

Yhdyssanat eri taitotasoilla S2-teksteissä

Pihla Metsänoja
Kandidaatintutkielma
Humanististen tieteiden kandidaatti, suomen kieli
Kieli- ja käännöstieteiden laitos
Humanistinen tiedekunta
Turun yliopisto
Toukokuu 2026

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

Humanististen tieteiden kandidaatti, suomen kieli

Pihla Metsänoja

Yhdyssanat eri taitotasoilla S2-teksteissä

Sivumäärät: 25 s, liitteet 1 s

Tiivistelmä

Tässä kandidaatintutkielmassa tarkastellaan yhdyssanojen käyttöä suomi toisena kielenä -oppijoiden kirjoittamissa teksteissä eri CEFR-taitotasoilla (A1-C2). Tutkimuksen tavoitteena on saada selvitettyä, kuinka paljon yhdyssanoja eri taitotasoilla esiintyy sekä millaisia ne ovat rakenteeltaan ja merkitykseltään. Aineistona tutkimuksessa käytetään Kansainvälistä oppijansuomen korpusta, josta on analysoitu kunkin taitotason tekstejä sekä kvantitatiivisesti että kvalitatiivisesti.

Tutkimuksessa yhdyssanojen esiintyvyyttä tarkastellaan niiden suhteellisena määränä tuhatta sanaa kohden. Lisäksi yhdyssanojen sanaluokkia, rakenteellista kompleksisuutta ja semanttisia piirteitä analysoidaan. Tulokset osoittavat, että yhdyssanojen suhteellinen määrä kasvaa CEFR-taitotason noustessa. Kehitys ei kuitenkaan ole täysin lineaarista. Kaikilla taitotasoilla selkeästi suurimman ryhmän muodostavat substantiiviyhdyssanat. Yhdyssanojen rakenteellisessa tarkastelussa selviää, että kaksiosaiset yhdyssanat ovat moniosaisia yleisempiä. Moniosaisen määrä kuitenkin kasvaa ylemmillä taitotasoilla. Yhdyssanojen semanttinen luonne muuttuu alhaisemmilta tasoilta ylöspäin noustessa konkreettisista, arkisista ilmauksista abstrakteimmiksi.

Tutkimusten tulokset osoittavat, että yhdyssanojen käyttö heijastaa S2-oppijoiden leksikaalisen ja morfologisen kompetenssin kehittymistä. Yhdyssanojen voidaan tutkielman perusteella todeta tarjoavan käyttökelpoisen näkökulman kielenoppimisen tutkimukseen sekä kielenopetuksen kehittämiseen.

Avainsanat: yhdyssanat, suomi toisena kielenä, korpustutkimus, CEFR-taitotasot

Sisällysluettelo

1	Johdanto	3
2	Teoria	5
2.1	Yhdyssanojen rakenne ja määritelmät	6
2.2	Yhdyssanat S2-oppijoiden kielenkehityksessä	7
3	Aineiston keruu ja käsittely	8
4	Yhdyssanojen määrät eri taitotasoilla	9
5	Yhdyssanojen tyypit ja rakenteet eri taitotasoilla	11
5.1	A1- ja A2-tasojen yhdyssanat	12
5.2	B1- ja B2-tasojen yhdyssanat	13
5.3	C1- ja C2-tasojen yhdyssanat	13
6	Yleisimmät yhdyssanat eri taitotasoilla	14
6.1	A1- ja A2-tasojen yleisimmät yhdyssanat	15
6.2	B1- ja B2-tasojen yleisimmät yhdyssanat	16
6.3	C1- ja C2-tasojen yleisimmät yhdyssanat	17
7	Yhteenveto ja päätelmät	18
7.1	Päätelmät tutkimustuloksista	19
7.2	Lopuksi	22
	Lähteet	22
	Liitteet	23

1 Johdanto

Tarkastelen tässä tutkimuksessa suomi toisena kielenä -oppijoiden kirjoittamissa eri taitotasolle luokitelluissa teksteissä esiintyviä yhdyssanoja. Jatkossa viitataan suomi toisena kielenä -termiin lyhenteellä S2. Yhdyssanat ovat kahdesta tai useammasta itsenäisestä sanasta muodostuvia yhtä käsitettä ilmaisevia sanoja, joissa määriteosa täsmentää perusosaa. Yhdyssanat kirjoitetaan kiinteästi yhteen. (Eronen – Maamies – Räikkälä 1996.) Aineistoni koostuu Kansainvälisen oppijansuomen korpuksesta löytyvistä teksteistä, joita tutkimalla selvitän yhdyssanojen esiintyvyyttä sekä niiden rakenteellisia muotoja, kompleksisuutta ja semantiikkaa. Tutkimukseni on soveltavaa kielentutkimusta ja liittyy myös kielen oppimisen tutkimiseen.

