÄVÄ ja PŘÍTEL tilannetta. Toisin sanoen YSTÄVÄTÄR on tšekiksi useimmiten PŘÍTELKYNĚ, ja sen lisäksi esiintyy monesti myös käännös KAMARÁDKA. Toisinpäin sanan PŘÍTELKYNĚ vastakohta oli aineistossa useimmiten YSTÄVÄTÄR, jonka lisäksi löytyy muutama eri käännös.

Tämän käännöskorpuksen analyysin perimmäisenä ajatuksena oli tarkastella tutkimuksen havaintoyksikköjen keskinäistä vastaavuutta (*mutual correspondence*). Keskinäisen vastaavuuden laskemiseksi käytetään seuraavaa kaavaa,

$$\frac{A_t + B_t}{A_s + B_s} \times 100$$

jossa A_t ja B_t ovat vertailtavien alkioiden frekvenssit käännöksissä ja A_s ja B_s niiden frekvenssit alkuperäisissä teksteissä (Altenberg 1999). Jos siis yhden kielen sana A on käännöksessä aina käännetty sanaksi B ja sana B on vastaavasti aina käännetty A:ksi, niiden keskinäinen vastaavuus olisi 100 %. Toisinpäin, jos kahta sanaa ei ole aineistossa koskaan käännetty toistensa kautta, niiden vastaavuus olisi 0 %. (Altenberg & Granger 2002: 17–18.) Tätä kaavaa soveltamalla saadaan sanaparin YSTÄVÄ/PŘÍTEL keskinäiseksi vastaavuudeksi 74 % ja jos lasketaan mukaan myös sanan YSTÄVÄ käännökset PŘÍTELKYNĚ-sanaksi sekä suomen yhdyssanat, saavutetaan jopa **85 %** vastaavuus. Tämä on suhteellisen korkea, joten voi sanoa, että YSTÄVÄ ja PŘÍTEL osoittautuvat ylivoimaisesti suosituimmiksi toistensa käännösvastineiksi.

Lopulta olisi kiinnostavaa nähdä, miten tilanne heijastuu suomi–tšekki ja tšekki–suomi sanakirjoihin (taulukko 5). Sanakirjoista vanhempi vuoden 1984 taskusanakirja kääntää leksemit hyvin yksiselitteisesti, eli niin, että YSTÄVÄ käännetään sanaksi PŘÍTEL ja PŘÍTEL taas YSTÄVÄKSI.

Taulukko 5. Kaksikielisten sanakirjojen YSTÄVÄ/TÄR ja PŘÍTEL/KYNĚ sana-artikkelit.

	Suomi–tšekki–suomi taskusana- kirja	Finsko-český, česko-finský šikovný slovník
YSTÄVÄ	přítel, přítelkyně	1. kamarád, přítel, přítelkyně 2. milovník, vyznavač ¹³
YSTÄVÄTÄR	-	-
PŘÍTEL	ystävä	1. ystävä 2. poikaystävä
PŘÍTELKYNĚ	ystävä(tär)	1. ystävä 2. tyttöystävä

Hieman erilaiset käännösvaihtoehdot on kuitenkin kuvattu uudemmassa vuoden 2016 sanakirjassa. Ensinnäkin on huomionarvoista, että sanan YSTÄVÄ ensimmäisenä käännösvastineena esitellään puhekielisempi KAMARÁD ja vasta sen jälkeen PŘÍTEL ja PŘÍTELKYNĚ. Lekseemin PŘÍTEL ensimmäisen merkityksen suomennos on pelkästään YSTÄVÄ, mikä on samassa linjassa myös käännöskorpuksen tuloksien kanssa siinä mielessä, ettei PŘÍTEL-sanaa käännetä KAVERIKSI juuri lainkaan. Toinen merkittävä huomio on se, että uudempi sanakirja tarjoaa toisenkin sanojen PŘÍTEL/KYNĚ käännösvaihtoehdon, joka on POIKA- tai TYTTÖYSTÄVÄ. Sanakirjojen mukaisesti sanoille PŘÍTEL ja PŘÍTELKYNĚ on siis tapahtunut semanttinen muutos, jossa niiden merkitysalue on laajentunut, tarkoittaen läheisen luotettavan ihmisen lisäksi myös puolisoa. Tämä muutos ei kuitenkaan näy käännöskorpuksessa, mikä mahdollisesti johtuu käännöskorpuksen aineiston rajallisesta laajuudesta sekä ajankohdasta. Sanaparin PŘÍTEL/YSTÄVÄ erilaisia merkityksiä analysoin tarkemmin seuraavassa luvussa.

4.2 Merkitysjako ja polyseemisyys

Tämän alaluvun ytimessä on lekseemien YSTÄVÄ ja PŘÍTEL eri merkitykset sekä se, kuinka vastaava niiden merkitysjako on suomen ja tšekin kielen välillä. Ensin esittelen, miten tutkittavia lekseemejä on määritelty valituissa sanakirjoissa (sanakirjoista tarkemmin luvussa 3.1.2). Sen jälkeen käsittelen korpusaineistoa ja tutkin, miten autenttinen kielenkäyttö vastaa tai eroaa sanakirjojen merkitysjaoittelusta. Tarkoituksena on ymmärtää paremmin lekseemien polyseemisyttä ja saada käsitys siitä, kuinka yleisiä yksittäiset alamerkitykset ovat. Tämä tieto toimii jälkeinpäin pohjana kollokaatioanalyysille.

¹³ harrastaja, kannattaja

4.2.1 Merkitysjako sanakirjoissa

Suomen kielen puolesta tarkastelen merkitystä Kielitoimiston sanakirjasta (KS) ja Nykysuomen sanakirjasta (NS). Uudempi KS luettelee lekseemille YSTÄVÄ kaksi päämerkitystä. Ensimmäinen (1.) on määritelty seuraavasti: 'ihmisestä suhteessa toiseen, jota kohtaan hän tuntee mieltymystä ja luottamusta ja joka suhtautuu häneen samoin'. Toisessa (2.) merkityksessään YSTÄVÄ on 'jonkin harrastaja, arvostaja, kannattaja, suosija, suojelija, auttaja tai tuki' (*Hyvän ruoan ystävä*)¹⁴. Ensimmäiselle merkitykselle vielä laajemmin selitetään, millaisella neljällä käyttötavalla sana tämän lisäksi esiintyy: **a.** YSTÄVIKSI voi merkitä samaan ryhmään kuuluvat samanmieliset ihmiset (*Ystävät ja viholliset.*); **b.** sanaa voi käyttää varoittavassa sävyssä toiselle puhuessa (*Älä hyvä ystävä luota siihen lieroon!*); **c.** YSTÄVÄLLÄ voi myös viitata seurustelukumppaniin (*poika-, tyttöystävä*) ja **d.** YSTÄVÄ voi olla muukin kuin ihminen, etenkin kuvaannollisissa ilmaisuissa (*Vanha ystävämmme kevät on taas tullut.*).

Nykysuomen sanakirjan yleiskatsaukseen verrattuna suurin ero näkyy siinä, että NS luettelee sanalle YSTÄVÄ kolme päämerkitysryhmää, joista ensimmäinen ja kolmas suurin piirtein vastaavat KS:n kahta päämerkitystä. NS:n toinen merkitys puolestaan koostuu useanlaisista käyttötavoista, jotka löytyvät jo KS:n ykkösmerkityksen alta, kuten YSTÄVÄ varoittavana puhuttelusanana, aatteellisen yhteenkuulumisen osoitus ja rakastetun nimeäminen (*sydämen ystävä*). Suurempia merkityseroja siis näiden kahden sanakirjan välillä ei löydy, vaan molemmissa luetellaan samat merkitykset ja käyttötavat. Eroja on kuitenkin sana-artikkelin kokonaisrakenteessa ja esittelytavassa. Tällaisesta merkitysten erillisestä painottamisesta voisi päätellä, että kyseiset merkitykset, jotka NS erottaa itsenäiseksi merkitysryhmäksi, ovat nykyisin marginaalisempia kuin NS:n julkaisemisen aikana.

Lekseemin YSTÄVÄTÄR osalta molempien sanakirjojen kuvaus on paljon suppeampi. KS:n mukaan YSTÄVÄTÄR on 'naisen naispuolinen ystävä' tai toisessa merkityksessä 'rakastajatar'. NS samalla tavalla mainitsee YSTÄVÄTTÄREN olevan toisen naisen ystävä sekä senkin mukaan sanaa käytetään kiertoilmauksena rakastajattarelle. Jo sanakirjojen selvitysten perusteella on selkeää, että sanalla YSTÄVÄTÄR on tiukemmin rajatut käyttötavat, eikä sitä käytetä samanlaisessa yleisessä merkityksessä kuin YSTÄVÄ-sanaa. Sisällytän lekseemin YSTÄVÄTÄR kuitenkin analyysiin, jotta tšekki ja suomen aineiston välinen tasapaino ei rikkoontuisi.

Sanan PŘÍTEL merkitysten selventämiseksi käytän sanakirjoja Slovník spisovného jazyka českého (SSJČ) ja Slovník současné češtiny (SSoučČ). Näistä ensimmäinen määrittelee

¹⁴ Esimerkkilauseet Kielitoimiston sanakirja s.v. *ystävä*.

hakusanalle PŘÍTEL neljä päämerkitystä seuraavasti: PŘÍTEL on (1.) ’joku, joka tulee jonkun toisen kanssa hyvin toimeen, jota kohtaan hän tuntee myötätuntoa ja he ovat suhteessa toisiinsa läheisiä’; tai (2.) ’kannattaja, suosija’; (3.) puhekielessä voi olla kyseessä tuttavallinen puhuttelusana, joka voi olla alentava; ja (4.) monikossa *přátele* on vanhahtava ja murteellinen ilmaisu tarkoittaen sukulaisia.

SSoučĚ-sanakirjan vuoden 2013 sähköinen versio esittelee lekseemin PŘÍTEL SSJČ:n kolmen ensimmäisen merkityksen lisäksi myös joksikuksi, ’joka on toisen kanssa läheisessä, seksuaalisessa suhteessa’ esimerkikilauseella: *Má přítele, není volná*. (Hänellä on poikaystävä, hän ei ole vapaa). Tämän merkitysjärjestyksessä toissijaisen määrittelyn sekä oman kielitajuni mukaan sanan PŘÍTEL uuden merkityksen parempi suomen vastine olisi POIKAYSTÄVÄ, AVO-MIES tai PUOLISO. Tšekin sanan PŘÍTEL tapauksessa on siis kyse semanttisesta muutoksesta ja merkityksen laajenemisesta, jota on aiemmin tutkittu tietyillä yhteisöä ja ihmisten välistä suhdetta ilmaisevilla sanoilla (ks. Wierzbicka 1997; Yavorska & Zymovets 2016). Lekseemin PŘÍTEL merkityssiirtymä vaikuttaa olevan suhteellisen uusi, sillä sitä ei ole kuvailtu viime vuosisadan lopun sanakirjoissa. Tarkastelemani tšekin korpusaineisto on peräisin 1990-luvulta ja 2000-luvun alusta, joten jää nähtäväksi heijastuuko kyseinen muutos sanaparin PŘÍTEL/KYNĚ kollokaatteihin.

Lekseemiä PŘÍTELKYNĚ käsittelevät tšekin sanakirjat vielä suppeammin kuin suomen sanakirjat lekseemiä YSTÄVÄTÄR. SSJČ:n mukaan PŘÍTELKYNĚ on sanan PŘÍTEL feminiinijohdos ja sitä voi käyttää sanan PŘÍTEL ensimmäisessä ja toisessa merkityksessä. SSoučĚ tarjoaa pelkästään sanan kieliopillista tietoa. PŘÍTELKYNĚ on siis suomen YSTÄVÄTTÄREN tavoin rajautuneempi käyttötavoissaan. Ei kuitenkaan vaikuta sanakirjojen mukaan siltä, että sen ystävyyssuhteessa oleva toinen henkilö olisi välttämättä myös naispuolinen.

Yhteenvetona voi sanoa, että sanakirjojen tarkastelun perusteella lekseemit YSTÄVÄ ja PŘÍTEL vastaavat toisiaan useammalla merkitystasolla. Molempia voi pitää hyvin samankaltaisina niiden perusmerkityksessä, eli jonakuna, joka on johonkin toiseen läheisessä suhteessa oleva, jota kohtaan koetaan positiivisia tunteita sekä luottamusta. Molemmissa kielissä sanalla myös ilmaistaan jonkin kannattajaa, oli se sitten aatteen, harrastuksen tai jonkin muun asian. Sen lisäksi kummankin sanan käyttöalueisiin kuuluu toisen puhuttelemine, joka suomessa saa varoittavan ja tšekissä alentavan sävyn. Suurin merkitysero lekseemien välillä lienee tšekin taipumuksessa käyttää sanaa PŘÍTEL poika- tai miesystävän tavoin. Vaikka Kielitoimiston sanakirja mainitsee, että YSTÄVÄ-lekseemillä viitataan myös seurustelukumppaniin, sen esimerkikilauseesta (*Poika-, tyttöystävä. Etsiä ystävää kirjeenvaihtoilmoituksella.*) tämä ei selvästi nouse esille. Sanojen YSTÄVÄ/PŘÍTEL välillä on siis sanakirja-analyysin perusteella

päällekkäinen polysemia, mikä tarkoittaa, että lekseemit joissakin merkityksissä vastaavat toisiaan ja joissakin ne eroavat (ks. François 2008).

4.2.2 Merkitysjako verrannollisessa korpuksessa

Tässä kappaleessa käsittelen, vastaako verrannollisen korpuksen aineisto (ks. lukua 3.1.3.2) edellisen luvun merkitysjakoa sekä sitä, kuinka yleinen kunkin sanan alamerkitys on. Tarkasteltavaksi poimin satunnaisella otannalla kummankin kielen osakorpukselta 500 konkordanssia, joiden nooissa sijaitsevat sanat PŘÍTEL/YSTÄVÄ tai niiden feminiinijohdokset YSTÄVÄTÄR/PŘÍTELKYNĚ. Suomen YSTÄVÄTÄR-lekseemin suhteelliseen harvinaisuuteen ja toisaalta tšekin PŘÍTELKYNĚ-lekseemin yleisyyteen vastaavat myös niiden lukumäärät otetuissa konkordansseissa. Yhdeksän YSTÄVÄTÄR-konkordanssia edustaa vain 1,8 % otoksesta, kun taas tšekissä PŘÍTELKYNĚ ilmeni 86 konkordanssissa, eli 17,2 % tapauksista näytteessä.

Kaikki konkordanssit, ja tarvittaessa myös niiden laajempi konteksti, on näytteenoton jälkeen analysoitu ja jaoteltu tutkittavien lekseemien merkitysten mukaisesti. Merkitysryhmiä muotoillessa otin mallin sanakirjojen alamerkitysten järjestyksestä ja hiukan muokkasin niitä aineiston tarkastelun jälkeen. Yhteensä olen muodostanut viisi ryhmää, joista toisesta ryhmästä löytyy vielä kolme alaluokkaa seuraavasti:

1. läheinen ihminen
2. a. harrastaja yhdistyksen nimessä
b. harrastaja, suosija
c. liittolainen, kannattaja
3. puoliso
4. puhuttelusana
5. kuvaannollinen ilmaus

Tässä kohdassa ensin esittelen ryhmittelyä tarkemmin sekä selitän, minkälaisia konkordansseja olen jaotellut kuhunkin ryhmään ja miksi. Jälkeenpäin arvioin tuloksia ja vertailen eri merkityksiä suomen ja tšekin välillä.

Ensimmäinen ryhmä (1) edustaa sanojen peruserkitystä ja oletettavasti myös kaikista yleisintä merkitystä. Kaikkien tarkasteltujen sanakirjojen välillä on konsensus, sanojen YSTÄVÄ/PŘÍTEL ollen ensisijaisesti läheinen ihminen, johon voi luottaa ja jonka kanssa tulee hyvin toimeen, kuten esimerkeissä (8) ja (9).

8) *Mohamed al-Fayed byl přítelem mého otce.* [Mohamed al-Fayed oli isäni ystävä.]

9) *Petite Morte on tarina kahdesta ystävättärestä, saman miehen rakastajattaresta ja vaimosta.*

Tapaukset, joissa YSTÄVÄ/PŘÍTEL ilmenevät elokuvien, kirjojen tai näytelmien nimissä osoittautuivat hankaliksi eritellä. Etenkin tšekin otoksessa oli ilman kontekstia ongelmallista päätellä, onko kysymyksessä PŘÍTEL perusmerkityksessään vai tarkoitetaanko sillä poika/tyttöystävää. Suomeksi tämäntyyppisiä epäselvyyksiä oli paljon vähemmän, mutta esimerkiksi Pentti Holapan kirjan nimi *Ystävän muotokuva* aiheutti päänvaivaa. Tällöin päätin luokitella sen 'puoliso'-ryhmään, koska kirja käsittelee kahden ihmisen romanttista suhdetta.

Lisäksi olen lajitellut perusmerkitysryhmään myös koiraan ja muihin lemmikkeihin viitattavia ilmaisuja (10), joita löytyy jonkin verran.

10) *Lemmikki todellakin on paras ystävä.*

Vaikka kuvaannollisia ilmauksia varten onkin varattu jaottelun viimeinen ryhmä, ihmisen suhde lemmikkiin on tietyllä tasolla verrattavissa ihmisystävien suhteeseen ja näin sanan käyttötapa eroaa pelkästä vertauskuvallisesta merkityksestä. Perusmerkitysryhmään olen lisännyt myös pienen viidestä konkordanssista koostuvan joukon, jossa YSTÄVÄ on vapaaehtoisen palvelun, esim. vanhainkodissa työskentelevän, nimike (11).

11) *Elinkaaren toisessa päässä moni on jälleen yksin, ja ystäväpalvelujen ystävät saattavat olla lähes ainoat kontaktit ulkomaailmaan.*

Kyseessä on Suomen Punaisen Ristin niin sanottu Ystävätoiminta, jonka tehtävänä on avustaa asiakasta ja toimia tämän ystävänä. Avustettavan ja vapaaehtoisen välinen suhde on verrattavissa perinteiseen ystäväsuhteeseen, joten näidenkin konkordanssien paikka on perusmerkityksen ryhmässä.

Toisen (2.) ryhmän konkordanssien merkitys pohjautuu sanakirjojen määritelmiin, joiden mukaan sanat YSTÄVÄ/PŘÍTEL voivat esittää jonkin harrastajaa, kannattajaa, suosijaa ja arvostajaa. Konkordansseja tutkiessa koin tarpeelliseksi erotella tämän merkityksen vielä kolmeksi alaryhmäksi, jotka tosin ovat lähellä toisiaan mutta samalla tietyssä mielessä eroavat. Ensimmäkin huomaa, että YSTÄVÄ/PŘÍTEL harrastajan tai kannattajan merkityksessä ilmenee joidenkin järjestöjen ja kerhojen virallisissa nimissä kuten *Raittiuden Ystävät*, *Käsityön Ystävät ry.* tai *Klub přátel Českých Budějovic* (České Budějovicen¹⁵ ystävien kerho). Niitä (2.a) kannattaa käsitellä erikseen, jotta paremmin nousee esiin, millainen 'harrastaja'-merkityksen edustus

¹⁵ Tšekkiläinen kaupunki

tšekissä ja suomessa on silloin, kun ilmaus on jonkin järjestön nimen osana tai kun se on yleisen kielenkäytön ilmentämä. Jälkimmäistä tapausta jaottelen edelleen kahdeksi ryhmäksi, joiden välillä hahmottuu hieno merkitysvivahde. Osassa konkordansseista (2.b) tulee selvästi esiin nimenomaan harrastaminen tai jonkin asian suosiminen. Täten joku (yleensä ihminen, muttei välttämättä) pitää jostakin tai harrastaa jotakin kuten esimerkkilauseissa (12) ja (13).

12) *Hän on myös teatterin suuri ystävä.*

13) *Ponrepo nikdy nebyl přítelem bombastické reklamy.* [Ponrepo¹⁶ ei koskaan ollut mahdottoman mainoksen ystävä.]

Toiset lauseet (2.c) herättävät hiukan erilaisia konnotaatioita. Näissä YSTÄVÄ/PŘÍTEL - sanat ilmaisevat pikemminkin liittolaisuutta, useimmiten poliittisessa kontekstissa. Konkordansseissa (14) ja (15) ei ole kyse harrastamisesta, vaan kannattamisesta, ja jo mainittu liittolaisuus nousee vahvasti esiin.

14) *Ruotsissa Lipposesta on kuva, joka on yhdistelmä harvasananaisen jäyhää ja leppoisaa Ruotsin ystävää.*

15) *Miksi antaa Koiviston ajaa vain ystäviensä etua, kun koko valtakunnan asioissakin riittäisi fundeeraamista?*

On kuitenkin syytä mainita, että konkordanssien ryhmittely merkityksen mukaan ja yli-päätään merkityksen hahmottaminen sekä ymmärtäminen on aina jossain määrin subjektiivista. Sen lisäksi monessakin tapauksessa lauseet ovat useamman merkityksen rajoissa. Pyrin kuitenkin mahdollisimman yhtenäiseen ja järjestelmälliseen luokitteluun.

Kolmas pääryhmä (3.) kuvaa sanojen YSTÄVÄ/PŘÍTEL merkitystä, jolla viitataan puolisoon, eli tilanteeseen, jossa henkilöiden välillä on jonkinlainen romanttinen suhde. Vain uusin tšekin sanakirjoista huomioi tämän alamerkityksen, ja suomen sanakirjat mainitsevat YSTÄVÄ-sanan tarkoittavan puolisoa vain rajallisesti. Useimmiten joko sanan kontekstista tai konkordanssin laajemmasta kontekstista selviää, onko kyseessä puoliso vai ”pelkkä” ystävä kuten esimerkkilauseessa (16). Joskus merkityksen tarkka tunnistaminen on kuitenkin hyvin hankalaa, koska laajempaa kontekstia ei ole aina saatavilla, ja ilman sitä voi olla mahdotonta tietää, millainen on kyseisten henkilöiden suhde. Esimerkiksi lauseen (17) YSTÄVÄTÄR-sanan tulkitsin tarkoittavan naisystävää, vaikka näin ei välttämättä ole. Tämänäköisiä epäselviä tapauksia

¹⁶ Prahalainen elokuvateatteri

on ollut kuitenkin suhteellisen vähän, joten mahdollisten epätarkkuuksien ei pitäisi vaikuttaa kokonaistuloksiin.

16) *Krátce před Vánoce mi se rozešla se svým přítelem.* [Hän erosi (poika)ystävänsä kanssa vähän ennen joulua.]

17) *Toukokuun 9. ostin ystävättärelle kukkakimpun.*

Neljänteen ryhmään (4.) kuuluvissa lauseissa YSTÄVÄ/PŘÍTEL -lekseemiä käytetään puhuttelusanoina (18). Suomen sanakirjojen mukaan ilmaisussa ilmenee silloin varoittava sävy, kun taas tšekin sanakirjassa huomautetaan, että tällainen puhuttelu voi olla sävyiltään alentava.

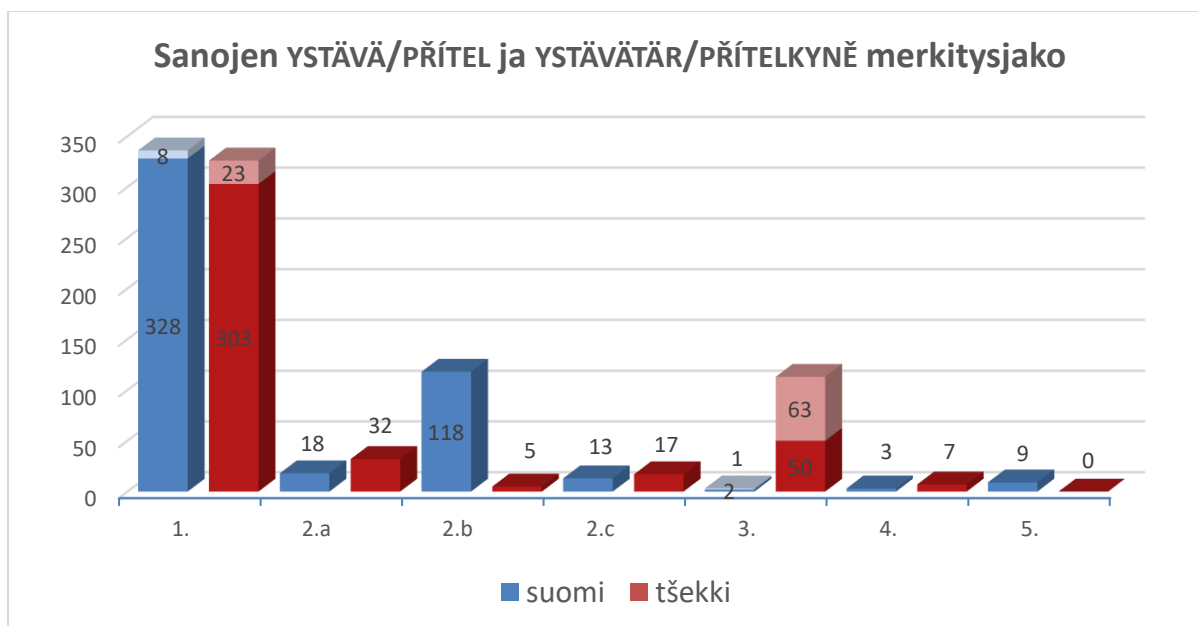
18) *Vážení přátelé, [...].* [Arvoisat ystävät, [...].]

Lopulta viidenneksi ryhmäksi (5.) luokittelin ne konkordanssit, joissa sanat YSTÄVÄ/PŘÍTEL esiintyvät kuvaannollisina ilmaisuina. Näissä tapauksissa siis ei ole kyseessä ihminen, vaan ystävän oletettavat ominaisuudet metaforisesti siirretään toisenlaiseen kohteeseen kuten esimerkissä (19). Sanojen YSTÄVÄ/PŘÍTEL mahdollisesta metaforisesta käytöstä mainitsevat pelkästään suomen sanakirjat.

19) *Oireet ovat ihmisen ystäviä, jotka yrittävät suojella kehoa [...].*

Wierzbicka (1996: 244–245) korostaa, että sanan metaforista merkitystä ei kannata eritellä itsenäiseksi merkitykseksi, koska metaforisessa käytössä ei ole kyse erilaisesta leksikaalisesta merkityksestä, vaan sen käyttöä voi selittää alkuperäisen merkityksen kautta. Tämä huomio pätee jaotteluni viimeiseen kahteen ryhmään. En siis pidä niitä täysin erillisinä merkityksinä, eikä niitä ole kuvailtu näin myöskään sanakirjoissa. Erottelon kuitenkin tässä analyysissä myös lekseemien YSTÄVÄ/PŘÍTEL erilaisia käyttötapoja, jotta niiden merkitys- ja käyttöprofiileja voi helpommin selvittää sekä vertailla kielten välillä.

Alla olevassa kuviossa 6 näkyy merkitysjaon lukumääräiset tulokset ryhmittäin. Ensimmäinen ja kolmas ryhmä ovat ainoat merkitysryhmät, joissa otoksesta esiintyvät feminiinijohdokset PŘÍTELKYNĚ ja YSTÄVĀTĀR. Nämä ovat kuviossa merkattuna pylvään vaaleampana värinä.



Kuvio 6: Sanojen YSTÄVÄ/PŘÍTEL ja YSTÄVÄTÄR/PŘÍTELKYNĚ merkitysjaako verrannollisen korpuksen satunnaisen otannan (n = 500) perusteella.

1. läheinen ihminen; **2.a** harrastaja – yhdistyksen nimessä; **2.b** harrastaja, suosija; **2.c** liittolainen, kannattaja; **3.** puoliso; **4.** puhuttelusana; **5.** kuvaannollinen ilmaus

Tuloksista voi tulkita, että perusmerkitysryhmään odotuksenmukaisesti kuuluu suurin osa konkordansseista niin suomessa kuin tšekissä. Kiinnostavampia ovat kunkin kielen toiseksi yleisimmät merkitykset, jotka ovat suomen otoksessa 'harrastaja' tai 'suosija' ja tšekissä 'puoliso'. Mikäli kyseessä on kuitenkin yhdistysten nimiin liittyvät YSTÄVÄ/PŘÍTEL sanat, löytyy niitä jonkin verran enemmän tšekin näytteestä. Kaikki muut ryhmät ovat kummassakin kielessä melko vähäisiä. Seuraavaksi analysoin tuloksia tarkemmin kategoria kerrallaan.

Perusmerkityksen ryhmä (**1.**) muodostaa kummassakin kielessä melkein yhtä korkean pylvään (esimerkkilauseet 20–21). Suomen puolesta YSTÄVÄTÄR esiintyy kahdeksassa konkordanssissa ja näistä jokaisessa kappaleessa on kyse toisen naisen ystävästä (22), mikä vastaa sanan sanakirjamääritystäkin. Vaikka tšekin sanakirjat eivät eksplisiittisesti mainitse, että PŘÍTELKYNĚ olisi välttämättä toisen naisen ystävä, näin kuitenkin näytteestä usein kävi ilmi (23). Tässä yhteydessä on syytä tarkentaa, että mikäli ystävyysuhteen toisella puolella oli mieshenkilö, paljastui taipumus tulkita sanaa PŘÍTELKYNĚ mieluummin miehen naisystäväksi kuin kaveriksi. Sanojen merkityksiä analysoidessa kuitenkin pyrin olemaan vailla heteronormatiivisia ennakkokäsityksiä ja käsitellä sanojen merkityksiä vain niiden leksikaalisen ympäristön perusteella, ja mikäli mahdollista, aina tarkastella laajempaa kontekstia. Merkityksen tulkitseminen pysyy kuitenkin aina jossain määrin subjektiivisena aivan kuin sanojen merkitykset eivät myöskään ole yksiselitteisiä.

20) *Keskustelin erään ystävän kanssa näistä sunnuntaikirjoituksista.*

21) *Někdy musím poopravit i některé názory mých přátel či příbuzných.* [Joskus minun täytyy korjata minun ystävieni ja sukulaisteni mielipiteitäkin.]

22) *Toden kertoakseni kuulin hänestä ensin ystävättäreltäni, joka ilmoitti löytäneensä minulle hyvän miesehdokkaan.*

23) „[...] *Na to nejsou zvyklí“*, říká její přítelkyně, advokátka Jana Hamplová. [”He eivät ole tottuneet siihen,” kertoo hänen¹⁷ ystävättärensä, asianajaja Jana Hamplová.]

Toisin kuin lekseemien YSTÄVÄ/PŘÍTEL peruserkityksessä, siirryttäessä toiseksi suurimpaan ryhmään (2.b), tilanne ei enää näytä yhtä tasapainoiselta. Suomen otoksessa YSTÄVÄ-sanalla tarkoitetaan jonkin harrastajaa melkein joka neljännessä konkordanssissa. Tämä on erittäin huomattava määrä verrattuna tšekkiin, jossa tulosten mukaan PŘÍTEL tarkoittaa jonkin harrastajaa vain harvoin. Asioihin, joiden ystävä voi olla, kuuluu suomen aineistossa ennen kaikkea käsitteet, jotka liittyvät taiteeseen (teatteri, musiikki, tanssi), urheiluun ja ruokaan (24–25), mutta suosiossa olevien asioiden skaala on tosin kirjavampi (26). Mainitsemisen arvoista joka tapauksessa on, että kyseisistä asioista voi suomen aineistossa pääosin pitää tai niitä voi harrastaa. Toisaalta tšekin konkordansseissa useammin ilmenee, että PŘÍTEL on jonkin aatteen tai käsitteen tukija, eli kannattamisen kohteena on useimmiten abstraktimpi asia (27–28). Toki tšekissä tätä ryhmää edustaa ainoastaan 5 kappaletta, mikä on liian vähän yleistyksen tekemiseen. Tämä hienojakoinen eroavuus suomen ja tšekin välillä ei ole niin yllättävää, sillä tšekin sanakirjatkaan eivät mainitse PŘÍTELI-lekseemin merkityksenä olevan ’harrastaja’ kuten suomen kielessä, vaan ’kannattaja’ ja ’suosija’. On tosin merkittävää, että kyseessä oleva ryhmä on suomen otoksessa toiseksi suurin, kun taas tšekistä löytyy vain muutama kappale.

24) *Paikallisille samban ystäville tuttu artisti on paljastunut tangon ja humpan ystäväksi.*

25) *Koska Tuija on intohimoinen valkosipulin ystävä, hän kokeili aikoinaan [...].*

26) *Tiettävästi myös itse Napoleon oli jojojen suuri ystävä.*

27) *Jsem spíš přítelem široce propracovaného systému práv lidských a občanských, který [...].* [Olen enemmänkin hienostuneen ihmis- ja kansalaisoikeusjärjestelmän ystävä, joka [...].]

28) *Doporučujeme [knihu] všem přátelům bombardování Jugoslávie, kterým dochází dech.* [Suosittelemme [kirjaa] kaikille Jugoslavian pommitusten ystävälle, joilta on loppumassa argumentit.]

¹⁷ Tšekin *její*-pronominista tulee selväksi, että lauseen kertoja on myös naispuolinen.

Silloin kun 'harrastaja'-merkitys liittyy yhdistysten ja järjestöjen nimiin (2.a) tulokset näyttävät toisenlaisina (29–30). Tällöin suhde on tasaisempi, tapauksia ollen tšekissä jopa enemmän. On huomion arvoista, että kun sanalla PŘÍTEL tarkoitetaan 'harrastajaa', se on hyvin todennäköisesti juuri jonkin järjestön nimenä. Lekseemien YSTÄVÄ/PŘÍTEL yhteyttä järjestöjen nimiin käsittelen myös kollokaattien yhteydessä (ks. luku 4.3.1).

29) *Käsityön Ystävät on toimittanut viidenkymmen nelion suuruisen ryijyn Montanan yliopistoon.*

30) [...] *řekl předseda Klubu přátel exotického ptactva Miroslav Procházka.* [...] kertoi Miroslav Procházka, Eksoottisten lintujen ystävien puheenjohtaja.]

Konkordansseja, joissa merkityksenä oli 'liittolainen' (2.c) on kappaleittain kummankin kielen otoksessa lähes saman verran. Sanat PŘÍTEL/YSTÄVÄ tässä ryhmässä esiintyvät poliittisessa kontekstissa yleensä erilaisten valtioiden ystävänä (31) tai viitaten liiton työtoveriin (32).

31) *Tässä on Pakistanin vankumaton ystävä Kiina ratkaisevassa asemassa.*

32) *První, jak také jinak, jsou na řadě přátelé z koalice.* [Ensimmäisinä, ei mikään yllätys, palkitaan ystävät koalitiosta.]

Tšekin otoksen toiseksi yleisimmäksi merkitysryhmäksi sijoittuu 'puoliso'-alamerkitys, jota on suomen näytteessä vain mitätön määrä. Kiinnostava havainto on myös se, että tässä merkityksessä esiintyy useammin feminiinijohdos PŘÍTELKYNĚ kuin pelkkä PŘÍTEL, vaikka PŘÍTELKYNĚ muodostaa vain 17 % näytteestä. Se myöskin tarkoittaa, että 73 % sanan PŘÍTELKYNĚ esiintymisistä on merkityksessä 'puoliso' tai 'tyttöystävä'.

Feminiinijohdoksen PŘÍTELKYNĚ korkean esiintymismäärän kolmannessa merkitysryhmässä selittää juuri ihmisen puolison yksikköluonne. Kun sanoja YSTÄVÄ/PŘÍTEL käytetään perusmerkityksessään, eli tarkoittaen läheistä ihmistä, on tšekissä tapana käyttää tunnusmerkitöntä miespuolista muotoa PŘÍTEL, mikäli puhutaan ystävästä tai ystävästä yleisesti. Toisaalta kun halutaan ilmaista, että jollakulla on kumppani, viitataan useammin tiettyyn ihmiseen, joten tämän sukupuoli on paljon tärkeämmässä roolissa. Sen lisäksi tšekissä on kieliopillisen suvun takia periaatteessa mahdotonta puhua ihmisestä ilmaisematta hänen sukupuoltaan. Tuloksia katsoessa herää kysymys, esiintyykö näytteessä nainen useammin tyttöystävän roolissa (33), kuin mies poikaystävän roolissa (34)? Toisaalta PŘÍTEL- ja PŘÍTELKYNĚ -sanojen esiintymismäärien ero ei ole niin suuri, että voisi varmasti päätellä asian olevan niin. Jatkotutkimus voisi joka tapauksessa tuoda mielenkiintoisia havaintoja. Varmaa kuitenkin on, että suomessa

YSTÄVÄ 'puoliso'-merkityksessä ilmenee vain harvoin, sillä suomen näytteestä sitä löytyy pelkästään kolme kappaletta (35).

33) *Pokud se hráči chtějí setkat s manželkami a přítelkyněmi, mohou tak učinit ve volném čase.* [Mikäli pelaajat haluavat tavata vaimojensa ja avovaimojensa kanssa, ne voivat tehdä niin vapaa-ajallaan.]

34) *Topmodelka Naomi Campbellová podepsala předmanželskou dohodu s italským přítelem Flaviem Briatorem.* [Supermalli Naomi Campbell allekirjoitti avioehdon poikaystävänsä Flavio Briatoren kanssa.]

35) *Prinssi Charlesin pitkäaikainen ystävä Camilla Parker Bowles remontteeraa uuden maaseutusuntonsa lähellä sijaitsevasta ulkorakennuksesta piilopirttiä [...]*

Konkordansseissa, joissa YSTÄVÄ/PŘÍTEL -lekseemejä käytetään toiselle puhumiseen (4.), löytyy kummankin kielen aineistosta vain muutamia. Suomen otoksessa puhuttelusanana YSTÄVÄ voi kyllä kantaa varoittavaa sävyä (36), muttei välttämättä (37). Tšekin näytteessä ei ilmene lauseita, joissa PŘÍTEL-sanalla olisi alentava sävy, mutta konkordansseja on tässä ryhmässä vain niukasti.

36) *Álkää hyvät ystävät tehkö sellaista virhettä.*

37) *Rakkaat ystävät, paljon kiitoksia.*

Viimeisen vertauskuvista koostuvan ryhmän (5.) osalta voi todeta, että suomenkielinen aineisto sisältää melko kirjavan listan siitä, mitä kaikkea voi kuvaannollisesti merkitä ystäväksi. Kyseessä on useimmiten elottomia olioita kuten esimerkkilauseessa (38). Tšekin otoksesta ei toisaalta löytynyt yhtäkään tämän ryhmän edustajaa, mikä ei poikkea sanakirjojen määritelmistä, joista lekseemin PŘÍTEL puhtaasti metaforinen käyttötapa puuttuu kokonaan. Tämä ei itsessään tarkoita, ettei tšekiksi voisi sanoa elottomia asioita vertauskuvallisesti ystäviksi, mutta tämä käyttötapa näyttää olevan paljon harvinaisempi kuin suomessa.

38) *Vaikka idässä puhutaan vastasta ja meillä täällä lännessä vihdasta, on rakkaus lehtevään ystäväämme yhteinen.*

Kaiken kaikkiaan voi merkitysanalyysin pohjalta todeta, että lekseemien YSTÄVÄ ja PŘÍTEL välillä on päällekkäinen polyseemisyys niin sanakirjojen kuin korpusaineiston mukaan. Joissakin merkityksissä sanat ovat varsin samankaltaiset ja toisissa merkityksissä niiden käyttö poikkeaa. Eroavuus ei niinkään näyttäyty siinä, että toista lekseemiä käytettäisiin merkityksessä, jota toisella lekseemillä ei olisi ollenkaan, vaan sanat eroavat yksittäisten merkitysten yleisyydessä.

Sanojen YSTÄVÄ ja PŘÍTEL peruserkitys vaikuttaa olevan melko identtinen sekä sanakirjojen määritelmässä, että korpusnäytteen taajuuksissa. Lekseemien toiseksi yleisimmässä merkityksessä näkyvät jo huomattavat eriävyydet. Suomenkielisessä aineistossa toiseksi yleisin on luokka, jossa YSTÄVÄ-sanalla tarkoitetaan jonkin harrastajaa ja kannattajaa. Tämän tyyppisiä konkordansseja on tšekissä näytteessä niukasti ja kannattamisen kohde yleensä abstraktimpi kuin suomessa. Tämä sanojen YSTÄVÄ/PŘÍTEL alamerkitys on etenkin tšekissä varsin yleinen järjestöjen nimissä. Tšekissä esiintyvyydeltään toiseksi suurin ryhmä edustaa merkitystä, jossa PŘÍTEL tai PŘÍTELKYNĚ tarkoittavat puolisoa. Etenkin sitten feminiinijohdoksen PŘÍTELKYNĚ esiintymät kertyvät tähän merkitykseen ja muodostavat yli puolet tämän luokan konkordansseista. Toisaalta suomen YSTÄVÄ/TÄR -lekseemeillä viitataan puolisoon vain poikkeuksellisesti. Kaikki muut merkitysryhmät ovat aineiston perusteella suhteellisen harvinaisia molemmissa kielissä.

Voi todeta, että sanakirjojen määritelmät ja korpusaineiston analyysin tulokset eivät eroa merkittävästi toisistaan. Ainoana poikkeuksena on PŘÍTEL/KYNĚ -lekseemien puolisoimerkityksen poissaolo vanhemmassa tšekissä sanakirjassa, mikä todistaa kyseisten sanojen semanttisen muutoksen. Täytyy kuitenkin pitää mielessä, että johtopäätökset on tehty vain rajoitetun otoksen pohjalta, joten päätelmät ovat ennemminkin suuntaa antavia kuin yleistäviä. Ottaen huomioon kollokaatioanalyysin, voi olettaa, että kollokaatit semanttisesti liittyvät kolmeen yleisimpään merkitysryhmään, eli ystävään läheisenä ihmisenä sekä suomessa että tšekissä, harrastajaan suomessa, ja puolisoon tšekissä.

4.3 Kollokaatioanalyysi

Tässä luvussa analysoin korpusaineistoa ja tulkitsen saamiani tuloksia. Jaottelen kollokaatioanalyysin kolmeen alalukuun. Ensin esittelen *MI*-testin tilastollisesti merkitseviä kollokaatteja ominaisuuksineen ja luokittelen ne kolmeen perusryhmään. Seuraavaksi käsittelen yksityiskohtaisemmin 30 tilastollisesti merkitsevintä kollokaattia. Viimeisessä alaluvussa keskityn kollokaattien merkitykseen. Luokittelen niitä semanttisiin ryhmiin, joista jokaista käsittelen tarkemmin pystyäkseni ymmärtämään paremmin sanojen YSTÄVÄ ja PŘÍTEL välisiä eroja ja yhtäläisyyksiä.

4.3.1 MI-testin kollokaattien luokittelu

Luvuissa 3.2.1 ja 3.2.2 kuvattujen menetelmien avulla sain lemmalistan kollokaateista, joiden *I*-arvo on korkeampi kuin 3,0 ja esiintymisfrekvenssi noodin ympärillä on vähintään 30. Verrannollisen korpuksen suomen aineistosta kollokaatteja löytyy yhteensä 360 ja tšekin osakorpuksesta 797 kappaletta. Tämä kollokaattimäärien ero johtuu siitä, että kumpaakin aineistoa käsiteltiin eri käsittelyohjelmalla hiukan eri menetelmällä sekä siitä, että osakorpuksset ovat erikokoisia.

Seuraavan analyysin tarkoituksena on ottaa käsiteltäväksi ensimmäiset 250 kollokaattia kummankin kielen kollokaattilistalta. Tšekin tuloksista pitää ensin poistaa ne kollokaatit, jotka saivat korkean *I*-arvon korpuksessa esiintyvistä toistosta. Toistoa ei esiinny yhtä paljon kuin suomen korpuksessa, mutta suomen konkordanssit käytiin jo alussa läpi poistaen samalla duplikaatit. Sama käytäntö ei ollut mahdollista tšekin puolella, koska tšekin aineistosta ei eroteltu tutkittavaa lekseemiä sisältäviä konkordansseja, vaan kollokaattien *I*-arvo laskettiin koko korpuksen perusteella. Toiston vallitsevaksi syyksi osoittautuivat TV- ja teatteriohjelmiston kuvaukset, jotka sanomalehdissä ilmenivät muuttumattomina jopa kymmeniä kertoja. Näin useampikin sana näistä kuvauksista nousee merkitseväksi kollokaatiksi. Nämä yhteensä 26 tapusta poistettiin ja korvattiin listan seuraavilla kollokaateilla saadakseni yhteensä 250 kollokaattia. Lopullisten kollokaattilistojen *I*-arvot vaihtelevat suomen puolesta arvoissa 9,9–3,7 ja tšekissä välillä 10,9–4,7 (ks. liite 1 ja 2).