Yhdyssanojen käyttöä S2-oppijoiden teksteissä on tutkittu melko vähän, vaikka yhdyssanat ovat yleisiä suomen kielessä: jopa 65% nykysuomen sanakirjan hakusanoista on yhdyssanoja (Eronen ym. 1996). Samaa aihepiiriä käsittelevä aiemmin tehty pro gradu -tutkielma *Voisitko apua? Suomi toisena kielenä -oppijoiden sananmuodostustaitojen jäljillä* (Penttinen 2010) tarkastelee suomenoppijoiden sananmuodostustapoja. Tutkimuksessa tulee ilmi, että sananmuodostustaidon hallinta etenee muun kielellisen taitotason kasvaessa. Virpi Masonen taas nostaa artikkelissaan *Leksikaalinen kompetenssi osana kielitaitoa* (2013) esiin sitä, kuinka yhdyssanojen hallinta liittyy sekä oppijan sanaston laajuuteen että syvyyteen. Masosen mukaan olisi tärkeää mitata yhdyssanoja erillisenä osana testeissä jotka mittaavat kielitaitoa. Vain siten yhdyssanojen hallintaa voidaan kehittää, esimerkiksi huomioimalla se paremmin kielenopetuksessa. Yhdyssanojen käyttöä S2-oppijoiden teksteissä on kiinnostavaa ja tarpeellista tutkia, sillä tutkimuksella voidaan tuottaa tietoa siitä, miten kielenoppijoiden leksikaalinen ja morfologinen kompleksisuus kehittyy ja miten se tulisi ottaa huomioon esimerkiksi kielenopetuksessa.

Tutkimukseni aineistona käytän Kansainvälistä oppijansuomen korpusta (International Corpus of Learner Finnish, ICLFI), joka koostuu eri äidinkielisten S2-opiskelijoiden kirjoittamista teksteistä. Korpuksen tekstit on kerätty eri puolilta maailmaa suomen kielen opettajien toimesta. Aineistossa on sekä fiktiivisiä- että asiatekstejä, kuten esseitä, tarinoita ja kirjeitä. Korpus tarjoaa runsaasti taustatietoa tekstien kirjoittajista ja oppimistilanteista, kuten oppijan äidinkielestä, iästä ja asuinpaikasta sekä opettajan äidinkielestä ja tekstin kirjoituspaikasta. En kuitenkaan tutkimuksessani ota huomioon yksittäisistä kirjoittajista annettuja taustatietoja. Hyödynnän tutkimuksessani ainoastaan taitotasoluokituksia, joihin kaikki käyttämäni korpuksen tekstit on jaoteltu.

Korpuksessa S2-oppijoiden tekstit on jaoteltu taitotasoihin A1-C2. Tutkimuksessani keskityn korpuksesta saamaani sataan ensimmäiseen hakutulokseen jokaisella taitotasolla. Tulosten analysoinnissa yhdistän määrällisen tiedon ja laadullisen tulkinnan. Käytän kvantitatiivista menetelmää, sillä tavoitteenani on selvittää konkreettisia lukuja yhdyssanojen esiintyvyydestä. Lisäksi tarkastelen ja analysoin poimimiani yhdyssanoja kvalitatiivisesti, esimerkiksi niiden semanttisten merkitysten näkökulmasta.

Tutkimukseni päätavoitteena on siis selvittää, miten yhdyssanojen käyttö eroaa eri taitotasoilla olevien S2-oppijoiden kirjoittamissa teksteissä. Selvitän, kuinka usein ja millaisia yhdyssanoja kielenoppijat käyttävät ja pohdin niiden avulla eroja kielen sanaston ja rakenteen hallitsemisessa oppimisen eri vaiheissa.

Vastaan tutkielmassani seuraaviin tutkimuskysymyksiin:

1. Kuinka paljon yhdyssanoja esiintyy eri taitotasoille jaetuissa suomi toisena kielenä -oppijoiden kirjoittamissa teksteissä?
2. Millaisia eri taitotasoilla olevien tekstien sisältämät yhdyssanat ovat ja miten ne eroavat toisistaan?