Tšekin Kontextissa sekä suomen Korpissa pystyy suhteellisen helposti hakea kaikki yksittäisen kollokaatin esiintymät kontekstissaan. Kävin kaikki kollokaatit läpi, jotta voisin tarkastaa, johtuvatko niiden korkeat *I*-arvot luonnollisesta taipumuksesta esiintyä hakusanojen YSTÄVÄ/PŘÍTEL ympärillä, vai ovatko ne osana jonkin järjestön nimeä. Erilaisten yhdistysten, kerhojen ja järjestöjen nimissä ilmenee sanoja YSTÄVÄ/PŘÍTEL aineistossani jonkin verran ja näitä on syytä erotella muista niin sanotuista luonnollisista kollokaateista ja käsitellä analyysissä erikseen. Esimerkiksi tšekin kaikista yleisin kollokaatti, sana ŽEH (tuhkaus), liittyy kremointia kannattavaan yhdistykseen.¹⁸ Sanan poikkeuksellisen korkea *I*-arvo ei johdu siitä, että PŘÍTEL-lekseemin kontekstuaaliseen valintaan kuuluisi nimenomaan kollokaatti ŽEH, vaan se viittaa siihen, että tietyn yhdistyksen nimi mainitaan aineistossa syystä tai toisesta hyvin usein. Kiinnostavaksi havainnoksi kyllä osoittautuu tutkittavien lekseemien tendenssi esiintyä

¹⁸ Společnost přátel žehu (Tuhkauksen ystävät -yhdistys)

järjestöjen nimissä, mutta sillä, onko kyseessä kissojen tai käsityön ystävät, ei ole tutkielman näkökulmasta paljoa väliä.

Toinen ilmiö, joka nousee kollokaattilistoja tarkasteltaessa esille, on lekseemien YSTÄVÄ/PŘÍTEL tapa ilmetä erisnimien lähellä. Prinsessa Dianan kuolemasta kirjoitettiin sanomalehtiartikkeleissa aikoinaan huomattavasti. Lauseissa, kuten esimerkissä (39), mainitaan myös hänen miesystävänsä (tšek. *přítel*) nimi. Tämänkaltaisten konkordanssien iso määrä johtaa siihen, että tšekin hyvin merkitseviksi kollokaateiksi nousevat erisnimet *Diana*, sekä *Dodi*, *Al* ja *Fayed*.

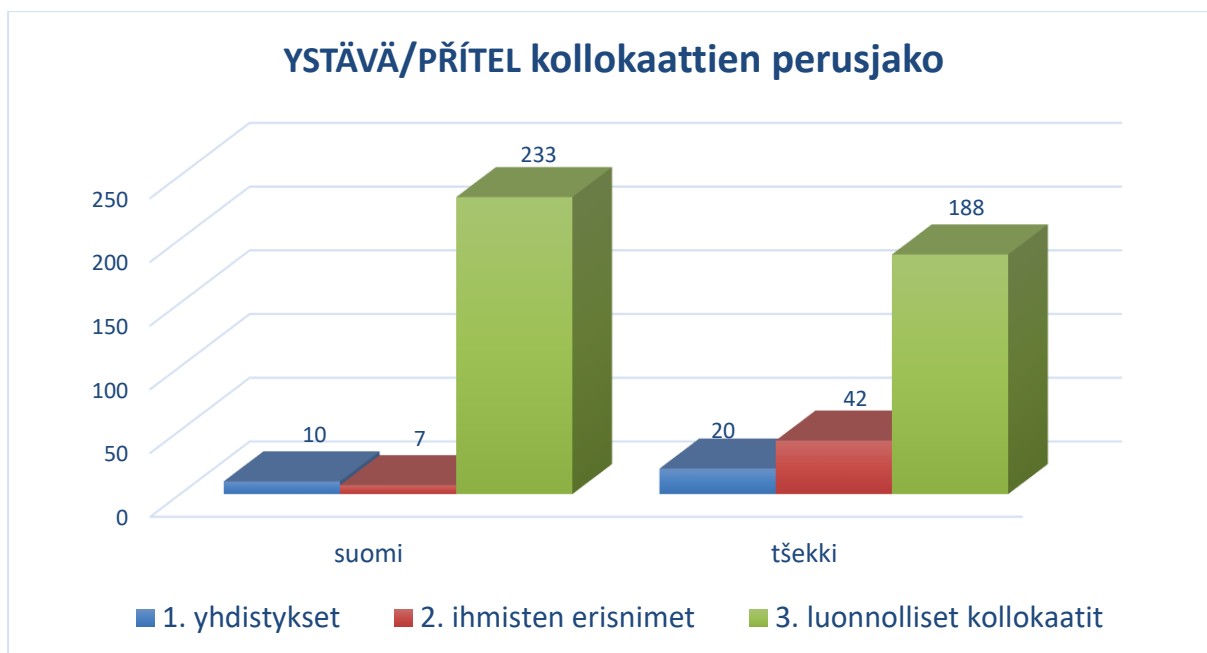
39) *Dvoupalcové titulky oznamují tragickou zprávu o úmrtí princezny Diany a jejího přítele Dodi Al Fayed.*

[Kaksituumaiset otsikot ilmoittavat traagisen uutisen prinsessa Dianan ja hänen (mies)ystävänsä Dodi Al Fayedin kuolemasta.]

Näissä tapauksissa tietyt nimet eivät ole kovin tärkeitä, sillä ne eivät paljasta YSTÄVÄ/PŘÍTEL -lekseemeistä muuta, kuin niiden yleisen tavan yhdistyä ihmisten erisnimien kanssa. Yksittäisiin nimiin vaikuttaa ennen kaikkea lehdissä juuri käsitelty aihe. Erikoistapauksia ovat myös erisnimet *Pentti* ja *Holappa*, jotka löytyvät kollokaattilistasta Holapan *Ystävän muotokuva* -nimisen romaanin vuoksi, joka mainitaan aineistossa useasti.

Kollokaatit olen jaotellut kolmeen perusryhmään, jotka nousevat selvästi esille listoja katsoessa ja kollokaatteja tarkastellessa kontekstissaan. Kontekstisidonnainen tarkastelu on tärkeää, jotta kollokaatteja voisi tarkemmin analysoida. Ensimmäiseen ryhmään kuuluvat yhdistysten ja järjestöjen nimiin liittyvät kollokaatit, kuten esimerkiksi sana KISSA, joka esiintyy useimmiten yhdistyksessä *Hämeenlinnan Kissojen Ystävät*. Toiseen ryhmään erottelin henkilöiden erisnimet ja kolmanteen pääryhmään kaikki muut kollokaatit, joita kutsun luonnollisiksi (kuvio 7). Viimeiseen ryhmään kuitenkin jätin kollokaatit, kuten YHDISTYS ja JÄRJESTÖ, koska nämä liittyvät pikemminkin sanojen YSTÄVÄ/PŘÍTEL kielenkäyttöön, eivätkä edusta pelkästään tiettyä yhdistystä.

Kollokaatit voisivat usein kuulua kahteenkin ryhmään ja näissä tapauksissa lajittelin ne sen mukaan, mitä ryhmää kollokaatti edustaa useammin. Esimerkiksi kollokaatti LUONTO ilmenee useimmiten ilmauksessa *luonnon ystävä/t*, kuitenkin sekä erisnimenä (*Luonnon Ystävät -yhdistys*), että yleisnimenä (*Hän on luonnon ystävä ja matkustaa junalla.*). Jälkimmäisen edustajia löytyy selvästi enemmän, joten luokittelin kollokaatin tämän mukaisesti luonnollisten kollokaattien ryhmään.



Kuvio 7: *I*-arvon merkitsevien kollokaattien perusjako (n = 250).

Tuloksia (kuviokuva 6) katsellessa huomaa, että yhdistyksiin liittyvät kollokaatit muodostavat molemmassa kielessä melko pienen osan kollokaattinäytteestä, kun taas erisnimien osuus edustaa tšekissä melkein 17 prosenttia kollokaateista. Suomessa tämä osuus pysyy kuitenkin hyvin matalana. Esittelen tässä lyhyesti kaksi ensimmäistä ryhmää ja seuraavissa luvuissa keskityn kollokaatteihin, jotka olen luokitellut luonnollisiksi, eli niihin, jotka eivät ole erisnimiä eivätkä liity tiettyihin järjestöihin.

Yhdistyksiin ja järjestöihin liittyvät kollokaatit vastaavat suomessa neljää prosenttia ja tšekissä kahdeksaa prosenttia kollokaattinäytteestä. Ne ovat itsessään suhteellisen pieniä osuuksia, mutta huomattavaa on, että useimmat näistä kollokaateista saavat varsin korkeat *I*-arvot ja sijoittuvat kollokaattilistan kärkeen (taulukko 6). Tähän ilmiöön on päässyynä aineistoni tekstilaji eli sanomalehtitekstit. Jos jonkin järjestön toiminta tai siihen liittyvä tapahtuma on herättänyt huomiota, on luontaista, että sen nimi esiintyy jonain ajankohtana artikkeleissa toistuvasti. Vaikka *I*-arvot ovat toiston vuoksi tavallaan vääristyneitä, yhdistysten nimet ovat väistämättä yhtenä ympäristönä, jossa lekseemit YSTÄVÄ/PŘÍTEL esiintyvät tyypillisesti molemmassa kielessä. Taulukossa 6 esittelen tämän alaryhmän kymmenen yleisintä kollokaattia.

Taulukko 6. Kymmenen yleisintä yhdistyksiin liittyvää kollokaattia *I*-arvon mukaisesti.

YSTÄVÄ				PŘÍTEL			
#	kollokaatti	frekv.	MI	#	kollokaatti	frekv.	MI
3	raittius	38	9,75	1	žeh [tuhkaus]	44	10,94
8	kansantanssi	36	8,38	14	lokálka [lähijuna]	37	8,36
11	käsityö	112	8,08	23	nerost [kivennäinen]	30	7,59
43	runo	94	6,09	30	Hradčany ¹⁹	43	7,28
53	kissa	52	5,93	32	vlastenecký [isänmaallinen]	58	7,21
88	laulu	105	5,45	43	dechovka [puhallinorkesteri]	43	6,93
122	Kreikka	34	5,02	82	zdravotně [terveydellisesti]	71	6,24
137	kirjasto	49	4,80	84	rodák [alkuasukas]	113	6,19
193	Israel	44	4,25	101	příroda [luonto]	362	5,93
211	nuori	269	4,10	115	Jeseníky ²⁰	34	5,75

Yleisesti ottaen molemmissa kielissä järjestöt, joiden osana on sana YSTÄVÄT/PŘÁTELÉ, liittyvät pääosin kulttuuriin ja harrastuksiin sekä erilaisiin maa-alueisiin. Jotkut kollokaatit koskevat yhtä tiettyä järjestöä, kuten *Raittiuden ystävät*, *Joensuun Kreikan ystävät* tai *Klub přátel lokálky* (Lähijunan ystävien kerho), kun taas toiset kollokaatit ilmenevät useammassa eri yhdistyksen nimessä. Näin esimerkiksi kollokaatti KANSANTANSSI voi viitata sekä *Turun, Helsingin* että *Suomen Kansantanssin ystäviin*. Vaikka yleensä kyseessä onkin jokin järjestö, yhdistys tai kerho, luettelo on odotettua kirjavampi ja konkordansseista löytyy esimerkiksi *Mieskuoro Laulun ystävät* tai *Klub přátel dechovky* [Puhallinorkesterin ystävien kerho], joka oikeastaan viittaa televisio-ohjelmaan. Suomen yhdistysten nimet, joiden osana on sana YSTÄVÄ, ilmenevät ylipäänsä muodossa, jossa kannatuksen kohde on genetiivissä, ja jota seuraa lekseemi YSTÄVÄ monikossa sekä mahdollisesti, muttei välttämättä lyhenne *ry*. Tšekin aineistossa toisaalta PŘÍTEL-monikon ja kohteen genetiivimuotoja tavallisesti edeltää tarkempi määrite kuten *společnost*, *sdržení*, *klub* tai *organizace* (yhdistys, järjestö, kerho, organisaatio). Kaikki tämän ryhmän kollokaatit kuuluvat sanojen YSTÄVÄ/PŘÍTEL alamerkitykseen, joka ilmaisee jonkin harrastuksen tai aatteen kannattajia.

Ihmisten erisnimien joukko edustaa suomen kollokaateista noin kolmea prosenttia ja tšekin kollokaateista n. 17 prosenttia 250 merkitsevimmästä kollokaatista. Erisnimet pääosin

¹⁹ Prahan kaupunginosa

²⁰ Vuoristoalue Tšekin Sleesiassa

ilmenevät joko ilmauksissa, missä kyseessä on itse ystävän nimi (40) tai sitten nimetyn ihmisen ystävä (41). Osa nimistä myös liittyy teokseen, jonka nimen osana on YSTÄVÄ/PŘÍTEL kuten jo mainittu Pentti Holapan *Ystävän muotokuva* tai John Irvingin *Ystäväni Owen Meany*.

40) *Hänen ystäviinsä lukeutuivat mm. Jean Sibelius, Akseli Gallen-Kallella ja Eliel Saarinen.*

41) *Jean Sibelius pakeni ystäviään, koska ei kestänyt heitä ilman viinaa.*

MI-testi nostaa harvinaisemmatkin kollokaatit merkitseviksi, muttei selitä miksi tšekin kollokaattilistasta löytyy erisnimiä melkein kuusinkertaisesti enemmän (kuvio 7). Tšekin kollokaattien kontekstin tarkastelusta selviää, että usein kyseessä on jonkin tunnetun henkilön tyttö/naisystävä (*přítelkyně*). Esimerkkilauseen (42) ja monien tämäntyyppisten konkordanssien vuoksi merkitseviksi kollokaateiksi nousevat nimet *Andrea* sekä *Verešová*.

42) *Asi zapomněl [...] na dramatické okolnosti rozchodu Jaromíra Jágra s přítelkyní Andreou Verešovou.*

[Kenties hän unohti [...] Jaromír Jágrin ja naisystävänsä Andrea Verešován eron dramaattiset seikat.]

Tämänkaltaisessa kontekstissa esiintyvät tšekin aineistossa myös miesten nimet, mutta paljon harvinaisemmin kuin naisten etu- ja sukunimet. Oletettavasti siis tšekin sanojen PŘÍTEL ja etenkin PŘÍTELKYNĚ alamerkitys 'puoliso' aiheuttaa ihmisten erisnimien korkean esiintymismäärän tšekin kollokaattilistassa.

4.3.2 Tilastollisesti merkitsevimmät kollokaatit

Jatkossa keskityn analyysissä kollokaatteihin, jotka eivät ole erisnimiä ja joiden korkea *I*-arvo ei johdu tavasta ilmetä jonkin järjestön nimen osana. Koska mainittuihin kahteen ryhmään kuuluu enemmän tšekin kuin suomen kollokaatteja, jälkimmäisen analyysin kollokaattimäärät olivat epätasapainossa (233:188). Sen takia lisäksi tšekin kollokaattijoukkoon alkuperäisen listan seuraavat 45 kollokaattia, poimien luontevasti vain ne kollokaatit, jotka eivät lukeutuisi ensimmäiseen kahteen ryhmään. Tällä tavoin saadaan uudestaan tasapainotetut kollokaattilistat koostuen kummankin kielen 233 kollokaatista.

Taulukko 7. Lekseemien YSTÄVÄ/PŘÍTEL 30 merkitsevintä kollokaattia *MI*-testissä

YSTÄVÄ			PŘÍTEL		
kollokaatti	frekvensi (L/R)	I	kollokaatti	frekvensi (L/R)	I
vannoutua ²¹	45 (38/7)	9,94	čtyřnohý [nelijalkainen]	180 (176/4)	10,46
sukulainen	543 (316/227)	9,90	nerozlučný [erottamaton]	39 (38/1)	9,16
tuttava	268 (52/216)	9,62	dlouholetý [pitkäaikainen]	660 (641/19)	8,86
työtoveri	153 (47/106)	9,30	příbuzný [sukulainen]	695 (451/244)	8,39
tunnustautua	72 (53/19)	9,27	důvěrný [luottamuksellinen]	164 (162/2)	8,39
uskollinen	90 (78/12)	8,23	posedět [istua (iltaa)]	52 (43/9)	8,22
läheinen	317 (272/45)	8,16	disent [vastarinta]	41 (2/39)	7,91
rakas	110 (90/20)	7,92	posezení [istuskelu]	98 (89/9)	7,90
vihollinen	105 (28/77)	7,82	rádce [neuvonantaja]	46 (24/22)	7,81
omainen	157 (127/30)	7,63	rozejít [erota]	135 (108/27)	7,67
perheenjäsen	37 (26/11)	7,50	kruh [piiri]	534 (502/32)	7,64
runous	40 (36/4)	7,34	spolek [seura]	439 (420/19)	7,61
merkkipäivä	31 (18/13)	7,31	obdivovatel [ihailija]	55 (12/43)	7,42
luona	74 (5/69)	7,29	oddaný [uskollinen]	34 (33/1)	7,40
toveri	39 (5/34)	7,27	vážený [arvoisa]	142 (139/3)	7,37
edesmennä ²²	38 (26/12)	7,17	rozchod [ero]	74 (64/10)	7,33
tervehtiä	41 (26/15)	7,13	věrný [uskollinen]	218 (180/38)	7,23
ry	232 (16/216)	7,11	blízký [läheinen]	1545 (1455/90)	7,21
tukija	48 (12/36)	7,10	pohádat [riidellä]	33 (22/11)	7,17
ilahduttaa	46 (32/14)	7,08	nepřítel [vihollinen]	265 (74/191)	7,16
kollega	50 (16/34)	6,91	oženit [mennä naimisiin]	78 (71/7)	7,15
isänmaa	36 (32/4)	6,81	milenska [rakastajatar]	68 (28/40)	7,10
syntymäpäivä	33 (10/23)	6,79	stýkat [tavata/tapailla]	43 (26/17)	7,00
lahja	88 (54/34)	6,76	milý [rakas]	185 (149/36)	6,97
yksinäinen	50 (32/18)	6,69	popovídat [jutella]	34 (28/6)	6,95
olut	77 (62/15)	6,66	mileneč [rakastaja]	86 (36/50)	6,90
innokas	60 (52/8)	6,61	spolužák [koulukaveri]	90 (45/45)	6,89
jutella	31 (16/15)	6,49	popíjet [juoda alkoholia]	37 (21/16)	6,88
tervetullut	52 (10/42)	6,48	spolupracovník [työkaveri]	265 (88/177)	6,86
lähin	48 (47/1)	6,45	modelka [naismalli]	92 (23/69)	6,84

²¹ Lemma VANNOUTUA esiintyy aineistossa lähes aina *vannoutunut*-sananmuotona.

²² Lemma EDESMENNÄ esiintyy aineistossa lähes aina *edesmennyt*-sananmuotona.

Taulukossa 7 on lueteltu kummankin kielen *MI*-testin antamat 30 tilastollisesti merkitsevintä kollokaattia. *I*-arvon lisäksi taulukossa on tarkennettu kollokaattien esiintymisfrekvenssiä noodin ympärillä sekä kuinka usein kollokaatti on ilmennyt noodin vasemmassa ja oikeassa kontekstissa.

Tuloksia tarkastellessa joidenkin kollokaattien korkeat *I*-arvot ehkä hieman yllättävät ja muistuttavat, että aineisto on 90-luvun sanomalehtikieltä. Suomen kaikista suurimman *I*-arvon saanut kollokaatti VANNOUTUA, joka ilmenee ainoastaan partisiippimuodossaan *vannoutunut*, voi havaita virkkeissä, joissa YSTÄVÄ-sanalla tarkoitetaan harrastuksen tai aatteen kannattajaa (43). Samankaltaista kontekstia suosivaksi osoittautuu myös kollokaatti TUNNUSTAUTUA (44).

43) *Tätä tosiasiaa ei vannoutuneinkaan sinikeltaisen jääkiekon ystävä voi kieltää.*

44) *Piispat tunnustautuvat eurooppalaisen humanismipainotteisen liberalismiin ystäviksi ja sanovat, että [...].*

Tšekin puolella kollokaatin DISENT (vastarinta, toisinajattelu) hyvin korkea *I*-arvo ennen kaikkea heijastaa 90-luvun sanomalehtien tiettyä aihealuetta. Sanalla viitataan totalitarismin aikana kiellettyyn toimintaan ja liikkeeseen, joka vastusti valtion virallista ajatussuuntaa (45).

45) *Havel v sobotu nepřijel do Bratislavy, kde se dríve po různých klubech a nadacích „ukrývali“ jeho starí přátelé z dob disentu a listopadové revoluce z roku 1989.*

[Havel ei tullut lauantaina Bratislavaan, missä aiemmin eri kerhoissa ja säätiöissä ”piileskelivät” hänen vanhat ystävänsä vastarinnan- ja vuoden 1989 marraskuun vallankumouksen ajoista.]

Molempien sanojen YSTÄVÄ/PŘÍTEL *I*-arvot ovat yllättävän identtisiä siinä mielessä, että taulukon sekä alku- että loppupään kummankin kielen kollokaattien *I*-arvot vastaavat toisiaan melkein täysin. Ei siis näy merkkejä siitä, että yksi sana kollokoisi omien kollokaattiansa kanssa voimakkaammin kuin toinen sana, vaan tilanne näyttäytyy hyvin tasaiseksi. *MI*-testin minimiarvo oli asetettu kolmeksi, ja vielä kollokaatit 30. sijalla ylittävät tämän arvon enemmän kuin kaksinkertaisesti. Vaikka korkeisiin *I*-arvoihin vaikuttavat myös korpusten hyvin laajat koot, kuitenkin se, että tšekin korpus on kaksi kertaa isompi kuin suomen korpus ei tunnu vaikuttavan kollokaattien *I*-arvoihin, toisin kuin niiden esiintymisfrekvensseihin.

Kollokaattien frekvenssit nousevat etenkin tšekin puolella hyvin korkealle, mutta merkitsevien kollokaattien joukosta löytyy myös niitä, jotka ovat juuri ja juuri ylittäneet minimifrekvenssin rajan. Korpusten eroavien kokojen vuoksi ei voi vertailla keskenään itse kollokaattien esiintymisfrekvenssejä, mutta voi tarkastella kuinka usein kollokaatit asettuvat noodin

vasemmalle ja oikealle puolelle. Jos otetaan huomioon vain ne kollokaatit, jotka selvästi profiloituvat yksipuoliseksi, kuten OMAINEN tai ODDANÝ (uskollinen), käy ilmi, että molemmissa kielissä kollokaatit suosivat sanojen YSTÄVÄ/PRÍTEL vasenta ympäristöä. Suomen aineistossa tämä preferenssi on kuitenkin jokseenkin lievempi kuin tšekissä. Toisin sanoen, lekseemin YSTÄVÄ merkitsevimpien kollokaattien listalta löytyy myös monta sellaista, jotka esiintyvät useammin oikealla puolella kuten LUONA ja TOVERI. Tšekissä sanan PRÍTEL kollokaattien vasemman kontekstin preferenssi on paljon selkeämpi ja näitä, jotka ilmenevät useammin oikealla puolella löytyy vain muutama, kuten DISENT. Tämän lisäksi molemmissa kielissä havaitaan kollokaatteja kuten JUTELLA ja RÁDCE [neuvonantaja], joita löytyy melkein tasaisesti sekä vasemmasta että oikeasta kontekstista (46–47).

46) *Ajanvietteistä mainioin on juoda teetä ja jutella hyvän ystävän kanssa.*

47) *Hyvien ystävien kanssa jutellaan ja ollaan ulkona.*

Ei ole yllättävää, että vasenta kontekstia suosivat kollokaatit ovat sanaluokiltaan adjektiivieja, kun taas substantiivit ja verbit hajaantuvat nooiden ympärille tasaisemmin. Substantiivit muodostavat 30 merkitsevimmästä kollokaatista noin puolet sekä suomessa että tšekissä. Lukumäärältään substantiivien jälkeen seuraavina ovat adjektiivit ja sitten verbit. Sanaluokkien puolesta suomen ja tšekin väliltä ei löydy huomattavia eroja, lukuun ottamatta yhtä postpositiota LUONA suomen kollokaattien joukossa.

Substantiivikollokaatit kytkeytyvät semanttisesti ennen kaikkea ihmisen perheeseen (SUKULAINEN, PERHEENJÄSEN) ja lähipiirissä oleviin ihmisiin (TUTTAVA, SPOLUPRACOVNÍK [työkaveri]). Myös verbeistä johdetut substantiivit viittaavat toimintaan, jota yleensä yhdistetään läheiseen ystävyysuhteeseen (TUKIJA, RÁDCE [neuvonantaja], OBDIVOVATEL [ihailija]). Kummankin kielen kollokaattilistalla on lisäksi sanoja, jotka osoittavat joko ihmisjoukkoihin tai virallisiin yhdistyksiin (RY, KRUH [piiri], SPOLEK [seura]).

Adjektiivit ovat molemmissa kollokaattilistassa hyvin yleisiä, tšekissä ne jopa vievät kolme ensimmäistä paikkaa. Lähes kaikki adjektiivikollokaatit kantavat positiivisia sävyjä ja kuvailevat, minkälainen kyseinen ystävä on (USKOLLINEN, ČTYŘNOHÝ [nelijalkainen]) tai täsmentävät ystävyysuhteen luonnetta (LÄHEINEN, NEROZLUČNÝ [erottamaton], DLOUHOLETÝ [pitkäaikainen]).

Verbejä löytyy yleisten kollokaattien joukosta suhteellisen vähän molemmissa kielissä, ja ne lähinnä kuvaavat ystävien välistä toimintaa (JUTELLA, STÝKAT [tapaila]). Verbikollokaattien esiintymisfrekvenssit lekseemien YSTÄVÄ/PRÍTEL kontekstissa ovat myös keskimäärin matalampia kuin substantiivii- ja adjektiivikollokaattien frekvenssit. Tätä ilmiötä voi ainakin

osittain selittää sillä, että noodeihin liittyvien verbien ei välttämättä tarvitse sijaita valitun L4:R4 kollokaatiovälin puitteissa, toisin kuin esimerkiksi adjektiivien.

Kollokaatteja analysoidessa silmiinpistävää on, että useita sanoja löytyy molemmista listoista, kuten SUKULAINEN – PŘÍBUZNÝ, LÄHEINEN – BLÍZKÝ tai VIHOLLINEN – NEPŘÍTEL. Päätin tarkastella laajempaa kollokaattiluetteloa, eli kummankin kielen 233 kollokaattia, löytääkseni paljonko yhteisiä sanoja näiden välillä ilmenee. Oletus osoittautui oikeaksi, kun paljastui, että ylipäänsä 40 prosenttia valituista kollokaateista löytyy kummankin kielen listalta. Tämä melko korkea luku kertoo, että lekseemit YSTÄVÄ ja PŘÍTEL jakavat merkittävää osaa kontekstuaalisesta profiilistaan keskenään. Palaan laajempaan kollokaattijoukkoon ja niiden yhteyksiin ja eroihin seuraavassa luvussa, jossa käsittelen kollokaattien semanttisia ryhmiä.

Huomattavin ero suomen ja tšekin 30 merkitsevimmän kollokaatin osalta vaikuttaa olevan tšekin romanttiseen suhteeseen liittyvien kollokaattien määrä (ROZEJÍT [erota], ROZCHOD [ero], OŽENIT [mennä naimisiin], MILENEC/KA [rakastaja/tar]). Näitä suomen merkitsevimmistä kollokaateista ei löydy ollenkaan, ja nämä todennäköisesti useimmiten liittyvät PŘÍTEL-sanana 'puoliso'-merkitykseen (48), vaikka näin ei välttämättä aina ole (49).

48) *Dozvěděl se, že jeho milenka si našla nového přítele.*

[Hän sai selville, että hänen rakastajattarensa löysi uuden (mies)ystävän.]

49) „*Jsme jenom přátelé, ne milenci...*“

[”Olemme vain ystäviä, ei rakastajia...”]

Sekä suomen että tšekin kollokaateista nousee esiin muutama, joista ei ehkä heti saa selville, miten ne tarkalleen yhdistyvät YSTÄVÄ/PŘÍTEL-sanapariin, ja joiden kontekstia kannattaa hiukan tutkia. Suomen puolella ainakin kollokaatit RUNOUS, ISÄNMAA ja OLUT vaativat niiden ympäristön tarkastelua, jonka jälkeen tuli selväksi, että nämä kaikki esiintyvät lähes poikkeuksetta 'harrastaja'-merkityksessä (50–51). Tämänäköisiä kollokaatteja ei löydy tšekin merkitsevimpien kollokaattien joukosta lainkaan, lukuun ottamatta itse järjestöä tarkoittava sana SPOLEK.

50) *Se olisi tietenkin tervetullutta myös suomalaisen oluen ystävälle.*

51) *Tämä herra ilmoittautui spontaanisti runouden ystäväksi.*

Tšekin osalta jokseenkin ihmetyttää kollokaatti MODELKA (naismalli). Kun tarkastelee virkkeitä, joissa esiintyvät sanat MODELKA ja PŘÍTEL, tai tarkalleen ottaen PŘÍTELKYNĚ, selviää, miksi sen *I*-arvo sekä esiintymisfrekvenssi ovat niin korkeat.

51) *Prvního potomka čeká s přítelkyní, modelkou Larou Felthamovou, australský tenista Pat Rafter.*

[Australialainen tennispelaaja Pat Rafter odottaa esikoista (nais)ystävänsä mallin Lara Feltham kanssa.]

Konkordanssi (51) sekä aiempi esimerkki (34) edustavat aika yleistä ilmiötä tšekin aineistossa, joka on todennäköisesti peräisin lehtien viihdeosastolta, ja joka käsittelee julkisuuden henkilöiden elämää. Tässä kontekstissa usein mainitaan heidän puolisonsa ja mahdollisesti myös ammattinsa, mikä heijastuu sanojen PŘÍTEL/KYNĚ kollokaatteihin. Tätä, sekä muita tässä luvussa mainittuja semantiikkaan liittyviä seikkoja käsittelem tarkemmin seuraavassa luvussa.

4.3.3 Semanttinen analyysi

Tämän luvun kvalitatiivisessa analyysissä tutkin kollokaattien semanttisia profiileja ryhmitellen lähisyronymien YSTÄVÄ/PŘÍTEL kollokaatteja merkitysryhmiin. Kiinnostuksen kohteena on tarkastella laajempaa ja usein abstraktimpaa kontekstia, jossa valitut lekseemit toistuvasti esiintyvät sekä ymmärtää paremmin niiden yhteneviä ja eroavia piirteitä.

Sekä suomen että tšekin kollokaateista otan analysoitavaksi *MI*-testin antamat ensimmäiset 233 kollokaattia. Yksityiskohtaisempaa käsittelyä varten valitsin niitä alun perin 250 kappaletta, mutta kuten luvussa 4.3.1 kävi ilmi, karsin pois ihmisten erisnimet ja järjestöjen nimiin liittyvät kollokaatit.

Kvantitatiiviseen analyysiin verrattuna semanttinen tarkastelu on aina jonkin verran subjektiivista riippuen tutkijan omasta kielitajusta ja semanttisten avaruuksien käsityksistä. Tämä alkaa jo pelkästä yksittäisten merkitysryhmien nimeämisestä. Koen siis erittäin tärkeäksi perustella riittävästi valitsemaani lähestymistapaa ja periaatteita, joita seurasin kollokaattien luokitellussa. Etenin alussa niin, että jo kollokaattilistojen ensikatsauksella nousi selkeästi esille joitakin semanttisia yläkäsitteitä, joiden ympärille asetin niihin yksiselitteisesti kuuluvat kollokaatit. Kun kävin aineiston sanat läpi yksi kerrallaan, alkoi muodostua uusia merkitysryhmiä ja vanhoja piti välillä muokata ja järjestää uudestaan.

Yleiseksi hankaluudeksi osoittautui se, että yksittäinen kollokaatti sopisi kahteen tai useampaankin luokkaan. Joskus tämä johtui sanan useammasta merkityksestä, kuten esimerkiksi tšekin kollokaatti SCHÁZET, joka tarkoittaa suomeksi sekä *tavata* että *kaivata*. Tämänkaltaiset ongelmat yleensä ratkesivat melko helposti, kun selvitin korpuksen konkordansseista, missä merkityksessä ja semanttisessa kontekstissa kollokaatti esiintyy useammin. Välillä oli

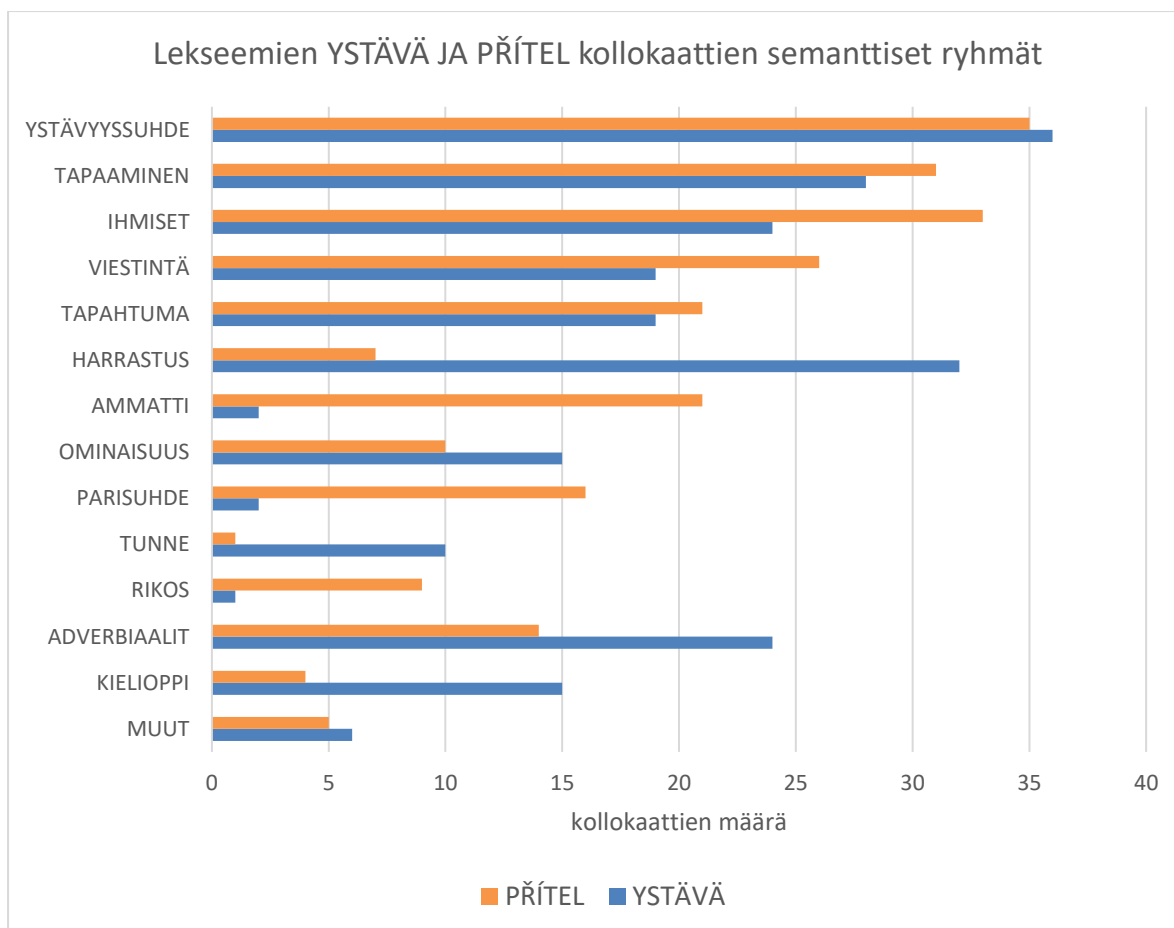
kuitenkin haastavaa luokitella kollokaatteja, vaikka niiden merkitys oli selkeä. Näin piti esimerkiksi päätellä, ilmaiseeko kollokaatti LUOTETTAVA, ilmauksissa *luotettava ystävä*, enemmän ihmisen ominaisuutta, kuuluen 'ominaisuus'-ryhmään, vai onko pikemminkin kyseessä ystävyysuhteen luonne ja täten kollokaatti sopisi 'ystävyyssuhde'-ryhmään. Tämän sekä vastaavankaltaisten kollokaattien puolesta tein jälkimmäistä ryhmää puoltavan päätöksen, mutta valinnat eivät aina olleet helppoja, ja toinen tutkija voisi hyvinkin olla eri mieltä.

Kuten jo edellisessä luvussa kävi ilmi, noin 40 % suomen ja tšekin kollokaateista osoittautui toistensa vastineiksi. Mielenkiintoiseksi osoittautui, että joissakin tapauksissa nämä eri kielten lähisyronyymit päättyivät eri merkitysryhmiin. Suomen kontekstissa ilmaus *entinen ystävä* tyypillisesti liittyy ystävyden loppuun (52), joten sen mukaisesti lajittelin kollokaatin ENTINEN 'ystävyyssuhde'-ryhmään. Toisaalta tšekin aineistossa vastaava ilmaus *bývalý přítel* useimmiten merkitsee parisuhteen loppua (53), jonka takia kollokaatin BÝVALÝ paikka on 'parisuhte'-ryhmässä. Tämä näennäinen epäloogisuus on oikeastaan melko luonteva: kuten sanat YSTÄVÄ ja PŘÍTEL ilmaantuvat osittain eri merkityksissä, on odotuksenmukaista, että vastaavasti näiden kollokaatit muotoilevat eri semanttisia ympäristöjä.

52) *Noina kuukausina monista entisistä ystävistä – tai ainakin työtovereista – tuli vihamiehiä.*

53) *Otcem syna je Petin bývalý přítel Damian Harris.* [Pojan isä on Petan entinen (mies)ystävä Damian Harris]

Muodostin kaiken kaikkiaan 13 merkitysryhmää, joiden nimet sekä niihin kuuluvien kollokaattien määrät näkyvät kuviossa 8 (merkitysryhmien täydellisen listan kollokaateista ks. liite 3).



Kuvio 8: Lekseemien YSTÄVÄ ja PŘÍTEL MI-testin kollokaattien semanttiset ryhmät (n = 233).

Joidenkin ryhmien kohdalla huomaa, että yksi merkitysryhmä on ikään kuin toisen ryhmän yläkäsitteenä. Esimerkiksi 'tapahtuma'-ryhmän kollokaatit liittyvät ihmisten yhteiseen ajanviettoon, mikä luonnollisesti edellyttää, että ihmiset ensin tapaavat. Näin ollen 'tapahtuma'-ryhmän kollokaatit sopisivat hyvinkin myös 'tapaaminen'-ryhmään. Katsoin kuitenkin parhaaksi erotella nämä mahdollisimman tarkkoihin luokkiin, mikäli löytyi tarpeeksi tietynlaisia kollokaatteja oman ryhmän muodostamiseen.

Kuvion 8 tuloksia tulkitessa nousee esille useampi merkitysryhmä, jotka sisältävät lähes saman verran sekä suomen, että tšekin kollokaatteja. Tasaisimmat tulokset saivat 'ystävyyssuhde'-, 'tapaaminen'- ja 'tapahtuma'-merkitysryhmät. Suomen kielen osalta 'harrastus', 'tunne' ja 'kielioppi' ovat ne ryhmät, joissa esiintyy huomattavasti enemmän kollokaatteja kuin tšekissä. Toisaalta tšekin kollokaatit ovat yliedustettuja merkitysryhmissä 'ammatti', 'parisuhde' ja 'rikos'. Lisäksi löytyi yhteensä 10 kappaletta kollokaatteja, jotka eivät sopineet mihinkään muotoilemaani ryhmään, ja nämä olen merkinnyt 'muut'-ryhmäksi.

Seuraavaksi käsittelen merkitysryhmät yksi kerrallaan ja keskityn siihen, miten määrittelen kyseessä olevia kategorioita ja perustelen, millaiset kollokaatit päätin luokitella kuhunkin. Tulkitsen jokaisen ryhmän tuloksia painopisteen ollen kummankin kielen eroavissa ja yhtenevissä seikoissa. Jokaisen merkitysryhmän kohdalla lukee suomen ja tšekin kollokaattien prosenttiosuus koko otannasta, sekä niiden tarkka kappalemäärä.

YSTÄVYYSSUHDE – suomi 15 % (36), tšekki 15 % (35)

Sanojen YSTÄVÄ sekä PŘÍTEL leksikaalisesta ympäristöstä löytyy lähes saman verran 'ystävyyssuhde'-merkitysryhmään kuuluvia kollokaatteja. Nämä kaikki yhteenlaskettuna tekevät 'ystävyyssuhde'-ryhmästä kollokaattien osuudeltaan suurimman.

'Ystävyyssuhde' on yksi kategorioista, joka ei tullut heti esille, kun aloin tarkastella kollokaatteja. Vasta tarkemmin analyysin edetessä tämä selvästi muodostui omaksi ryhmäkseen. Luokittelin siihen kollokaatit, jotka liittyvät ihmisiin tutustumiseen, ystävien etsimiseen, löytämiseen sekä menettämiseen, toisin sanoen ystävyuden eri vaiheisiin. Nämä ovat pääsääntöisesti verbejä (TUTUSTUA, LÖYTÄÄ, HLEDAT [etsiä], OPUSTIT [jättää]). Kollokaattiaineistossa esiintyi myös sanoja, jotka ilmaisevat ystävyyssuhteessa olevien ihmisten positiivista vuoro-vaikutusta, eli sitä, mitä ystävät keskenään tekevät (ARVOSTAA, LUOTTAA, AUTTAA, POMOC [apu], RADIT [antaa neuvoa]). Seuraava alaryhmä koostuu kollokaateista, jotka osoittavat ystävyuden kestävyyyden tai aikaan, milloin tämä syntyi (LAPSUUS, PITKÄAIKAINEN, NĚKDEJŠÍ [senaikainen]). Lopulta 'ystävyyssuhde'-merkitysryhmään kuuluvat adjektiivit, jotka kuvailevat ystävän tai ystävyuden laatua (LUOTETTAVA, AITO, VĚRNÝ [uskollinen]). Tämä viimeksi mainittu, pääosin adjektiiveista koostuva, alaryhmä voisi hyvinkin kuulua myös 'ominaisuus'-merkitysryhmään, johon erottelin ihmistä kuvailevia kollokaatteja. Kun tämän tyyppisiä kollokaatteja tutkii tarkemmin kontekstissaan, huomaa kuitenkin eron adjektiivien MUKAVA/HAUSKA/ILOINEN ja LUOTETTAVA/USKOLLINEN/HYVÄ välillä. Ensimmäinen joukko luonnehtii, millainen tai missä mielentilassa kyseinen ihminen on (54), eivätkä nämä välttämättä ilmaise ystävän ominaisuutta. Jälkimmäiset kollokaatit toisaalta viittaavat siihen, miksi joku on ystävä, ja kyseessä on nimenomaan ominaisuus, joka tekee ihmisestä hyvän ystävän (55).

54) Hän on *iloinen*, että elävistä *ystävistä* kaksi läheistä, [...] ovat mukana kantaesityksessä.

55) Janne Manninen on yksi postimies Mauri Ilvosen *uskollisimmista ystäväistä*.

'Ystävyysuhde'-merkitysryhmään kuuluvien kollokaattien yleisyys molemmissa kielissä ei tullut yllätyksenä. YSTÄVÄ/PŘÍTEL -lekseemien perusmerkitys onkin 'läheinen ihminen', ja ystävien keskeisen suhteen kuvaus, sekä sen muodostaminen ja ylläpito ovat tämän merkityksen ytimessä.

OMINAISUUS – suomi 6 % (15), tšekki 4 % (10)

'Ominaisuus'-merkitysryhmään päätyivät yleisesti ottaen adjektiivit, jotka kuvailevat, millainen ihminen on (yhdeällä poikkeuksella: ČTYŘNOHÝ [nelijalkainen]). Kuten jo mainittiin edellisen ryhmän kohdalla, ystävää luonnehtivat kollokaatit ovat melko lähellä ystävyysuhdetta kuvaavia adjektiiveja. Uskon kuitenkin olevan perusteltua erotella nämä omaksi merkitysryhmäkseen, sillä ne eivät välttämättä liity mitenkään itse ystävyteen. Konkordansseja tarkastellessa selviää, ettei ole aina välttämättä kyse henkilön ominaisuudesta (56), vaan kohteena voi olla myös ystävien yhteisestä toiminnasta syntynyt mieliala (57).

56) *Ja mikäs tässä olleessa, kun on uusi mukava ystäväkin löytynyt, [...].*

57) *On mukavaa tavata ystäviään edes kerran vuodessa vapaammissa merkeissä.*

Tässä melko pienessä kategoriassa esiintyy sävyiltään sekä positiivisia (ILOINEN, HAUSKA, HODNÝ [kiltti]), neutraaleja (ULKOMAALAINEN, TUNTEMATON, HOMOSEXUALNÍ [homoseksuaalinen]) että negatiivisia (KÖYHÄ, YKSINÄINEN) adjektiiveja. 'Ominaisuus'-ryhmän kollokaatteja löytyy hiukan enemmän suomen kielestä, jonka aineistossa ilmenee myös useampia paikkaan viittaavia adjektiiveja (AMERIKKALAINEN, SAKSALAINEN). Tämänäyttypistä kollokaattia tšekin puolella edustaa hieman yllättävästi ainoastaan adjektiivi PAŘÍŽSKÝ (pariisilainen).