Aikaisemmin tehtyyn tutkimukseen (esim. Penttinen 2010; Masonen 2013; Bulte – Roothoof 2020) perustuen tutkimukseni hypoteesina on, että yhdyssanojen määrä ja monimuotoisuus kasvavat S2-oppijoiden teksteissä korkeammille taitotasoille siirryttäessä. Uskon alempien taitotasojen yhdyssanojen koostuvan pääsääntöisesti kahdesta yleisestä perusmuotoisesta sanasta, ja korkeammille taitotasoille noustessa yhdyssanojen määrän kasvavan ja niiden rakenteiden monimutkaistuvan. Korkeimmalle taitotasolle luokitelluissa teksteissä arvelen yhdyssanojen lisääntyvän entisestään. Uskon yhdyssanojen myös olevan moniosaisempia ja vähemmän konkreettisia alempiin taitotasoisiin verrattuna. Hypoteesiani tukee myös CAF-malli, jonka esittelen alaluvussa 2.2.

2 Teoria

2.1 Yhdyssanojen rakenne ja määritelmät

Suomen kielessä yhdyssanat ovat keskeinen sananmuodostustapa. Yhdyssanassa kaksi tai useampi sana muodostaa yhden omaa merkitystään kantavan käsitteellisen kokonaisuuden. Yhdyssanassa loppu- eli perusosa määrittää sen sanaluokan ja keskeisen merkityksen, kun taas alku- eli määriteosa tai -osat täsmentävät perusosaa. (Hongisto 2023, 2.) Esimerkiksi yhdyssanan *kahvikuppi* perusosa *kuppi* määrittää sen sanaluokan substantiiviksi ja kertoo sen viittaavan astiaan. Määriteosa *kahvi* taas tarkoittaa millainen kuppi on kyseessä. Vastaavasti yhdyssanassa opiskelijaravintola perusosa *ravintola* kertoo yhdyssanan keskeisen merkityksen ja määriteosa *opiskelija* rajaa merkitystä viittaamalla ravintolan käyttäjäryhmään. Yhdyssanoja syntyy suomen kieleen jatkuvasti lisää ja ne muodostavat merkittävän ryhmän suomen kielen sanastossa. Tämä heijastuu myös niiden runsaaseen esiintymiseen sanakirjoissa (Eronen ym. 1996).

Yhdyssanoja voidaan luokitella esimerkiksi sanaluokan perusteella. Isoon suomen kielioppiin (Hakulinen – Vilkuna – Korhonen – Koivisto – Heinonen – Alho 2004, jatkossa ISK) perustuen voidaan todeta, että suomen kielen yhdyssanat rakentuvat tyypillisimmin substantiivien pohjalta. Yleisimpiä suomen kielessä ovat siis nimenomaan substantiiviyhdyssanat. Substantiiviyhdyssanojen keskeinen asema tekee niistä relevantteja myös tässä tutkimuksessa, tutkittaessa S2-oppijoiden kielitaitoa. ISK:ssa mainitaan yhdyssanatyypeiksi lisäksi adjektiiviyhdyssanat, esimerkiksi *monivärinen* ja *kolmevuotias*, yhdysnumeraalit, esimerkiksi *kaksitoista* ja *kolmesataa*, ja johdosyhdyssanat, esimerkiksi *pitkätukkainen* ja *päivänselvä*. (ISK 2004, 388.)

ISK:n mukaan yhdyssanat jakautuvat rakenteellisesti kaksiosaisiin ja moniosaisiin yhdyssanoihin. Kaksiosaiset yhdyssanat, kuten *taskulamppu* ja *käsipeili*, ovat yleisempiä ja niiden semanttiset merkitykset ovat usein helposti tulkittavissa, sillä niiden sisältämät sanat ovat suhteeltaan epäsymmetrisiä (ISK 2004, 388). Moniosaisten yhdyssanojen, esimerkiksi *päivähoitopaikka* tai *terveydenhuoltojärjestelmä*, rakenteet ovat kompleksisempia. Niiden semanttiset merkitykset voivat olla vaikeammin tulkittavissa, sillä niillä on useampia määriteosia. Yhdyssanojen rakenteellinen jaottelu on keskeistä tämän tutkimuksen analyysissa, kun yhdyssanojen kompleksisuutta tarkastellaan eri taitotasolla.