TAPAAMINEN – suomi 12 % (28), tšekki 13 % (31)

'Tapaaminen' on toiseksi suurin semanttinen ryhmä, kun lekseemien YSTÄVÄ ja PŘÍTEL kollokaatit lasketaan yhteen. Nämä voisi edelleen jakaa kahteen alaryhmään, jotka selkeästi tulevat esille. Ensiksi suuri määrä kollokaatteja liittyy jollakin tavalla vierailuun (58) sisältäen sanoja kuten KUTSU, VIERAILLA, TERVETULLUT ja ZVÁT (kutsua). Toiseksi kollokaatit ilmaisevat ihmisten tapaamista ja yhdessä olemista (59), johon liittyvät esimerkiksi kollokaatit SEURA, KANSSA, SEJÍT (tavata) ja SPOLEČNĚ (yhdessä).

58) *Marraskuussa sain vieraakseni tänne vanhoja ystäviäni, Fedja – sedän, [...].*

59) *„Minulý týden jsem potkala svého přítele Eli Wiesela.“ [”Viime viikolla tapasin ystävänäni Eli Wieselin.”]*

Suurin osa 'tapaaminen'-merkitysryhmän kollokaateista oli hyvin yksiselitteisiä, mutta löytyi myös muutama sana, kuten KAHVI, joiden oikeaan ryhmään kuulumista piti tarkastella kontekstista (60).

60) *Kerran olin kahvilla kauniin ystäväni kanssa ja kuulin, kuinka viereisen pöydän pojat puhuivat ystäväni kauneudesta.*

Kollokaattien yleiskatsauksessa mainitsin, että noin 40 % kaikista kollokaateista on toistensa vastineita. 'Tapaaminen'-ryhmän puolesta on huomionarvoista, että lähes 80 % kollokaateista on toistensa käännösekvivalentteja. Lähisyronyymit eivät tietenkään ilmene pelkästään suomen ja tšekin välillä, vaan myöskin yhden kielen puitteissa. Näin ollen esimerkiksi kollokaatille TAVATA löytyy tästä ryhmästä kuusi tšekin lähisyronyymia.

Yllä mainitut huomiot korostavat tosiasiaa, että tapaaminen ja vierailu ovat oleellinen osa ystävien välistä interaktiota. Kuten jo 'ystävyyssuhde'-ryhmässä, myös 'tapaaminen'-merkitysryhmässä painottuu läheisten ihmisten suhteeseen kuuluva sosiaalinen aspekti ja ystävien keskinäinen vuorovaikutus. Näiden seikkojen osalta ei suomen ja tšekin kielen osalta löydy eroja, vaan niissä päinvastoin ilmenee vahva yhtäläisyys.

TAPAHTUMA – suomi 8 % (19), tšekki 9 % (21)

'Tapahtuma'-merkitysryhmää voisi mieltä edellä käsitellyn 'tapaaminen'-merkitysryhmän alakäsitteenä, sillä ystävien välinen kohtaaminen ja vuorovaikutus ovat edellytyksiä tapahtumien ilmenemiselle. Kuten myös edellisessä kategoriassa, niin tässäkin on lähes saman verran tšekin ja suomen kollokaatteja. Jos molemmat ryhmät laskettaisiin yhteen, saisi tulokseksi yhteensä 21 % kaikista kollokaateista, mikä on melko huomattava osuus. On kuitenkin syytä käsitellä näitä kahta merkitysryhmää erikseen: jos 'tapaamisen' kollokaatit kertovat, että ystävät tapaavat ja viettävät aikaa yhdessä, niin 'tapahtuma' tarkoittaa mitä he tekevät.

Kollokaatit, jotka luokittelin tähän ryhmään, viittaavat pääosin juhlimiseen, viihteeseen, ajanviettoon, juomiseen ja syömiseen (SYNTYMÄPÄIVÄ, TANSSIA, RAVINTOLA, USPOŘÁDAT [järjestää], VÍNO [viini]). Tšekin puolella ilmenee useampi kollokaatti, joka viittaa kapakkaan ja alkoholin nauttimiseen (61), mikä ainakin tšekin kulttuurisessa kontekstissa on suosittu ajanviete ystävien kesken.

61) *Svíj volný čas tráví mladá učitelka s českými práteli v hospodě, kde se zdokonaluje v konverzaci.* [Vapaa-aikaansa nuori opettaja viettää ystävien kesken kapakassa, missä hän kehittää viestintätaitojaan.]

Myös 'tapahtuma'-ryhmässä esiintyvät kollokaatit, jotka voisivat kuulua toiseenkin kategoriaan. Esimerkiksi sanat YÖ ja ILTA ovat ajan kollokaatteja (jotka olen yleisesti ottaen luokitellut 'kielioppi'-ryhmään), mutta kontekstin mukaan nämä useimmiten viittaavat pikemminkin illanviettoon ja yöelämään (62).

62) *Muschter piti luennon ja lähti ystäviensä kanssa vielä istumaan iltaa.*

'Tapahtuma'-merkitysryhmän suhteellinen suuruus sekä tšekin että suomen aineistossa osoittaa ystävien kesken laatuajan viettämisen ja sosiaalisten tilanteiden merkitsevyydestä. Toisaalta ottaen huomioon sanomalehtien tekstilajin, on todennäköistä, että virallisista tapahtumista ja merkittävistä tilaisuuksista tiedotetaan useammin kuin muista arkielämän toiminnoista.

IHMISET – suomi 10 % (24), tšekki 14 % (33)

Kolmanneksi suurimman 'ihmiset'-merkitysryhmän kollokaatteja löytyy hieman enemmän tšekin aineistossa. Olen tähän luokkaan jaotellut ihmisiä tarkoittavia substantiiveja, jotka useimmissa tapauksissa ilmaisevat henkilön lähipiiriä. Ihmisiä merkitseviä kollokaatteja esiintyy laajalti myös 'ammatti'- ja jonkin verran 'parisuhde'-merkitysryhmissä. Nämä kollokaatit olen kuitenkin luokitellut mieluummin kyseisiin ryhmiin, sillä ne ilmaisevat spesifimpää merkitystä kuin ihmisiä yleisesti.

'Ihmiset'-ryhmän puitteissa muodostui useampi alaryhmä, joista lukuisin ilmaisee perheeseen kuuluvia ihmisiä (ĀITI, VELI, PŘÍBUZNÝ [sukulainen]). Perhepiirin osana on myös kollokaatti KOIRA ja sen tšekin vastine PES, sillä konkordansseista tulee selväksi, että koiraa pidetään perheenjäsenenä. Perheen lisäksi esiintyy kollokaatteja, jotka viittaavat läheisiin ihmisiin (OMAINEN, KAMARÁD [kaveri]) sekä henkilöihin, jotka kuuluvat ihmisen tuttavuuspiiriin (NAAPURI, KOLLEGA, SPOLUŽÁK [koulukaveri]). Näiden tiettyjen henkilöiden lisäksi ilmenevät substantiivit, jotka ilmaisevat ikään kuin henkilön tehtävää tai toimintaa ystäväänsä kohtaan (TUKIJA, OCHRÁNCE [puolustaja], PŘÍZNIVEC [kannattaja]). Kummankin kielen kollokaateista löytyy myös ystävän vastakohta (VIHOLLINEN/NEPŘÍTEL) ja ihmisryhmään viittaavat substantiivit (PERHE, PARTA [porukka], KRUH [piiri]).

Yksi syy siihen, miksi suhteellinen moni kollokaatti edustaa ihmisen lähipiiriin kuuluvia ihmisiä, johtunee siitä, että itse ystävä on erottamaton osa ihmisen sisäpiiriä, joten ystävä usein mainitaan muidenkin läheisten ihmisten yhteydessä (63). Tšekin kollokattilistalta löytyy enemmän 'ihmiset'-merkitysryhmän kollokaatteja mahdollisesti sen takia, että useampi kollokaatti esiintyy sekä feminiini-, että maskuliinimuodossa. Tämä johtunee siitä, että mikäli viitataan

tiettyyn henkilöön, on tšekissä pakollista ilmaista hänen sukupuoltaan. Keinona edistää sukupuolineutraalia kielenkäyttöä mainitsemalla sekä sanan maskuliini- että feminiinimuoto geneerisen maskuliinin sijaan esiintyy myös, vaikka 90-luvun teksteissä melko harvinaisesti (64).

63) *Kaikista henkilöistä ei enää tiedetä keitä he ovat, mutta lähipiiri tunnetaan, äiti, ystävät ja tuttavat.*

64) *Ale nakonec souhlasil, že se setká v hudbě a písničích se svými přáteli, kolegy a kolegyněmi na jednom pódiu.* [Hän lopulta suostui siihen, että hän esittää musiikkikappaleita samalla lavalla yhdessä ystäviensä, kollegoidensa (sekä naiskollegoidensa) kanssa.]

VIESTINTÄ – suomi 8 % (19), tšekki 11 % (26)

'Viestintä' on viimeinen suhteellisen iso semanttinen ryhmä, johon kuuluu lähes sama osuus sekä suomen että tšekin kollokaatteja. Tämän ryhmän sanat eivät rajoitu pelkästään ihmisten suulliseen vuorovaikutukseen (JUTELLA, PROSIT [pyytää]), vaan myös kirjalliseen tiedonvälitykseen (KIRJOITTA, LÄHETTÄÄ [korttia]). Lisäsin 'viestintä'-kategoriaan myös kollokaatteja, jotka viittaavat kognitioon (YMMÄRTÄÄ, VZPOMÍNAT [muistella]). Sanaluokaltaan ryhmän kollokaatit ovat ylivoimaisesti verbejä, mutta löytyy myös muutama substantiivi (PUHELIN, VZPOMÍNKA [muisto]).

Jo edellisistä merkitysryhmistä kävi ilmi, että mikäli kollokaatit liittyvät 'läheinen ihminen'-merkitykseen, esiintyy kummankin kielen välillä monta ekvivalenttia. Tšekin puolelta löytyy kuitenkin vähän enemmän suulliseen vuorovaikutukseen liittyviä kollokaatteja sekä lähisyronyymejä (SOITTA – VOLAT/ZAVOLAT/TELEFONOVALAT). Viestintään liittyvät kollokaatit korostavat ystävien välistä toimintaa ja ymmärtämistä. Ihmiset ovat jatkuvasti keskenään vuorovaikutuksessa, mihin kuuluu viestien tuottaminen sekä vastaanottaminen, ja kommunikaatio on perusedellytys niin ystävyysuhteen luomiselle kuin ylläpitämisellekin.

HARRASTUS – suomi 14 % (32), tšekki 3 % (7)

'Harrastus' on toiseksi suurin lekseemin YSTÄVÄ kollokaattien merkitysryhmä 'ystävyyssuhde'-ryhmän jälkeen. Lekseemin PŘÍTEL ympäristössä harrastuksiin liittyvät kollokaatit taas esiintyvät melko harvoin. Tähän tulokseen viittaa jo edellinen merkitysanalyysi (ks. luku 4.2.2, kuvio 5), joka paljasti, että 'harrastaja' on suomen kielessä YSTÄVÄ-sanana toiseksi yleisin merkitys. Kollokaattien semanttisten ryhmien analyysi vahvistaa tämän löydön. On kuitenkin muistettava, ettei kollokaattien 'harrastus'-merkitysryhmä ole ekvivalentti YSTÄVÄ-lekseemin 'harrastaja'-merkitykseen. Vaikka suurin osa kyseessä olevista kollokaateista edustaa

nimenomaan 'harrastaja'-merkitystä (65), YSTÄVÄ perusmerkityksessäänkään ei ole poikkeuksellinen (66).

65) *Mä olen suuri kirjojen ystävä ja divarit olivat halvin tapa saada itselleen halvalla kirjoja.*

66) *Yritin etsiä heikkonäköiselle ikäiselleni ystävälle kirjaa.*

'Harrastus'-merkitysryhmä koostuu pienestä alaryhmästä kollokaatteja, jotka ilmaisevat itse harrastamista (HARRASTAA, HARRASTUS) sekä yhdistyksiä, jotka muodostuvat harrastusten ympärille (RY, YHDISTYS, PERUSTAA). Tähän kategoriaan kuuluvat myös kaikki tšekin kollokaatit (KONÍČEK [harrastus], KLUB [kerho]). Kaikki muut suomen kollokaatit harrastamistoiminnan ulkopuolella edustavat tiettyä konkreettista harrastusta ja näitä tšekin kielen kollokaateista ei löydy lainkaan. Tämä on merkittävä tulos, sillä tšekiksi voi sanaa PŘÍTEL käyttää myös 'harrastaja'-merkityksessä, mutta kollokaatioanalyysin perusteella se ei ole suomen tavoin tavallista kielenkäyttöä.

Harrastukset, joihin kytketään ilmaus YSTÄVÄ kattavat isohkon skaalan erilaisia toimintoja. Melkein puolet kollokaateista edustaa kulttuuria (JAZZ, TANGO, KIRJA, ELOKUVA). Sen lisäksi ryhmästä löytyy urheiluun (LIIKUNTA, JÄÄKIEKKO), luontoon (LUONTO, LINTU), ruokaan sekä juomaan (KALA, OLUT) liittyviä kollokaatteja. 'Harrastus'-ryhmään lajittelin myös adjektiiveja kuten PERINTEINEN ja KLASSINEN, joiden tapauksessa vasta konteksti selventää harrastukseen kuulumista (67).

67) *Maakunnan laskettelukeskuksista löytyy vaihtoehtoja niin perinteisen laskattelun ystäville kuin lautailijoillekin.*

Harrastuksiin, kannatukseen ja vapaa-aikaan viittaavat kollokaatit edustavat suomen kielessä kieltämättä merkittävää osaa YSTÄVÄ-lekseemin semanttisista ympäristöistä. Samanlaista havaintoa ei voi tehdä tšekin kielestä ja näin ollen kyseessä on yksi keskeisimmistä eroista suomen ja tšekin kielen välillä.

PARISUHDE – suomi 1 % (2), tšekki 7 % (16)

'Parisuhde'-merkitysryhmä on edellisen 'harrastus'-ryhmän vastakohta siinä mielessä, että tšekin kollokaatit muodostavat selvän enemmistön tämän kategorian kollokaateista, kun taas suomen kollokaatteja on hädän tuskin yhtään. Samaa taipumusta voidaan tämän lisäksi havainnoida merkitysryhmissä 'ammatti' ja 'väkivalta', joilla molemmilla on osittainen yhteys 'parisuhde'-ryhmään.

Suomen puolelta löytyy vain kaksi tätä ryhmää edustavaa kollokaattia (VAIMO, PUOLISO), jotka yhtä hyvin sopisivat myös 'ihmiset'-ryhmään. Tšekin aineistossa osa kollokaateista viittaa ihmisiin (MANŽEL [aviomies], MILENKA [rakastajatar], PARTNER [puoliso]) ja osa liittyy ihmisten vuorovaikutukseen (ROZEJÍT [erota], OŽENIT [mennä naimisiin], HÁDKA [riita]). Tutkielmassani on jo useamman kerran todettu, että 'puoliso'-merkitys on vahva tšekin lekseemillä PŘÍTEL. Kuten huomautettiin edellisen ryhmän kohdalla, kollokaattien semanttisen analyysin ja merkitysanalyysin välillä ei kuitenkaan ole suoraa yhteyttä. Semanttisen kollokaattianalyysin tarkoituksena on tutkia millaisessa semanttisessa ympäristössä valitut noodit ilmenevät ja siihen tietenkin vaikuttaa noodin tarkka merkitys, mutta kannattaa pitää mielessä, ettei tämä aina vastaa semanttista ryhmää. Näin ollen etenkin ihmisiin viittaavien kollokaattien konkordansseista suurin piirtein puolet heijastaa 'puoliso'-merkitystä (68) ja puolet sanan PŘÍTEL perusmerkitystä (69). Muiden kollokaattien osalta kyseessä on lähes pääsääntöisesti puoliso (70).

68) *Partnerem se rozumí manžel i manželka, přítel i přítelkyně.* [Puolisolla tarkoitetaan aviomiestä sekä aviovaimoa, miesystävää sekä naisystävää.]

69) *Největším mým přítelem je můj manžel, ten mi vždycky poradí [...].* [Mieheni on minun paras ystäväni, hän aina auttaa minua [...].]

70) *Po loňském rozchodu s přítelem žila sama v Lomnici v rodinném domku.* [Hän asui yksin Lomnicessa omakotitalossa viime vuoden (mies)ystävänsä eron jälkeen.]

Lopulta voi todeta, että 'parisuhde' on yksi semanttisista ryhmistä, joiden kollokaatit toistuvasti esiintyvät lekseemin PŘÍTEL ympärillä, kun taas suomen aineistossa tämä ryhmä on periaatteessa olematon. Syynä siihen voi olla lekseemien PŘÍTEL/YSTÄVÄ eroavat merkitykset, tai esimerkiksi suomen ja tšekin aineistojen poikkeavat aihealueet.

AMMATTI – suomi 1 % (2), tšekki 9 % (21)

Seuraavaan 'ammatti'-merkitysryhmään olen luokittelut kollokaateissa esiintyviä ammattinimikkeitä. Suomen kollokaatteja löytyy aineistosta vain muutama kappale, mutta tšekin kollokaateista ammattinimikkeet edustavat yhdeksää prosenttia. Erilaisten ammattien listalta huomaa heti niiden yhteisen tekijän olevan jonkinlainen julkisuus. Suurin osa tämän ryhmän kollokaateista edustaa artistisia ammatteja (SPISOVATEL [kirjailija], SOCHAŘ [kuvanveistäjä], ZPĚVAČKA [laulajatar]), mutta löytyy myös poliittisia edustajia (PREZIDENT [presidentti], PRINC [prinssi]).

'Ammatti'-ryhmän kollokaatit yleisesti sanoen esiintyvät kahdenlaisessa lausetyypissä. Joko ammattia edustava henkilö on jonkun ystävä (71), tai hänellä itsellään on ystävä (72).

71) *Při té příležitosti se Kate svěřila, že její přítel, slavný americký zpěvák a herec Johnny Depp, [...].* [Siinä tilanteessa Kate paljasti, että hänen (mies)ystävänsä, kuuluisa amerikkalainen laulaja ja näyttelijä Johnny Depp, [...].]

72) *Zpěvák Bono a jeho přátelé vždycky poslouchali rockové desky, které [...].* [Laulaja Bono ja hänen ystävänsä ovat aina kuunnelleet rock-levyjä, jotka [...].]

Koska tšekki on vahvasti sukupuolitettu kieli, on mielenkiintoista havainnoida eroa nais- ja miesammattinimikkeiden kontekstin sekä lauserakenteen välillä. Mikäli kyseinen ammattilainen on nainen (ZPĚVAČKA [laulajatar], HEREČKA [näyttelijätär]), hän on tyypillisemmin mainittu jonkun (nais)ystävänä, kuin että kyse olisi hänen ystävästään. Samanlaista taipumusta ei esiinny miesammattinimikkeissä, joiden tapauksessa molemmat lausetyypit ovat jakautuneet tasapuolisemmin. Miespuoliset henkilöt ovat olleet sanomalehtien artikkeleissa naisiin verrattuna yliedustettuja (GMMP 2015), mikä voisi ainakin osittain selittää tätä ilmiötä. Tarkempi tulkinta kuitenkin vaatisi kattavampaa sukupuolentutkimuksellista analyysiä.

Onko kuitenkin mahdollista selittää taitelijoiden ammattinimikkeiden yleisyyttä PŘÍTEL/KYNĚ -lekseemien kontekstissa ja näiden olemassaoloa suomen YSTÄVÄ-lekseemin ympärillä? Tähän on vaikea antaa perusteltua vastausta, joten esittelen muutaman hypoteesin. Yhtenä tekijänä voi olla erot suomen ja tšekin aineistojen välillä. Vaikka korpusaineisto on koottu tasapaino mielessä pitäen, ei sitä voi koskaan saavuttaa täydellisesti ja on mahdollista, että tšekin aineistossa esiintyy enemmän kulttuuria ja viihdettä käsitteleviä artikkeleja. Toinen vaihtoehto on se, että suomen korpuksen lehdissä syystä tai toisesta ei yhtä paljoa mainita taitelijoiden välisiä ystävyys-suhteita tai heidän ystävyyspiirejensä. Ammattinimikekollokaattien tšekin yliedustavuuden ja suomen aliedustavuuden tarkempi selittäminen ei laajuutensa vuoksi kuitenkaan kuulu tämän tutkielman piiriin.

RIKOS – suomi 0 % (1), tšekki 4 % (9)

'Rikos' on melko pieni ryhmä kollokaatteja, jotka jollakin tavoin liittyvät väkivaltaan tai rikokseen (VRAŽDA [murha], ZABÍT [tappaa], ZASTŘELIT [ampua kuoliaaksi]). Suomen kielestä tähän ryhmään on päättynyt ainoastaan kollokaatti KADOTA, joka pääsääntöisesti ilmenee *kadonnut ystävä* -ilmauksissa. Toisaalta katoaminen ei välttämättä liity rikokseen, toisin kuin tšekin kollokaattien tapauksissa, joista melkein kaikki liittyvät suorasti murhaan (73). Tämä on hieman yllättävää, koska murha ja väkivalta eivät ole asioita, joita odotettaisiin ilmenevän

PŘÍTEL-sanan ympäristössä. Kun tarkastelin 'rikos'-ryhmään kuuluvia kollokaatteja kontekstissaan, kävi ilmi, että yli puolessa näiden konkordansseista PŘÍTEL/KYNĚ -lekseemiä on käytetty 'puoliso'-merkityksessä. Väkivalta ja tappo eivät kieltämättä ole harvinainen ilmiö puolisojen eikä ystävien kesken (74).

73) *Muž se pokusil zavraždit svoji přítelkyni zřejmě ze žárlivosti.* [Mies yritti murhata (nais)ystävänsä todennäköisesti mustasukkaisuudesta.]

74) *Mladík zabil v hotelu svého přítele nožem.* [Nuori mies tappoi hotellissa ystävänsä veitsellä.]

Tämän merkitysryhmän kollokaattien poissaoloon suomen aineistossa voi vaikuttaa puolisoon liittyvien kollokaattien puuttuminen, tai mahdollisesti korpusaineston aihealueiden eroavuus tšekin ja suomen välillä.

TUNNE – suomi 4 % (10), tšekki 0 % (1)

Ryhmään nimensä mukaisesti kuuluvat kollokaatit, jotka ilmaisevat tunteita (ILAHDUTTAA, SURU, PELÄTÄ). Ystävien väliseen suhteeseen ja toimintaan kytkeytyvät tunteet, joten on hie-man yllättävää, ettei tšekin kollokaateista löydy niitä kuin yksi (POTĚŠIT [ilahduttaa]). Toisaalta tšekin aineistossa esiintyvät myös sanat MILOVAT [rakastaa] ja LÁSKA [rakkaus], jotka kuitenkin kontekstinsa perusteella sopivat enemmän 'parisuhde'-merkitysryhmään. 'Tunne'-ryhmässä suomen kielenkin osuus on kuitenkin niin pieni, että on hankalaa tehdä yleistäviä päätelmiä.

Tunnesanat ovat kielessä kuitenkin hyvin frekventtejä, joten tämä suomen ja tšekin aineiston epäsuhtaisuus herätti mielenkiintoni. Tarkastelin tšekin laajempaa kollokaattilistaa, josta löytyi useampia tilastollisesti merkitseviä ($I > 3,0$) tunnesanoja, mutta näiden I -arvo ei ollut yhtä korkea kuin suomen kollokaateilla. Tätä voi tulkita kahdella tavalla: joko YSTÄVÄ-lekseemi kollokoi tunnesanojen kanssa vahvemmin kuin PŘÍTEL-lekseemi, tai tšekin aineistossa ilmenee enemmän muita voimakkaita kollokaatiosuhteita.

ADVERBIAALIT – suomi 10 % (24), tšekki 6 % (14)

'Adverbiaalit' ovat kieliopillinen ryhmä, ei niinkään merkitysryhmä. Jaottelen ne edelleen määrään, paikkaan ja aikaan viittaaviksi kollokaateiksi, ja näistä kolmesta suurin on molemmissa kielissä määräkategoria. Siihen kuuluu lukumäärää ilmaisevia adverbeja, adjektiiveja ja pronomineja (PALJON, LUKUISA, MONI) sekä substantiiveja (JOUKKO, SKUPINKA [ryhmä],

HRSTKA [kourallinen]). Kontekstissaan määrän adverbiaalit ilmenevät pääsääntöisesti lekseemien YSTÄVÄ/PRÍTEL määritteenä (75) tarkentamassa montako ystävää on kyseessä.

75) *Lisäksi he ovat saavuttaneet runsaasti ystäviä urallaan, Cowie kirjoittaa.*

Enemmistö suomen paikka-adverbiaaleista on lokatiivisia proadverbeja (TÄNNE, SIELLÄ), mutta esiintyy myös asumiseen viittavia sanoja (KOTONA, ASUA sananmuodossa *asuva*). Muutama tähän alaryhmään kuuluva tšekin kollokaatti viittaa hieman tarkempaan paikkaan tai liikkeeseen (CHATA [mökki], VÝLET [reissu]) kuten esimerkkilauseessa 76.

76) *Letos pojedeme k přátelům na chatu, kterou mají na Slapech.* [Tänä vuonna lähdemme ystävien mökille, joka sijaitsee Slapylla.]

Näin ollen, paikka-adverbiaalit osoittavat aineistossani liikkeeseen, yhdessä olemiseen ja asumiseen. Tämän lisäksi kollokaateista löytyy pienehkö ryhmä aikaa ilmaisevia adverbiaaleja (USEIN, JOSKUS, KDYSI [aikoinaan]).

KIELIOPPI – suomi 6 % (15), tšekki 2 % (4)

Kieliopilliset sanat muodostavat suhteellisen pienen ryhmän. Suomen kielessä osaa niistä edustavat persoonapronominit, jotka kontekstissa ilmenevät genetiivimuodossa (*minun, sinun, hänen*) ja ilmaisevat kenen ystävä on kyseessä. Tšekin kollokaateista vastaavasti löytyy possessiivipronomineja (MŮJ [minun], TVŮJ [sinun], SVŮJ [oma]). Lekseemin YSTÄVÄ kollokaattilistalla kuitenkin esiintyy myös muita pronomineja ja partikkeleja (JA, KUKA, JOKU, PAITSI).

MI-testi, jolla laskettiin tilastollisesti merkitsevät kollokaatit, suosii luonteeltaan sisältösanoja, mutta mitä pidemmälle listalla edetään, sitä useammin alkaa ilmetä myös kieliopillisiä sanoja. Suomen kollokaateista löytyy kieliopillisiä sanoja enemmän kuin tšekin kollokaateista, mikä voi johtua suomen pienemmästä korpusaineistosta. Toisaalta on myös mahdollista, että tšekin PRÍTEL vetää puoleensa enemmän sisältökollokaatteja kuin suomen YSTÄVÄ. Tätä on kuitenkin hankalaa todistaa sen vuoksi, että kummankin kielen kollokaatit on hankittu erilaisista aineistosta.

5 PÄÄTÄNTÖ

Maisterintutkielmassani tutkin suomen lekseemin YSTÄVÄ ja tšekin lekseemin PŘÍTEL merkityksiä sekä kontekstuaalista valintaa keskittyen kollokaatteihin. Tutkimuksen tasapainon vuoksi sisällytin analyysiin myös feminiinijohdokset YSTÄVÄTÄR ja PŘÍTELKYNĚ. Merkitys- ja kollokaatioanalyysin edellytyksenä tarkastelin lekseemien käänösvastaavuutta. Pääaineistonani käytin 90-luvun journalistisista teksteistä koostuvaa verrannollista korpusta, jonka suomen puolelta toimi Kielipankin Suomen kielen tekstikokoelma ja tšekin osalta Český národní korpus -hankeen SYNPUB2006-korpus.

Tutkielman alussa asetin kolme tutkimuskysymystä, joihin vastaavat kolme analyysin päälukua. Ensimmäinen analysoin, kuinka usein lekseemit YSTÄVÄ/PŘÍTEL ovat toistensa käänösvastineina rinnakkaiskorpuksessa InterCorp 11. Luku osoittautui olevan suurin piirtein 85 %, jota pidin tarpeeksi korkeana jälkimmäistä analyysia varten. Sen lisäksi mielenkiintoiseksi havainnoksi paljastui läheisempi yhteys lekseemin YSTÄVÄ ja tšekin puhekielisemmän KAMARÁD-lähisynonymien välillä, kuin vastaavan sanaparin PŘÍTEL ja KAVERI välillä.

Seuraavan analyysin painopisteenä oli selvittää, millaisia merkityksiä valituilla lekseemeillä esiintyy sekä sanakirjojen että verrannollisen korpuksen avulla, josta tutkin 500 konkordanssia kummankin kielen puolelta. Tarkastelusta kävi ilmi, että molemmat YSTÄVÄ/PŘÍTEL - sanat vastaavat hyvin toisiaan niiden perusmerkityksessään, joka on 'läheinen ihminen' (n. 60 %). Toiseksi yleisimmässä merkityksessä lekseemit jo eroavat täysin. Suomen aineistossa tämä osoittautui olevan jonkin 'harrastaja, kannattaja' (n. 30 %). Myös tšekin konkordansseissa kyseinen merkitys esiintyi, mutta pääsääntöisesti vain yhdistysten ja järjestöjen nimien osana. Tšekin PŘÍTEL-lekseemin toiseksi yleisin merkitys oli 'puoliso' (n. 23 %), joka puolestaan esiintyy suomen aineistossa todella harvoin. Yllätyksenä ilmeni, että 'puoliso'-merkitystä havaitaan suurimmassa osassa feminiinijohdoksen PŘÍTELKYNĚ esiintymistä. Lopulta kävi ilmi, että tarkasteltavien lekseemien tapauksessa, sanakirjojen merkitysten määritelmät heijastavat korpuksen havaintoja melko hyvin.

Tutkielmani keskipiste oli lekseemien YSTÄVÄ/PŘÍTEL sekä kvantitatiivisessa että kvalitatiivisessa kollokaatioanalyysissä. Hankin verrannollisesta korpuksesta *MI*-testin avulla 250 tilastollisesti merkitsevää kollokaattia. Näiden joukossa esiintyi jonkin verran ihmisten erisnimiä sekä järjestöjen nimiin liittyvät sanat. Molemmat alaryhmät jätin pois jälkimmäisestä semanttisesta analyysistä. Kollokaateista muodostin yhteensä 13 merkitysryhmää, joiden osalta löytyi kielten välillä sekä yhdistäviä että eroavia piirteitä. Ryhmät, joiden aiheena oli ystävien

välinen suhde ja vuorovaikutus ('ystävyyssuhde', 'tapaaminen', 'ihmiset', 'viestintä' ja 'tapahuma') näyttäytyivät varsin samankaltaisiksi sekä niiden suuruuden että niihin sisältyvien kollokaattien osalta. Yhtäläisyys oli silmiinpistävä, sillä kummankin kielen ryhmissä esiintyi iso määrä toisen kielen käänkösvastineita. 'Harrastus'-merkitysryhmän prosentiosuudessa havaittiin suurin ero suomen ja tšekin välillä, suomen kielessä osuuden ollen suurempi. Harrastamisen kohteiksi huomattiin pääosin kulttuuriin liittyviä kollokaatteja. Toista huomattavaa eroa edustaa 'parisuhde'-merkitysryhmä, jonka osalta löytyi odotetusti enemmän tšekin kollokaatteja. Vastaavanlainen eroavuus ilmeni myös 'ammatti'- ja 'rikos'-ryhmissä, jotka molemmat osittain liittyivät ihmisten parisuhteeseen. Viimeinen mielenkiintoinen havainto nousi esille kahdessa pikemminkin kieliopillisessa ryhmässä, joissa molemmissa oli yliedustettuina YSTÄVÄ-lekseemin kollokaatit. Voi todeta, että kollokaattien semanttisesta analyysistä johtuneet merkittävimmät tulokset ovat sovussa valittujen lekseemien merkitysjakauman kanssa. Kollokaatioanalyysissä kuitenkin tutkittiin muiden muassa merkityseroja tarkemmin ja nostettiin esille hienovaraisempiakin eroja kielten välillä.

Tutkimuksen tekemisen prosessissa hankalin vaihe oli aineiston käsittely, joka johtui kahden eri ohjelman käytöstä samalla yrittäen saavuttaa vertailukelpoista dataa. Lopulta tämä kuitenkin onnistui ja kollokaatit oli laskettu mahdollisimman läpinäkyvällä tavalla. Kahdesta itsenäisestä korpuksesta kokoamani aineiston vertailumahdollisuuksia tuli vastaan myös analyysivaiheessa. Miten voi olla varma, että YSTÄVÄ/PŘÍTEL -sanojen kollokaattien erot johtuvat lekseemien luonteista, eikä aineistojen erilaisuuksista? Mikäli kyseiset erot osoittautuvat tarpeeksi merkittäväksi, on mielestäni turvallista todeta, että havainnot heijastavat lekseemien eroavaa kontekstuaalista valintaa. Mikäli eroavuudet ovat pikemminkin vähäisiä, syynä todellakin voi olla esim. aineistojen eri aihealueet, korpusten erikokoisuus tai jokin toinen muuttuja. Vertaillen kahden kielen lekseemejä ei luonnollisesti voi käyttää samaa aineistoa ja verrannollinen korpusten laatiminen on aina jonkinlainen kompromissi. Suomen ja tšekin korpusten osat vastasivat toisiaan genressä, kielen muodollisuudessa ja aikakaudessa. Ne toisaalta eroavat huomattavasti koossaan ja tietysti myös sanomalehtikokoelmiensa sisällössä. Se, kuinka suuri vaikutus aineistojen eroavilla tekijöillä oikeasti on, voisi selkiytyä esimerkiksi vastaavassa tutkimuksessa käyttäen toisenlaista aineistoa ja tarkastelemalla löytyykö kollokaattien välillä samanlaisia eroavuuksia.

Aineisto on reilut 25 vuotta vanha, minkä ei pitäisi olla ongelma sanomalehtien kielen tapauksessa, sillä se kuitenkin muuttuu melko hitaasti. Toisaalta lekseemien YSTÄVÄ/PŘÍTEL merkitysanalyysistä ilmeni, että tšekin vastineessa tapahtui todennäköisesti viime vuosisadan toisella puolella semanttinen muutos, jolla sen merkitys laajentui tarkoittaen läheisen ihmisen

lisäksi myös puolisoa. Olisi mielenkiintoista tarkastella, olisiko tuoreemmissa teksteissä tämä muutos vielä selkeämpi.

Tilastollisen menetelmän osalta valitsin muuttujat kuten tilastollinen testi, sen arvon minimi, minimiesiintymisfrekvenssi ja kollokaatioväli. Nämä kaikki vaikuttavat siihen, kuinka moni kollokaatti nousi aineistosta merkitseväksi, ja millaisia ne ovat. Minimiesiintymisfrekvenssi ja kollokaatioväli olivat mielestäni sopivasti valittuja, ja ne hyvin vastasivat korpuksen kokoa ja kielten tyypillisiä kieliopillisia rakenteita. *MI*-testi myös vastasi hyvin tutkielmani tarpeisiin, mutta sen minimiarvo (3,0) olisi voinut olla korkeampi. Vielä 250. kollokaatin *I*-arvo oli tšekissä 4,6 ja suomessa 3,6. Korkeamman arvon määrittelemällä olisi voinut välttää kieliopillisten sanojen liiallista esiintymistä.

Tutkielmani viimeinen analyysin osa – kollokaattien semanttinen tarkastelu – oli luonteeltaan subjektiivinen ja näin ollen tutkijan omaan kielitajuun, intuitioon ja muihin kielen ja maailman käsityksiin nojautuva. Pyrin kuitenkin mahdollisimman objektiiviseen tarkasteluun ja valintojeni perusteluun. Silti on mahdollista, että erilaisia näkökulmia hyödyntävä tutkija voisi päästä jonkin verran eri tuloksiin.

Tutkielmani kytkeytyy laajempaan fraseologiseen viitekehykseen ja etenkin kontekstuaaliseen valintaan sekä kollokaatiotutkimukseen. Samalla sen pohjana on kielten vertaileva tutkimus, ja näiden kahden yhdistelmää näkee alan tutkimuksissa melko harvoin. On aina hyödyllistä tutkia asioita, jotka ovat muutenkin hieman sivuutettuja. Ehkäpä tutkielmani voisi toimia pohjana samankaltaiselle tutkimukselle, jolla samalla voisi edistää vertailevassa fraseologisessa tutkimuksessa käytettäviä menetelmiä. Tutkielmastani voi lisäksi olla hyötyä vieraan kielen opetuksessa, sillä se osoittaa, että kahteen eri kieliryhmään kuuluvalla sanalla voi olla laajoja merkityseroja, vaikka nämä ovat toistensa käännösvastineita. Voisi kuvitella, että monella muulla sanalla erot ovat vielä paljon huomattavampia. Kollokaatteihin keskittyminen vieraan kielen opetuksessa on edelleenkin varhaisessa vaiheessa, vaikka nimenomaan oikea kollokaattien käyttö edistää luonnolliselta vaikuttavaa kielen tuotantoa. Sillä tutkimuksessani käsitellään myös YSTÄVÄ/PRÍTEL -sanojen käännöksiä, tutkielmani aihe voisi olla kiinnostava käännöstieteen tutkijoille.

Jatkotutkimuksen mahdollisuuksina voisi miettiä lekseemien YSTÄVÄ/PRÍTEL analyysiä käyttäen eri genrestä koostuvaa korpusta, esim. puhutun kielen korpusta. Olisi kiinnostavaa tutkia, miten lekseemien kontekstuaalinen valinta vaihtelee riippuen kielen muodosta. Suomen ja tšekin lähisyronyymien vertaileva fraseologinen tutkimus on erittäin harvinaista, joten tarjolla on suuri määrä tutkittavia sanoja. Tähän tutkielmaan liittyen voisi lisäksi vertailla muita

läheisiin ihmisiin viittaavia sanoja. Esimerkiksi jo useamman kerran mainittu sanapari KAVERI/KAMARÁD tai kenties puolisoa tarkoittavia sanoja.

LÄHTEET

- Altenberg, Bengt 1999: Adverbial connectors in English and Swedish: Semantic and lexical correspondences. – Hilde Hasselgård & Signe Oksefjell (toim.), *Out of corpora: Studies in honour of Stig Johansson* s. 249–268. Amsterdam: Rodopi.
- Altenberg, Bengt & Granger, Sylviane 2002: Recent trends in cross-linguistic lexical studies. – Granger Sylviane & Bengt Altenberg (toim.), *Lexis in contrast: Corpus-based approaches* s. 3–51. Philadelphia/Amsterdam: John Benjamins.
- Arppe, Antti 2008: *Univariate, bivariate, and multivariate methods in corpus-based lexicography – a study of synonymy*. Publications of the Department of General Linguistics, University of Helsinki 44. Helsinki: University of Helsinki.
- Aston, Guy 2011: Applied Corpus Linguistics and the learning experience. – Vander Viana, Sonia Zyngier & Geoff Barnbrook (toim.), *Perspectives on Corpus Linguistics* s. 1–16. Amsterdam: John Benjamins.
- Baker, Mona 1993: Corpus linguistics and translation studies. Implications and applications. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology: in honour of John Sinclair* s. 233–250. Amsterdam: John Benjamins.
- 1998: *Routledge encyclopedia of translation studies*. London: Routledge.
- Barnbrook, Geoff 1996: *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- 2009: Sinclair on collocation. – Rosamund Moon (toim.), *Words, grammar, text: Revisiting the work of John Sinclair* s. 23–39. Philadelphia/Amsterdam: John Benjamins.
- Bečka, Josef Václav 1979: *Slovník synonym a frazeologizmů*. Praha: Novinář.
- Bednarek, Monika 2008: Semantic preference and semantic prosody re-examined. – *Corpus Linguistics and Linguistic Theory* 4 (2) s. 119–139.
- Bezdičková, Alžběta 2014: Portál věnovaný českému jazyku: www.nechybujte.cz (Lingea). – *Jazykovědné aktuality* 2014 (1–2) s. 35–44.
- Biber, Douglas, Conrad, Susan & Reppen, Randi 1998: *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- Blank, Andreas 2003: Polysemy in the lexicon and in discourse. – Brigitte Nerlich, Zazie Todd, Vimala Herman & David D. Clarke (toim.), *Polysemy: Flexible patterns of meaning in mind and language* s. 267–293. Berlin: de Gruyter.
- Carter, Ronald 2003: Introduction. – John Sinclair, *Trust the text. Language, corpus and discourse* s. 1–6. London: Routledge.
- Clear, Jeremy 1993: From Firth principles. Computational tools for the study of collocation. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology: In honour of John Sinclair* s. 271–292. Philadelphia: John Benjamins.
- Colson, Jean-Pierre 2008: Cross-linguistic phraseological studies. An overview. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. 191–207. Philadelphia/Amsterdam: John Benjamins.
- Conrad, Susan 2011: Variation in corpora and its pedagogical implications. – Vander Viana, Sonia Zyngier & Geoff Barnbrook (toim.), *Perspectives on Corpus Linguistics* s. 47–62. Amsterdam: John Benjamins.
- Cruse, Alan 1986: *Lexical semantics*. Cambridge: Cambridge University Press.
- 2000 [2004]: *Meaning in language: An introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- CSCb = Suomen sanomalehtikielen taajuussanasto. Kielipankki, CSC – Tieteen tietotekniikan keskus Oy. – <https://korp.csc.fi/suomen-sanomalehtikielen-taajuussanasto-B9996.txt> Viitattu 28.2.2018.

- Čermák, František 2004: *Frekvenční slovník češtiny*. Praha: Nakladatelství lidové noviny.
- 2007: *Jazyk a jazykověda*. Praha: Karolinum.
- Čermáková, Anna 2013: Jaký slovník uživatelé češtiny potřebují? O Slovníku současné češtiny nakladatelství Lingea. – *Slovo a slovesnost* 74 (3) s. 195–211.
- Ebeling Jarle & Ebeling, Signe Oksefjell 2013: *Patterns in contrast*. Amsterdam: John Benjamins.
- Ellis, Nick C. 2008: Phraseology: The periphery and the heart of language. – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 1–15. Amsterdam: John Benjamins.
- Firth, J. R. 1957: A synopsis of linguistic theory 1930–55. – F. R. Palmer (toim.), *Selected papers of J. R. Firth 1952–59* s. 168–205. London: Longmans.
- François, Alexandre 2008: Semantic maps and the typology of colexification. Intertwining polysemous networks across languages. – Martine Vanhove (toim.), *From polysemy to semantic change: towards a typology of lexical semantic associations* s. 163–217. Amsterdam: John Benjamins.
- GMMP = The Global Media Monitoring Project 2015. Saatavilla https://whomakesthe-news.org/wp-content/uploads/who-makes-the-news/Imported/reports_2015/global/gmmp_global_report_en.pdf Viitattu 12.7.2021.
- Granger, Sylviane & Meunier, Fanny 2008: Introduction. The many faces of phraseology. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. xix – xxviii. Philadelphia/Amsterdam: John Benjamins.
- Gries, Stefan Th. 2008: Phraseology and linguistic theory: A brief survey. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. 3–27. Philadelphia/Amsterdam: John Benjamins.
- 2011: Methodological and interdisciplinary stance in Corpus Linguistics. – Vander Viana, Sonia Zyngier & Geoff Barnbrook (toim.), *Perspectives on Corpus Linguistics* s. 81–98. Amsterdam: John Benjamins.
- Haarala, Risto 1994: Suomen kielen perussanakirja – Nykysuomen sanakirjan perillinen. *Kielikello* (3).
- Hoey, Michael 2005: *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- 2007: Lexical Priming and literary creativity. – Michael Hoey, Michaela Mahlberg, Michael Stubbs & Wolfgang Teubert (toim.), *Text, discourse and corpora* s. 7–30. New York: Continuum.
- Hundt, Marianne 2008: Text corpora. – Anke Lüdeling & Merja Kytö (toim.), *Corpus Linguistics. An International Handbook. Volume 1* s. 168–187. Berlin: Walter de Gruyter.
- Inkpen, Diana & Graeme, Hirst 2006: Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics* 32 (2) s. 223–262.
- Itkonen, Erkki 1992: *Suomen sanojen alkuperä: etymologinen sanakirja*. 1, A–K. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Itkonen, Esa 2001: *Maaailman kielten erilaisuus ja samuus*. Turku: Turun yliopisto.
- Jantunen, Jarmo 2001: ”Tärkeä seikka” ja ”keskeinen kysymys”: Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? – *Virittäjä* 105 (2) s. 170–192.
- 2004: *Synonyymia ja käännessuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännskielen leksikaalisiin erityispiirteisiin*. Väitöskirja. Joensuun yliopiston humanistisia julkaisuja 35.
- 2011: Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttajat ja annotointi. – *Lähivõrdlusi = Lähivertailuja* 21 s. 86–105.