Merkitykseltään yhdyssanat voivat olla konkreettisia tai abstrakteja. Konkreettisilla yhdyssanoilla viitataan arkisiin ja helposti havaittaviin ilmiöihin, esimerkiksi *kirjahylly* ja *polkupyörävarasto*. Abstraktien yhdyssanojen, esimerkiksi *kielitaito* ja *ajankäyttö*, merkitykset taas ovat enemmän käsitteellisiä tai muodollisia. Toisen kielen oppimista tutkittaessa on havaittu rakenteellisen kompleksisuuden korreloivan positiivisesti opittavan kielen taitotason kanssa (ks. Bulte – Roothoof 2020, luku 5). Tästä voidaan päätellä abstraktien yhdyssanojen käytön usein lisääntyvän kehittyneemmässä kielenkäytössä, mikä tekee myös tästä näkökulmasta mielenkiintoisen eri taitotasoja vertailtaessa.

2.2 Yhdyssanat S2-oppijoiden kielenkehityksessä

Tässä tutkimuksessa S2-oppijoiden kielitaitoa tarkastellaan erityisesti leksikaalisen ja morfologisen kompetenssin näkökulmasta. Kielitaito voidaan määritellä olevan kommunikatiivista kykyä, joka koostuu sekä kielen rakenteiden hallinnasta että niiden tarkoituksenmukaisesta käytössä erilaisissa kielenkäyttötilanteissa (Bachman 1990, 84–90). Leksikaalisella kompetenssilla tarkoitetaan kielenkäyttäjän kykyä käyttää ja ymmärtää sanoja tarkoituksenmukaisesti. Toisen kielen oppimisen tutkimuksessa leksikaalinen kompetenssi voidaan jakaa sanaston laajuuteen ja syvyyteen. Laajuus viittaa oppijan tuntemien sanojen määrään ja syvyys taas kertoo oppijan yksittäisten sanojen merkitysten tuntemuksesta. (Masonen 2013, 190–191.) Yhdyssanojen hallinta vaatii molempia, sillä niiden tuottamisen edellytyksenä on kyky yhdistää jo tunnettuja sanoja uusia merkityksiä sisältäviksi sanoiksi. Lisäksi tuottaminen vaatii kahden tai useamman sanan muodostamien kokonaisuuksien semanttisten suhteiden ymmärtämistä.

Morfologinen kompetenssi taas tarkoittaa kielenkäyttäjän kykyä hallita ja ymmärtää kielen sananmuodostus- ja taivutusjärjestelmää. Suomen kieli on morfologisesti rikas, mikä tekee morfologian hallitsemisesta hyvin olennaisen osan kielitaitoa (Rusava 2025, 4). Yhdyssanojen tuottaminen vaatii morfologista kompetenssia, sillä ne edellyttävät sanojen rajojen tuntemista sekä eri morfeemien yhdistämistä juuri tietyllä tavalla.

Kielen kompleksisuuden tarkastelu on mahdollista toisen kielen oppimisen tutkimuksessa vaikiintuneen kompleksisuuden (Complexity), tarkkuuden (Accuracy) ja sujuvuuden (Fluency) muodostaman kokonaisuuden avulla. Tätä kokonaisuutta kielentutkimuksessa kutsutaan CAF-malliksi. *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* -teoksessa (Housen – Kuiken – Vedder 2012) kompleksisuus määritellään kielen rakenteellisena ja leksikaalisena monipuolisuutena. Sen esitetään ilmenevän esimerkiksi

vaihtelevien rakenteiden, pidempien ilmaisujen sekä kehittyneemmän sanaston käytössä. Kompleksisuuden ei siis ajatella viittaavan vain kielen hankaluuteen, vaan siihen, kuinka monipuolisesti oppija pystyy ilmaisemaan kielen merkityksiä. Lisäksi kompleksisuuden on havaittu liittyvän kielelliseen tarkkuuteen ja sujuvuuteen: monimutkaisempia rakenteita tuottaessa kielenoppijan tarkkuus ja sujuvuus voivat tilapäisesti heikentyä. (Housen ym. 2012.) CAF-malli tarjoaa hyvän perustan myös tälle tutkielmalle.