- 2016: Corpora, phraseology and dictionaries: How does corpus research intersect language teaching and learning? – Begoña S. Vilas (toim.), *Collocations Cross-Linguistically*. Turenki: Hansaprint.
- Johansson, Stig 2007: *Seeing through multilingual corpora: On the use of corpora in contrastive studies*. Amsterdam: John Benjamins.
- 2008: Some aspects of the development of corpus linguistics in the 1970s and 1980s. – Anke Lüdeling & Merja Kytö (toim.), *Corpus Linguistics. An International Handbook. Volume 1* s. 33–53. Berlin: Walter de Gruyter.
- 2011: A multilingual outlook of corpora studies. – Vander Viana, Sonia Zyngier & Geoff Barnbrook (toim.), *Perspectives on Corpus Linguistics* s. 115–130. Amsterdam: John Benjamins.
- Kangas, Akseli 2018: *Adverbien täysin ja kokonaan kollokaatit ja semanttinen prosodia*. Maisterintutkielma. Jyväskylän yliopiston kieli- ja viestintätieteiden laitos.
- Kangasniemi, Heikki 1997: *Sana, merkitys, maailma. Katsaus leksikaalisen semantiikan perusteisiin*. Helsinki: Finn Lectura.
- Karlsson, Fred 2004 [2006]: *Yleinen kielitiede*. Helsinki: Yliopistopaino.
- Kennedy, Graeme 2008: Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology in Foreign Language Learning and Teaching* s. 21–42. Amsterdam: John Benjamins.
- Kenny, Dorothy 2001: *Lexis and creativity in translation*. Manchester: St. Jerome Publishing.
- Kielipankki: *Mikä on kielipankki?* – <https://www.kielipankki.fi/> Viitattu 19.1.2019.
- Koivusalo, Esko 1992: Sanomalehti ja äidinkieli. – *Kielikello* (2).
- Kuiri, Kaija 2012: *Johdatus semantiikkaan*. Helsinki: Finn Lectura.
- Laakkonen, Maria 2017: *Lähisynonyymien keho, vartalo ja ruumis kontekstuaalista semantiikkaa*. Pro gradu -tutkielma. Oulun yliopiston suomen kielen oppiaine.
- Leech, Geoffrey 1981: *Semantics. The study of meaning*. Harmondsworth: Penguin Books.
- 2011: Principles and applications of Corpus Linguistics. – Vander Viana, Sonia Zyngier & Geoff Barnbrook (toim.), *Perspectives on Corpus Linguistics* s. 155–170. Amsterdam: John Benjamins.
- Leino, Antti & Leino, Pirkko 1990: *Synonyymi-sanasto*. Keuruu: Otava.
- Lyons, John 1968: *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- 1995: *Linguistic semantics. An introduction*. Cambridge: Cambridge University Press.
- Machek, Václav 1968: *Etymologický slovník jazyka českého*. Praha: Nakladatelství Lidové noviny.
- Martin, Willy 2008: A unified approach to semantic frames and collocational patterns. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspectives* s. 51–67. Philadelphia/Amsterdam: John Benjamins.
- Mauranen, Anna 2002: Will 'translationese' ruin a contrastive study? – *Languages in Contrast* 2 (2) s. 161–186.
- Mauranen, Anna & Kujamäki, Pekka 2004: *Translation universals. Do they exist?* Amsterdam: John Benjamins.
- McEnery, Tony & Hardie, Andrew 2012: *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, Tony & Xiao, Richard 2008: Parallel and comparable corpora. What is happening? – Gunilla M. Anderman & Margaret Rogers (toim.), *Incorporating corpora. The linguist and the translator* s. 18–31. Buffalo: Multilingual Matters.
- McEnery, Tony, Xiao, Richard & Tono, Yukio 2006: *Corpus-based language studies: an advanced resource book*. London: Routledge.

- Mel'čuk, Igor A. 1998: Collocations and lexical functions. – Anthony Paul Cowie (toim.), *Phraseology. Theory, analysis and applications* s. 25–53. Oxford: Clarendon Press.
- Moon, Rosamund 1998: *Fixed expressions and idioms in English. A corpus-based approach*. Oxford: Clarendon Press.
- Morris, Charles W. 1938: *Foundations of the theory of signs*. Chicago: The University of Chicago Press.
- Mäkinen, Kirsti 1992: Muuttuuko lehtikieli? Muuttaako lehtikieli? – Valma Yli-Vakkuri, Maija Länsimäki & Aarre Nyman (toim.), *Yhteiskunta muuttuu. Kieli muuttuu* s. 211–235. Porvoo: WSOY.
- Nesselhauf, Nadja 2003: The use of collocations by advanced learners of English and some implications for teaching. – *Applied Linguistics* 24 (2) s. 223–242.
- 2005: *Collocation in a learner corpus*. Studies in corpus linguistics 14. Amsterdam: John Benjamins.
- Partington, Alan 1998: *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam/Philadelphia: John Benjamins.
- Piirainen, Elisabeth 2008: Figurative phraseology and culture. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspectives* s. 207–228. Philadelphia/Amsterdam: John Benjamins.
- Pirkola, Silja 2016: *Synonyymien ehkä ja mahdollisesti kollokaatit ja semanttiset preferenssit*. Maisterintutkielma. Jyväskylän yliopiston kielten laitos.
- Příručka ČNK = Václav Cvrček & Olga Richterová (toim.): cnk:intercorp. 30.1.2018 Saata-
villa <http://wiki.korpus.cz/doku.php?id=cnk:intercorp&rev=1517320887> Viitattu
19.3.2018.
- Ravin, Yael & Leacock, Claudia 2002: Polysemy: An Overview. – Yael Ravin & Claudia Leacock (toim.), *Polysemy. Theoretical and Computational Approaches* s. 1–29. Oxford: Oxford University Press.
- Robert, Stéphane 2008: Words and their meanings. Principles of variation and stabilization. – Martine Vanhove (toim.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* s. 55–92. Amsterdam: John Benjamins.
- Saeed, John I. 2011: *Semantics*. Malden, Mass: Wiley-Blackwell.
- Salminen, Saara 2008: *A corpus-based study on English and Swedish near-synonyms. The case of Environment, Circumstances and Surroundings, and Miljö, Omständigheter and Omgivning*. Pro gradu -tutkielma. Tampereen yliopiston kieli- ja käänntieteidien laitos.
- Saussure, Ferdinand de 1959 [1916]: *Course in general linguistics*. New York: Philosophical Library.
- Simon-Vandenberg, Anne-Marie 2013: English adverbs of essence and their equivalents in Dutch and French. – Karin Aijmer & Bengt Altenberg (toim.), *Advances in corpus-based contrastive linguistics: Studies in honour of Stig Johansson* s. 83–102. Philadelphia/Amsterdam: John Benjamins.
- Sinclair, John 1991: *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- 1996: The search for units of meaning. – *Textus* IX s. 75–106.
- 1998: The lexical item. – Edda Weigand (toim.), *Contrastive lexical semantics* s. 1–24. Amsterdam: John Benjamins.
- 2004: *Trust the text. Language, corpus and discourse*. London: Routledge.
- Jones, Susan, Daley, Robert & Krishnamurthy, Ramesh 2004: *English collocation studies: the OSTI report*. London; New York: Continuum.
- 2008: Preface. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An interdisciplinary perspective* s. xv–xviii. Philadelphia/Amsterdam: John Benjamins.
- Smejkalová, Kamila 2014: Přítel, nebo kamarád? – *Literární noviny*.

- Souček, Milan 2008: *Kontrastivní studie substantiv vyjadřujících čas ve finštině a češtině*. Pro gradu tutkielma. Prahan Kaarlen yliopiston lingvistiikan ja fennougristiikan laitoksessa.
- Stubbs, Michael 1993: British traditions in text analysis. From Firth to Sinclair. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology: In honour of John Sinclair* s. 1–32. Philadelphia/Amsterdam: John Benjamins.
- 1995: Collocations and semantic profiles: on the cause of the trouble with quantitative studies. – *Functions of Language* 2 (1) s. 23–55. – Uudelleen julkaistu osoitteessa <https://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf> s. 1–24. Viitattu 22.3.2019.
- 1996: *Text and corpus analysis: computer assisted studies of language and culture*. Oxford: Blackwell.
- 2001 *Words and phrases*. Corpus studies of lexical semantics. Oxford: Blackwell Publishing.
- 2007: On texts, corpora and models of language. – Michael Hoey, Michaela Mahlberg, Michael Stubbs & Wolfgang Teubert (toim.), *Text, discourse and corpora* s. 127–161. New York: Continuum.
- 2013: Sequence and order: The neo-Firthian tradition of corpus semantics. – Hilde Hasselgård, Jarle Ebeling & Signe Oksefjell Ebeling (toim.), *Corpus perspectives on patterns of lexis* s. 13–35. Amsterdam/Philadelphia: John Benjamins.
- Sweetser, Eve 1990: *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- SYN2005 taajuussanasto = *Český národní korpus: Srovnávací frekvenční seznamy*. Ústav Českého národního korpusu FF UK, Praha 2010. Saatavilla <http://ucnk.ff.cuni.cz/srovnani10.php> Viitattu 28.2.2018.
- Tognini-Bonelli, Elena 2001: *Corpus linguistics at work*. Amsterdam: John Benjamins.
- 2002: Functionally complete units of meaning across English and Italian. Towards a corpus-driven approach. – Bengt Altenberg & Sylviane Granger (toim.), *Lexis in contrast* s. 73–96. Amsterdam: John Benjamins.
- ÚČNK: *Profil ústavu*. – <https://ucnk.ff.cuni.cz/cs/ustav/profil-ustavu/> Viitattu 21.1.2019.
- Vanderveken, Daniel 2004: Success, satisfaction, and truth in the logic of speech acts and formal semantics. – Steven Davis & Brendan S. Gillon (toim.), *Semantics. A reader* s. 710–734. New York: Oxford University Press.
- Viaggio, Sergio 2008: Semantic and communicative translation: Two approaches, one method. – Mildred L. Larson (toim.), *Translation: theory and practice, tension and interdependence* s. 172–187. Amsterdam: John Benjamins.
- Viana, Vander, Sonia, Zyngier & Barnbrook, Geoff (toim.) 2011: *Perspectives on Corpus Linguistics*. Amsterdam: John Benjamins.
- VISK = Auli Hakulinen, Maria Vilkkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. – <http://scripta.kotus.fi/visk> Viitattu 7.3.2018.
- Wierzbicka, Anna 1996: *Semantics: Primes and Universals*. New York: Oxford University Press.
- 1997: *Understanding cultures through their key words: English, Russian, Polish, German, and Japanese*. New York: Oxford University Press.
- Wray, Alison 2002: *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Xiao, Richard & McEnery, Tony 2006: Collocation, semantic prosody, and near synonymy. A crosslinguistic perspective. – *Applied Linguistics* 27 (1) s. 103–129.
- Yarowsky, David 1993: One sense per collocation. – *Proceedings of the workshop on human language technology* s. 266–271. Princeton.

Yavorska, Galina & Zymovets, Galina 2016: Motivational scenarios and semantic frames for social relations in Slavic, Romance and Germanic languages – friends, enemies, and others. – Päivi Juvonen & Maria Koptjevskaja-Tamm (toim.), *The lexical typology of semantic shift* s. 335–354. Berlin: de Gruyter.

Sanakirjat

FČ = Finsko-český česko-finský šikovný slovník, 2016, Lingea.

KS = Kielitoimiston sanakirja. 2021. Helsinki: Kotimaisten kielten keskus. URN:NBN:fi:kotus-201433, Verkkojulkaisu www.kielitoimistonsanakirja.fi. Päivitetty 11.11.2021 Viitattu 20.11.2021.

NS = Nykysuomen sanakirja. Helsinki: WSOY 1951–1961.

SSJČ = Slovník spisovného jazyka českého 1960–1971, Bohumil Havránek (toim.), Ústavu pro jazyk český ČSAV. Praha: Academia.

SSoučČ = Slovník současné češtiny 2011, Lingea. Verkkojulkaisu www.nechybujte.cz/ Viitattu 10.3.2018.

ST = Lindroos-Čermáková, H. 1984: Suomi–Tšekki–Suomi taskusanakirja. Porvoo: WSOY.

Korpukset

CSC – Tieteen tietotekniikan keskus (1998). Suomen kielen tekstikokoelma. Kielipankki. – <http://urn.fi/urn:nbn:fi:lb-201403268> Viitattu 24.11.2017

Ústav českého národního korpusu FF UK Praha (2006). SYN2006PUB: korpus psané publicistiky. Český národní korpus. – <http://www.korpus.cz> Viitattu 24.11.2017.

Ústav českého národního korpusu FF UK Praha (2005). InterCorp11. Český národní korpus. – <http://korpus.cz> Viitattu 20.10.2019.

LIITTEET

LIITE 1

Lekseemien YSTÄVÄ ja YSTÄVÄTÄR *MI*-testin nostamat kollokaatit

(ryhmät: 1. yhdistyksiin liittyvät kollokaatit 2. ihmisten erisnimet 3. luonnolliset kollokaatit)

#	lemma	frekvenssi	I-arvo	ryhmä
1	vannoutunut	45	9,944	3
2	sukulainen	543	9,900	3
3	raittius	38	9,749	1
4	tuttava	268	9,616	3
5	työtoveri	153	9,299	3
6	tunnustautua	72	9,271	3
7	Holappa	30	8,593	2
8	kansantanssi	36	8,380	1
9	uskollinen	90	8,231	3
10	läheinen	317	8,163	3
11	käsityö	112	8,076	1
12	rakas	110	7,918	3
13	vihollinen	105	7,821	3
14	omainen	157	7,628	3
15	perheenjäsen	37	7,501	3
16	luokse	42	7,483	3
17	runous	40	7,340	3
18	merkkipäivä	31	7,307	3
19	luona	74	7,289	3
20	toveri	39	7,274	3
21	edesmennä	38	7,169	3
22	tervehtiä	41	7,126	3
23	ry	232	7,110	3
24	tukija	48	7,095	3
25	ilahduttaa	46	7,076	3
26	kollega	50	6,907	3
27	isänmaa	36	6,812	3
28	syntymäpäivä	33	6,792	3
29	lahja	88	6,758	3
30	yksinäinen	50	6,694	3
31	olut	77	6,665	3
32	innokas	60	6,607	3
33	jutella	31	6,494	3
34	tervetullut	52	6,479	3
35	lähin	48	6,454	3

36	toivottaa	39	6,431	3
37	naapuri	135	6,410	3
38	tavata	275	6,399	3
39	viini	71	6,363	3
40	pitkäaikainen	68	6,233	3
41	harrastus	94	6,198	3
42	suru	32	6,097	3
43	runo	94	6,093	1
44	luotettava	32	6,091	3
45	ylin	46	6,090	3
46	perhe	387	6,023	3
47	yhdistys	202	6,009	3
48	rock	47	5,992	3
49	klassinen	56	5,991	3
50	urheilu	113	5,964	3
51	tango	30	5,952	3
52	sinä	84	5,938	3
53	kissa	52	5,928	1
54	häätä	38	5,895	3
55	kirjallisuus	66	5,884	3
56	lapsuus	38	5,822	3
57	rakastaa	42	5,782	3
58	musiikki	298	5,776	3
59	parissa	44	5,771	3
60	kaveri	78	5,763	3
61	puoliso	38	5,729	3
62	tapaaminen	58	5,719	3
63	tuttu	192	5,707	3
64	jazz	59	5,698	3
65	juhlia	85	5,698	3
66	eräs	207	5,679	3
67	kutsua	152	5,654	3
68	kutsu	33	5,644	3
69	suku	45	5,634	3
70	vaimo	102	5,625	3
71	hyvä	1795	5,609	3
72	yhteistyökumppani	31	5,605	3
73	jalkapallo	50	5,589	3
74	kanssa	1079	5,573	3
75	eläin	96	5,549	3
76	kirje	66	5,549	3
77	kokoontua	84	5,542	3
78	muistaa	167	5,542	3
79	lukuisa	49	5,532	3

80	viettää	138	5,518	3
81	lainata	30	5,489	3
82	taide	118	5,486	3
83	ilo	69	5,473	3
84	äiti	138	5,473	3
85	kesken	115	5,471	3
86	ooppera	61	5,458	3
87	jääkiekko	49	5,455	3
88	laulu	105	5,450	1
89	tanssia	47	5,429	3
90	mukava	63	5,424	3
91	auttaa	150	5,418	3
92	iloinen	48	5,398	3
93	muistella	64	5,367	3
94	todellinen	113	5,337	3
95	kahvi	30	5,306	3
96	nauttia	78	5,292	3
97	kaivata	90	5,287	3
98	keskuudessa	35	5,278	3
99	veli	43	5,258	3
100	tytär	56	5,255	3
101	vanha	501	5,233	3
102	tänne	43	5,223	3
103	koira	77	5,203	3
104	ruoka	94	5,198	3
105	kortti	45	5,193	3
106	minä	421	5,188	3
107	henkilökohtainen	60	5,184	3
108	suositella	31	5,172	3
109	muisto	44	5,155	3
110	soittaa	115	5,136	3
111	luonto	107	5,128	3
112	kirja	221	5,125	3
113	yhdessä	129	5,113	3
114	kuollut	53	5,108	3
115	houkutella	48	5,103	3
116	seura	162	5,069	3
117	tuntea	260	5,068	3
118	aito	35	5,050	3
119	keskenään	31	5,039	3
120	tuntematon	33	5,026	3
121	lämmin	45	5,022	3
122	Kreikka	34	5,021	1
123	tutustua	66	5,021	3

124	kuolema	61	5,020	3
125	hauska	38	5,006	3
126	kulttuuri	108	5,003	3
127	löytää	201	4,960	3
128	vierailla	60	4,957	3
129	kuunnella	56	4,938	3
130	tanssi	32	4,914	3
131	sydän	47	4,889	3
132	viihtyä	31	4,882	3
133	Sibelius	34	4,855	2
134	ympäri	41	4,841	3
135	elokuva	165	4,816	3
136	asua	143	4,800	3
137	kirjasto	49	4,797	1
138	vapaaehtoinen	30	4,795	3
139	hevonen	40	4,793	3
140	ulkomaalainen	42	4,792	3
141	unohtaa	45	4,780	3
142	köyhä	33	4,750	3
143	Minna	30	4,741	2
144	luottaa	37	4,718	3
145	tilaisuus	70	4,715	3
146	koti	122	4,707	3
147	kirjoittaa	136	4,700	3
148	menettää	79	4,657	3
149	harrastaa	30	4,651	3
150	John	36	4,633	2
151	ravintola	49	4,627	3
152	lintu	33	4,621	3
153	liikunta	31	4,608	3
154	ranskalainen	45	4,600	3
155	kala	44	4,597	3
156	hän	2058	4,592	3
157	kotona	38	4,590	3
158	Clinton	45	4,584	2
159	tarjolla	40	4,574	3
160	tarjota	154	4,570	3
161	oppia	64	4,567	3
162	ilta	101	4,565	3
163	lähettää	83	4,565	3
164	kadota	44	4,558	3
165	etsiä	94	4,549	3
166	arvostaa	38	4,546	3
167	paikalle	37	4,536	3

168	ihmetellä	49	4,531	3
169	puhelin	32	4,520	3
170	saapua	78	4,504	3
171	tosi	53	4,497	3
172	kysyä	88	4,485	3
173	perustaa	117	4,482	3
174	täällä	84	4,482	3
175	tyttö	82	4,450	3
176	joskus	66	4,449	3
177	he	518	4,439	3
178	saksalainen	53	4,434	3
179	turkulainen	34	4,430	3
180	moni	377	4,429	3
181	pyytää	72	4,415	3
182	onneksi	60	4,411	3
183	vierailu	33	4,392	3
184	paitsi	42	4,389	3
185	kuolla	72	4,386	3
186	kohdata	42	4,383	3
187	konsertti	57	4,379	3
188	keskellä	30	4,376	3
189	apu	151	4,349	3
190	joukko	161	4,333	3
191	vanhus	36	4,297	3
192	museo	30	4,293	3
193	Israel	44	4,253	1
194	teatteri	71	4,247	3
195	amerikkalainen	78	4,241	3
196	tai	578	4,234	3
197	istua	57	4,234	3
198	joku	97	4,225	3
199	puolesta	54	4,177	3
200	usein	92	4,177	3
201	erityisesti	87	4,169	3
202	järjestää	168	4,167	3
203	suomalainen	348	4,150	3
204	kautta	103	4,149	3
205	Koivisto	32	4,148	2
206	kuka	76	4,135	3
207	kylä	63	4,129	3
208	lapsi	269	4,124	3
209	vieras	64	4,114	3
210	aikuinen	36	4,103	3
211	nuori	269	4,099	1

212	kerätä	60	4,098	3
213	löytyä	166	4,092	3
214	yö	65	4,082	3
215	ymmärtää	69	4,033	3
216	yhteys	70	4,016	3
217	runsaasti	37	4,016	3
218	tuki	109	4,005	3
219	laji	53	4,004	3
220	väittää	34	4,001	3
221	kertoa	456	3,997	3
222	kokea	71	3,992	3
223	palkita	31	3,992	3
224	tuoda	164	3,990	3
225	taas	109	3,961	3
226	perinteinen	51	3,961	3
227	kaikki	499	3,951	3
228	paljon	442	3,927	3
229	ja	6776	3,900	3
230	yllättää	40	3,898	3
231	siellä	77	3,895	3
232	entinen	155	3,890	3
233	hankkia	58	3,866	3
234	opettaja	58	3,857	3
235	varmasti	48	3,850	3
236	Pentti	33	3,835	2
237	asunto	57	3,830	3
238	omistaa	41	3,812	3
239	järjestö	47	3,797	3
240	lähteä	182	3,790	3
241	käydä	244	3,783	3
242	pelätä	40	3,769	3
243	keskustella	40	3,748	3
244	joka	1582	3,733	3
245	jälleen	39	3,722	3
246	muutama	114	3,711	3
247	tässä	36	3,711	3
248	toivoa	77	3,711	3
249	oikea	60	3,694	3
250	haluta	236	3,685	3

LIITE 2

Lekseemin PŘÍTEL ja PŘÍTELKYNĚ *MI*-testin nostamat kollokaatit

(ryhmät: 1. yhdistyksiin liittyvät kollokaatit 2. ihmisten erisnimet 3. luonnolliset kollokaatit)