Aiemmin tehdyssä tutkimuksessa on havaittu morfologisen kompleksisuuden lisääntymistä kielitaidon kehittyessä. Tämä ilmenee oppijoiden teksteissä esimerkiksi pidempinä sanamuotoina ja moniosaisina rakenteina sekä abstraktimpien ilmaisujen yleistymisenä. (Bulte – Root-hoof 2020). Tässä tutkielmassa yhdyssanojen käyttöä tarkastellaan erityisesti morfologisen kompleksisuuden näkökulmasta, sillä sen hallintaa voidaan pitää yhtenä indikaattorina S2-opiskelijan kielen kehittyneisyydestä. Yhdyssanojen esiintyvyyden tarkastelua ja vertailua eri taitotasoilla voidaan näin ollen pitää käyttökelpoisena näkökulmana kielenkehitystä tutkittaessa myös kokonaisuutena.

3 Aineiston keruu ja käsittely

Tutkimukseni aineistona käytän Kansainvälistä oppijansuomen korpusta, joka sisältää eri äidinkielisten eri taitotasolla olevien S2-oppijoiden kirjoittamia tekstejä. Oppijoiden kielitaito on määritelty CEFR-luokituksen (Common European Framework of Reference for Languages) tasojen A1-C2 mukaisesti. Tasot kuvaavat oppijan kielitaitoa usealla eri osa-alueella, esimerkiksi sanaston laajuuden, lauseiden monimutkaisuuden, kieliopin hallinnan sekä kielen ymmärtämisen ja tuottamisen taidon mukaan. Esimerkiksi A1-A2 tason oppija kykenee ymmärtämään ja tuottamaan yksinkertaisia lauseita, kun taas B1-B2 tasoilla taitotaso kasvaa. C1-C2 taitotasot ovat korkeimmat. Näillä tasoilla oppija hallitsee kielen lähes natiivitasolla, eli pystyy esimerkiksi käyttämään monimutkaisiakin lauserakenteita sekä monipuolista sanastoa ja ymmärtämään kielen abstraktejakin sisältöjä. (Council of Europe 2026).

Käyttämässäni korpuksessa kirjoittajien CEFR-taso on määritelty kirjoitettujen tekstien perusteella. Tasoluokitus mahdollistaa oppijoiden tekstien vertailun analyysissäni. Tässä tutkimuksessa tarkastelen kvantitatiivisesti sekä analysoin kvalitatiivisesti yhdyssanoja käyttämällä korpuksesta saamiani hakutuloksia jokaisella eri taitotasolla. Liitteessä 1 on aineiston keräämiseen käyttämäni tarkka CQP-hakulause, joka vaihtelee vastaavasti kaikille tasoille.

Yhdyssanojen esiintyvyyttä teksteissä mitaan yhdyssanojen määränä tuhatta sanaa kohden, jotta tulokset ovat vertailukelpoisia keskenään. Yhdyssanojen suhteellinen määrä teksteissä on laskettu kaavan 1 osoittamalla tavalla.

$$\text{Yhdyssanat per 1000 sanaa} = \frac{\text{yhdyssanojen määrä}}{\text{kaikkien sanojen määrä taitotasolla}} \times 1000$$

KUVIO 1. Yhdyssanojen suhteellisen määrän laskukaava.

Analysoin myöhemmin lisäksi sataa (100) ensimmäistä hakutulosta jokaisella taitotasolla, sekä kymmentä yleisimmin esiintyvää yhdyssanaa taitotasoin. Tämän analyysin tavoitteena on tarkastella muun muassa esiintyvien yhdyssanojen rakennetta ja semanttisia merkityksiä.