#	lemma	frekvenssi	I-arvo	suomennos	ryhmä
1	žeh	44	10,938	tuhkaus	1
2	Verešová	40	10,610		2
3	čtyřnohý	180	10,463	nelijalkainen	3
4	Fayed	69	10,426		2
5	Páleník	58	10,180		2
6	Dodiho	52	9,921		2
7	Dodi	45	9,501		2
8	Macháčková	38	9,195		2
9	nerozlučný	39	9,157	erottamaton	3
10	Camilla	95	9,112		2
11	dlouholetý	660	8,855	pitkäaikainen	3
12	příbuzný	695	8,392	sukulainen	3
13	důvěrný	164	8,391	luottamuksellinen	3
14	lokálka	37	8,361	lähijuna	1
15	posedět	52	8,219	istua (iltaa)	3
16	Dianin	38	7,917		2
17	disent	41	7,913	vastarinta, toisinajattelu	3
18	posezení	98	7,904	istuskelu	3
19	rádce	46	7,811	neuvonantaja	3
20	rozejít	135	7,671	erota	3
21	kruh	534	7,638	piiri	3
22	spolek	439	7,606	seura	3
23	nerost	30	7,594	kivennäinen	1
24	Jennifer	39	7,521		2
25	obdivovatel	55	7,424	ihailija	3
26	Linda	51	7,410		2
27	oddaný	34	7,400	uskollinen	3
28	vážený	142	7,371	arvoissa	3
29	rozchod	74	7,330	ero	3
30	Hradčany	43	7,276		1
31	věrný	218	7,232	uskollinen	3
32	vlastenecký	58	7,212	isänmaallinen	1
33	blízky	1545	7,205	läheinen	3
34	pohádat	33	7,166	riidellä	3
35	nepřítel	265	7,159	vihollinen	3
36	oženit	78	7,152	mennä naimisiin (miehenä)	3
37	Diana	146	7,150		2
38	milenka	68	7,102	rakastajatar	3
39	Andrea	87	7,090		2
40	stýkat	43	6,995	tapailta	3
41	milý	185	6,971	rakas	3
42	popovídat	34	6,952	jutella	3

43	dechovka	43	6,931	puhallinorkesteri	1
44	mileneec	86	6,899	rakastaja	3
45	spolužák	90	6,892	koulukaveri	3
46	popíjet	37	6,876	juoda (alkoholia)	3
47	spolupracovník	265	6,861	työkaveri	3
48	modelka	92	6,835	naismalli	3
49	Gott	77	6,809		2
50	přemluvit	32	6,804	suostutella	3
51	Al	62	6,771		2
52	večírek	76	6,741	bileet	3
53	celoživotní	64	6,731	elinikäinen	3
54	hrstka	35	6,724	kourallinen	3
55	večeře	136	6,712	illallinen	3
56	obklopit	43	6,691	ympäröidä	3
57	zemřelý	59	6,681	kuollut	3
58	zesnulý	52	6,623	edesmennyt	3
59	zavražděný	56	6,617	murhattu	3
60	Havlův	76	6,601		2
61	homosexuální	30	6,598	homoseksuaalinen	3
62	kamarádka	63	6,597	naiskaveri	3
63	známá	33	6,572	naistuttava	3
64	Marián	67	6,509		2
65	letitý	42	6,471	pitkäaikainen	3
66	rodina	1326	6,433	perhe	3
67	Lewinská	31	6,424		2
68	Suk	38	6,416		2
69	pozvat	224	6,414	kutsua	3
70	kamarád	283	6,411	kaveri	3
71	manželka	478	6,406	vaimo	3
72	těhotný	39	6,372	raskaana oleva	3
73	intimní	42	6,334	intiimi	3
74	sdužení	698	6,328	järjestö	3
75	Veronika	46	6,272		2
76	Clintonův	39	6,265		2
77	dětství	100	6,253	lapsuus	3
78	Hrabal	35	6,250		2
79	oslovovat	37	6,238	puhutella	3
80	zdravotně	71	6,236	terveydellisesti	1
81	rozloučit	86	6,231	sanoa hyvästi	3
82	rodák	113	6,192	alkuasukas	1
83	trávit	116	6,139	viettää	3
84	zavraždit	68	6,123	murhata	3
85	princezna	107	6,107	prinsessa	3
86	rodinný	408	6,084	perheellinen	3
87	Charles	71	6,081		2
88	rodič	581	6,079	vanhempi (vanhemmat)	3
89	spojenec	105	6,065	liittolainen	3
90	Radka	41	6,052		2
91	potkávat	32	6,015	kohdata	3

92	kolega	500	5,997	työkaveri, kollega	3
93	vlivný	59	5,964	vaikutusvaltainen	3
94	opravdový	82	5,958	todellinen	3
95	zvat	84	5,950	kutsua	3
96	telefonovat	44	5,940	soittaa (puhelimella)	3
97	autor	53	5,933	kirjoittaja	3
98	příroda	362	5,931	luonto	1
99	svěřit	143	5,925	luottaa	3
100	spousta	283	5,922	paljon	3
101	svěřovat	30	5,908	luottaa	3
102	herečka	141	5,903	näyttelijätär	3
103	Max	40	5,900		2
104	kolegyně	49	5,896	naiskollega	3
105	doprovázet	111	5,886	saattaa/seurata	3
106	okruh	232	5,878	piiri	3
107	dávný	261	5,810	pitkäaikainen	3
108	navštěvovat	78	5,801	vierailla	3
109	pozdání	53	5,797	kutsu	3
110	mládí	65	5,768	nuoruus	3
111	soused	164	5,759	naapuri	3
112	Jeseníky	34	5,751		1
113	Harry	48	5,728		2
114	zavolat	98	5,722	soittaa	3
115	princ	86	5,719	prinssi	3
116	Strana	96	5,713		1
117	dcera	265	5,711	tytär	3
118	parta	85	5,693	porukka	3
119	tvůj	585	5,686	sinun	3
120	dárek	96	5,665	lahja	3
121	zajít	54	5,659	käydä	3
122	betlém	67	5,656	Betlehem	1
123	Martina	77	5,648		2
124	setkání	340	5,647	tapaaminen	3
125	duha	34	5,646	sateenkaari	1
126	skupinka	40	5,644	ryhmä	3
127	poezie	91	5,638	runous	1
128	doprovod	75	5,630	seura	3
129	spolu	535	5,629	yhdessä	3
130	společně	307	5,622	yhdessä	3
131	Lucie	102	5,620		2
132	můj	3697	5,617	minun	3
133	svůj	8657	5,604	oma	3
134	narozeniny	85	5,591	syntymäpäivä	3
135	koníček	33	5,589	harrastus	3
136	vyprávět	159	5,588	kertoa/tarinoida	3
137	Ivana	72	5,575		2
138	hádky	30	5,567	riita	3
139	půvabný	31	5,548	sievä	3
140	darovat	39	5,537	lahjoittaa	3

141	zpěvačka	100	5,498	laulajatar	3
142	setkávat	61	5,490	tapailia	3
143	sběratel	44	5,471	keräilijä	1
144	básník	119	5,466	runoilija	3
145	manžel	241	5,456	aviomies	3
146	pořádat	209	5,456	järjestää	3
147	jeho	8258	5,448	hänen (M)	3
148	malíř	133	5,414	taidemaalari	3
149	navštívit	215	5,408	vierailla	3
150	Monika	43	5,396		2
151	vzpomínka	102	5,393	muisto	3
152	oslavit	77	5,378	juhlia	3
153	někdejší	152	5,362	senaikainen	3
154	bývalý	981	5,358	entinen	3
155	zastřelit	62	5,352	ampua kuoliaaksi	3
156	přátelství	34	5,352	ystävyyss	3
157	povídat	49	5,351	jutella	3
158	uspořádat	113	5,346	järjestää	3
159	lva	44	5,340		2
160	svědectví	49	5,315	todistajanlausunto	3
161	Jana	198	5,314		2
162	přivést	98	5,294	tuoda mukaan	3
163	známý	866	5,285	tuttava	3
164	scházet	78	5,276	tapailia	3
165	bydlet	161	5,271	asua	3
166	fotograf	81	5,255	valokuvaaja	3
167	založit	201	5,253	perustaa	3
168	dobrý	1931	5,252	hyvä	3
169	hudba	407	5,246	musiikki	1
170	prezident	55	5,245	presidentti	3
171	matka	243	5,203	äiti	3
172	soukromí	35	5,203	yksityisyys	3
173	maminka	87	5,198	äiti	3
174	seriál	181	5,185	tv-sarja	3
175	Honza	30	5,182		2
176	Petra	65	5,173		2
177	potkat	87	5,165	kohdata	3
178	klub	1050	5,156	kerho	3
179	miss	36	5,152	missi	3
180	vydávat	122	5,139	julkaista	3
181	bratr	168	5,125	veli	3
182	pohřeb	32	5,116	hautajaiset	3
183	bok	48	5,105	vieressä/ohella	3
184	nůž	67	5,098	veitsi	3
185	příznivec	113	5,093	kannattaja	3
186	nazývat	47	5,092	kutsua	3
187	spisovatel	132	5,088	kirjailija	3
188	kavárna	57	5,077	kahvila	3
189	sestra	112	5,071	sisko	3

190	poznat	140	5,064	tutustua/tunnistaa	3
191	otec	242	5,062	isä	3
192	osobní	357	5,060	henkilökohtainen	3
193	opustit	175	5,048	jättää	3
194	Eva	136	5,041		2
195	Olga	39	5,031		2
196	půjčit	45	5,027	lainata	3
197	hodný	39	5,018	kiltti	3
198	studentka	31	5,018	naisopiskelija	3
199	spolehlivý	34	4,989	luotettava	3
200	protivník	44	4,983	vastustaja	3
201	syn	225	4,980	poika	3
202	příslušník	107	4,970	jäsen	3
203	radit	47	4,962	neuvoa	3
204	Kateřina	74	4,954		2
205	vzpomínat	140	4,945	muistella	3
206	společník	30	4,939	seuralainen	3
207	seznámit	87	4,932	tutustua	3
208	postižený	171	4,930	vajaakykyinen	1
209	strávit	88	4,926	viettää	3
210	Bohumil	48	4,926		2
211	opouštět	38	4,916	jättää	3
212	dopis	206	4,914	kirje	3
213	Alexandr	37	4,909		2
214	kdysi	113	4,906	aikoinaan	3
215	odjíždět	30	4,904	lähteä/olla lähdössä	3
216	sebevražda	44	4,901	itsemurha	3
217	pivo	147	4,899	olut	1
218	Paul	41	4,882		2
219	najít	552	4,879	löytää	3
220	kněz	32	4,872	pappi	3
221	bar	54	4,861	baari	3
222	pomáhat	156	4,853	auttaa	3
223	Jitka	47	4,845		2
224	sochař	33	4,832	kuvanveistäjä	3
225	umění	315	4,817	taide	1
226	tehdejší	106	4,807	silloinen	3
227	prozradit	90	4,805	paljastaa	3
228	vyrazit	75	4,802	ryhtyä/lähteä	3
229	herec	173	4,798	näyttelijä	3
230	víno	98	4,797	viini	3
231	spoluhrač	58	4,786	tiimitoveri	3
232	volat	86	4,782	soittaa	3
233	muzikant	45	4,782	muusikko	3
234	setkat	116	4,773	tavata	3
235	úzký	65	4,765	kapea	3
236	pes	176	4,761	koira	3
237	ochránce	38	4,756	puolustaja	3
238	citovat	34	4,756	siteerata	3

239	pár	278	4,754	pari	3
240	Lenka	40	4,746		2
241	Boris	51	4,732		2
242	milovat	54	4,721	rakastaa	3
243	hledat	263	4,714	etsiä	3
244	odjet	77	4,705	lähteä	3
245	pomoc	340	4,702	apu	3
246	ekologický	99	4,697	ekologinen	1
247	láska	146	4,688	rakkaus	3
248	asociace	65	4,684	yhdistys	3
249	zpěvák	82	4,678	laulaja	3
250	oběd	35	4,668	lounas	3
251	pařížský	35	4,667	pariisilainen	3
252	sejít	130	4,655	tavata	3
253	bavit	92	4,651	jutella/viihtyä	3
254	návštěva	191	4,644	vierailu	3
255	hospoda	60	4,633	kapakka	3
256	starý	694	4,631	vanha	3
257	cestovat	46	4,596	matkustaa	3
258	výlet	52	4,596	reissu	3
259	posílat	52	4,595	lähettää	3
260	mnoho	399	4,593	moni	3
261	poradit	56	4,586	neuvoa	3
262	partner	174	4,578	puoliso, kumppani	3
263	potěšit	45	4,570	ilahduttaa	3
264	mrtvý	111	4,569	kuollut	3
265	prosit	35	4,560	pyytää	3
266	s	9423	4,546	kanssa	3
267	prožít	48	4,545	kokea	3
268	vymyslet	34	4,539	keksiä	3
269	občas	110	4,516	välillä	3
270	dovolená	98	4,504	loma	3
271	přijet	193	4,493	tulla	3
272	styk	53	4,485	yhteys	3
273	slíbit	65	4,483	luvata	3
274	dívka	141	4,478	tyttö	3
275	žít	322	4,470	elää/asua	3
276	nazvat	32	4,470	kutsua	3
277	portrét	37	4,452	muotokuva	3
278	hudebník	37	4,441	muusikko	3
279	oslava	91	4,437	juhla	3
280	doporučení	32	4,435	suositus	3
281	vražda	107	4,426	murha	3
282	chata	41	4,425	mökki	3
283	domů	112	4,421	kotiin	3
284	zabít	72	4,413	tappaa	3
285	vítat	37	4,393	toivottaa tervetulleeksi	3
286	výtvarník	61	4,392	kuvataiteilija	3
287	požádat	128	4,392	pyytää	3

288	přítomnost	49	4.382	mukanaolo	3
289	vrah	41	4.370	murhaaja	3
290	nalézt	50	4.362	löytää	3
291	mezitím	35	4.353	sillä välin	3
292	skladatel	40	4.343	säveltäjä	3
293	líčit	32	4.339	kuvaila	3
294	restaurace	109	4.316	ravintola	3
295	sedět	89	4.307	istua	3

LIITE 3

Lekseemien YSTÄVÄ ja PŘÍTEL *MI*-testin antamien kollokaattien semanttiset ryhmät

Semanttinen ryhmä	YSTÄVÄ	PŘÍTEL	PŘÍTEL - suomennos
YSTÄ-VYYS-SUHDE	(36) aito, apu, arvostaa, auttaa, entinen, etsiä, hankkia, henkilökohdainen, hyvä, hätä, kaihata, lainata, lapsuus, luotettava, luottaa, lähin, lämmin, löytyä, löytää, menettää, oikea, palkita, pitkäaikainen, rakas, sydän, todellinen, tosi, tuki, tunnustautua, tuntea, tutustua, unohtaa, uskollinen, vanha, vanhoutunut, ylin	(35) celoživotní, darovat, dávný, dětství, dlouholetý, dobrý, důvěrný, hledat, letitý, milý, mládí, najít, nalézt, někdejší, nerozlučný, pomáhat, pomoc, poznat, půjčit, oddaný, opouštět, opravdový, opustit, osobní, přátelství, radit, rodinný, scházet, seznámit, spolehlivý, starý, svěřit, svěřovat, tehdejší, věrný	elinikäinen, lahjoittaa, pitkäaikainen, lapsuus, pitkäaikainen, hyvä, luotamuksellinen, etsiä, pitkäaikainen, rakas, nuoruus, löytää, löytää, senaikainen, erottamaton, auttaa, apu, tutustua/tunnistaa, lainata, uskollinen, jättää, todellinen, jättää, henkilökohtainen, ystävyys, antaa neuvoa, perhe-, tapailla, tutustua, luotettava, vanha, luottaa, luottaa, silloinen, uskollinen
OMINAI-SUUS	(15) amerikkalainen, edesmennyt, hauska, iloinen, innokas, köyhä, kuollut, mukava, ranskalainen, saksalainen, suomalainen, tuntematon, turkulainen, ulkomaalainen, yksinäinen	(10) čtyřnohý, hodný, homoseksuální, mrtvý, pařížský, půvabný, vážený, vlivný, zemřelý, zesnulý	nelijalkainen, kiltti, homoseksuaalinen, kuollut, pariisilainen, sievä, arvoisa, vaikutusvaltainen, kuollut, edesmennyt
TAPAAMINEN	(28) kahvi, kanssa, kerätä, kesken, keskenään, keskuudessa, kohdata, kokoontua, kutsu, kutsua, kylä, käydä, luona, lähteä, saapua, seura, tapaaminen, tavata, tervehtiä, tervetullut, toivottaa, tuoda, vierailu, vierailu, vieras, yhdessä, yhteys, ympäri	(31) bok, cestovat, doprovázet, doprovod, návštěva, navštěvovat, navštívít, obklopit, odjet, odjíždět, potkat, potkávají, pozvání, pozvat, přijet, přítomnost, přivést, rozloučit, s, sejit, setkání, setkat, setkávají, společně, společník, spolu, styk, stýkat, vítat, zajít, zvat	vieressä, matkustaa, saattaa, seura, vierailu, vierailu, vierailu, ympäröidä, lähteä, lähteä, kohdata, kohdata, kutsu, kutsua, tulla, mukanaolo, tuoda, sanoa hyvästi, kanssa, tavata, tapaaminen, tavata, tapailla, yhdessä, seuralainen, yhdessä, yhteys, tapailla, toivottaa tervetulleeksi, käydä, kutsua

TAPAH- TUMA	(19) ilta, istua, juhlia, jär- jestää, konsertti, lahja, merkkipäivä, ravin- tola, ruoka, syntymä- päivä, tanssi, tanssia, tarjolla, tarjota, tilai- suus, viettää, viihtyä, viini, yö	(21) bar, dárek, hospoda, kavárna, narozeniny, oběd, oslava, oslavít, popíjet, pořádat, po- sedět, posezení, res- taurace, sedět, strávit, trávit, večere, večírek, víno, vyrazit, uspořádat	baari, lahja, kapakka, kahvila, syntymäpäivä, lounas, juhla, juhlia, juoda (alkoholia), järjes- tää, istua (iltaa), istus- kelu, ravintola, istua, viettää, viettää, illallinen, bileet, viini, ryhtyä/läh- teä, järjestää
IHMISET	(24) aikuinen, kaveri, koira, kollega, lapsi, läheinen, naapuri, omainen, perhe, per- heenjäsen, suku, suku- lainen, toveri, tukija, tuttava, tuttu, tyttö, ty- tär, työtoveri, vanhus, veli, vihollinen, yh- teistyökumppani, äiti	(33) blízký, bratr, dcera, dívka, kamarád, ka- marádka, kolega, ko- legyně, kruh, maminka, matka, nepřítel, obdi- vovatel, ochránce, otec, parta, pes, příbuzný, příslušník, protivník, příznivec, rádce, rodič, rodina, sestra, soused, spojenec, spoluhrač, spolupracovník, spo- lužák, syn, známá, známý	läheinen, veli, tytär, tyttö, kaveri, naiskaveri, kol- lega, naiskollega, piiri, äiti, äiti, vihollinen, ihai- lija, puolustaja, isä, po- rukka, koira, sukulainen, jäsen, vastustaja, kannat- taja, neuvonantaja, van- hempi, perhe, sisko, naa- puri, liittolainen, tiimito- veri, työkaveri, kouluka- veri, poika, naistuttava, tuttava
VIESTINTÄ	(19) jutella, kertoa, keskus- tella, kirje, kirjoittaa, kortti, kuunnella, ky- syä, lähettää, muistaa, muistella, muisto, op- pia, puhelin, pyytää, soittaa, suositella, väittää, ymmärtää	(26) bavit, citovat, dopis, do- poručení, líčit, nazvat, nazývát, oslovovat, pohádat, popovídát, po- radit, posílat, povídat, požádat, prosit, prozra- dit, přemluvit, slíbit, te- lefonovat, volat, vy- dávat, vymyslet, vy- právět, vzpomínat, vzpomínka, zavolat	jutella, siteerata, kirje, suositus, kuvailla, kutsua, kutsua, puhutella, rii- dellä, jutella, neuvoa, lä- hettää, keskustella, pyy- tää, pyytää, paljastaa, suostutella, luvata, soit- taa, soittaa, julkaista, keksiä, kertoa, muistella, muisto, soittaa

HARRAS- TUS	(32) elokuva, eläin, harras- taa, harrastus, hevo- nen, jalkapallo, jazz, järjestö, jääkiekko, kala, kirja, kirjalli- suus, klassinen, kult- tuuri, laji, liikunta, lintu, luonto, museo, musiikki, olut, oop- pera, perinteinen, perustaa, rock, runous, ry, taide, tango, teat- teri, urheilu, yhdistys	(7) asociace, klub, koniček, sdružení, seriál, spolek, založit	yhdistys, kerho, harras- tus, järjestö, sarja, seura, perustaa
PARI- SUHDE	(2) puoliso, vaimo	(16) bydlet, bývalý, hádka, intimní, láska, manžel, manželka, milenec, mi- lenka, milovat, oženit, partner, rozejít, rozchod, těhotný, žít	asua, entinen, riita, in- tiimi, rakkaus, aviomies, aviovaimo, rakastaja, ra- kastajatar, rakastaa, mennä naimisiin, puo- liso, erota, ero, raskaana oleva, asua
AMMATTI	(2) vapaaehtoinen, opet- taja	(21) autor, básník, fotograf, herec, herečka, hudeb- ník, kněz, malíř, miss, modelka, muzikant, prezident, princ, prin- cezna, skladatel, sochař, spisovatel, studentka, výtvarník, zpěvačka, zpěvák	kirjoittaja, runoilija, va- lokuvaaja, näyttelijä, näyttelijätär, muusikko, pappi, taidemaalari, missi, naismalli, muu- sikko, presidentti, prinssi, prinsessa, säveltäjä, ku- vanveistäjä, kirjailija, naisopiskelija, kuvataitei- lija, laulajatar, laulaja
RIKOS	(1) kadota	(9) nůž, sebevražda, svědectví, vrah, vražda, zabít, zastřelit, zav- ražděný, zavraždit	veitsi, itsemurha, todista- janlausunto, murhaaja, murha, tappaa, ampuu kuoliaaksi, murhattu, murhata

TUNNE	(10) haluta, ihmetellä, ilahduttaa, ilo, nauttia, pelätä, rakastaa, suru, toivoa, yllättää	(1) potěšit	ilahduttaa
ADVERBIAALIT	(24) asua, asunto, erityisesti, eräs, joskus, joukko, kaikki, keskellä, koti, kotona, lukuisa, moni, muutama, myöhemmin, paikalle, paljon, parissa, runsaasti, siellä, tänne, tässä, täällä, usein, varmasti	(14) domů, dovolená, hrstka, chata, kdysi, mezitím, mnoho, občas, okruh, pár, skupinka, spousta, úzký, výlet	kotiin, loma, kourallinen, mökki, aikoinaan, sillä välin, moni, välillä, piiri, pari, ryhmä, paljon, kaipa, reissu
KIELIOPPI	(15) hän, he, ja, jälleen, joka, joku, kautta, kuka, minä, onneksi, paitsi, puolesta, sinä, taas, tai	(4) jeho, můj, svůj, tvůj	hänen (M), minun, oma, sinun
MUUT	(6) houkutella, isänmaa, kokea, kuolema, kuolla, omistaa	(5) disent, pohřeb, portrét, prožít, soukromí	vastarinta, hautajaiset, muotokuva, kokea, yksityisyys