4 Yhdyssanojen määrät eri taitotasoilla

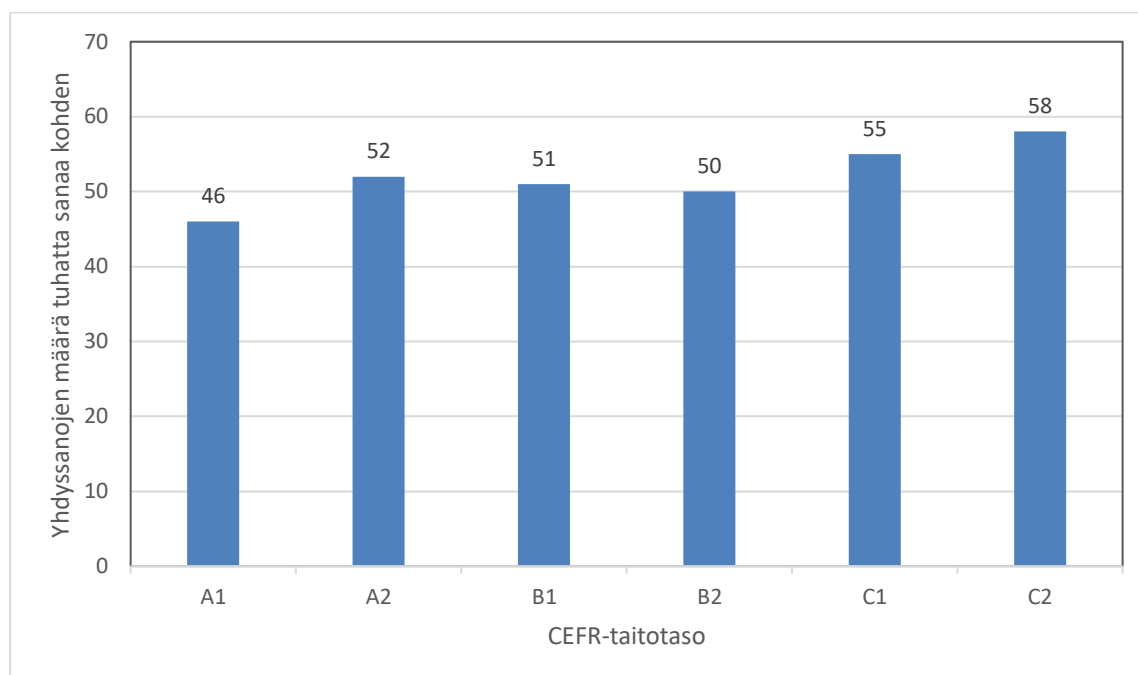
Tässä osassa tarkastelen kvantitatiivisesti yhdyssanojen esiintyvyyttä eri CEFR-tasoilla (A1-C2) S2-opiskelijoiden teksteissä. Aineistoni koostuu Kansainvälisen oppijansuomen korpukseen kerätyistä erilaisista teksteistä.

A1-taitotason korpusaineiston teksteissä kaikkia sanoja oli yhteensä 8 315 sanaa, joista hakuehdot täyttäviä yhdyssanoja oli 381 kappaletta. Laskukaavan mukaan tämän taitotason teksteissä tuhannesta sanasta yhdyssanoja oli 45,82, eli noin 46 kappaletta. A2-tasolla sanojen kokonaismäärä oli 95 821 ja hakuehdot täyttäviä yhdyssanoja esiintyi 5 010. Yhdyssanojen määrä oli siis 52,28, eli noin 52 kappaletta tuhatta sanaa kohden.

B1-taitotasolla sanoja oli yhteensä 407 215, joista 20 662 oli yhdyssanoja. Tämän perusteella yhdyssanojen määrä oli 50,74, eli noin 51 tuhatta sanaa kohden. B2-tason teksteissä sanoja oli kokonaisuudessaan 302 383. Yhdyssanoja näistä oli 15 157, jolloin yhdyssanojen esiintyvyys oli 50,13, eli noin 50 kappaletta tuhatta sanaa kohden.

C1-tason korpusaineistossa oli yhteensä 94 182 sanaa, joista 5 148 oli hakuehdot täyttäviä yhdyssanoja. Tämä vastaa laskukaavaan perustuen 54,66, eli noin 55 yhdyssanaa tuhatta sanaa kohden. C2-taitotasolla taas sanojen kokonaismäärä oli 19 651. Yhdyssanoja niistä oli 1 141 kappaletta. Yhdyssanojen esiintyvyys tällä taitotasolla oli siis 58,06, eli noin 58 kappaletta tuhatta sanaa kohden.

KUVIO 1. Yhdyssanojen määrä taitotasoin.



Tuloksista, jotka on koottu taulukkoon 1, käy ilmi, että yhdyssanojen määrä vaihtelee taitotasoin. A1-tasolla yhdyssanojen määrä tuhatta sanaa kohden on kaikista tarkastelluista tasoista matalin. Seuraavalla taitotasolla A2 yhdyssanojen esiintyvyys nousee selvästi. Tästä voidaan päätellä sanaston ja sananmuodostustaidon kehittyvän jo oppimisen varhaisessa vaiheessa.

B1-tason aineisto on huomattavasti laajempi verrattuna muihin taitotasoihin, mutta yhdyssanojen suhteellinen määrä pysyy melko tasaisena A2-tasoon verrattuna. B2-tasolla yhdyssanojen määrä tuhatta sanaa kohden pysyy edelleen melko tasaisena alempiin taitotasoihin verrattuna. Korkeammilla, C1- ja C2-taitotasolla yhdyssanojen suhteellinen esiintyvyys jälleen hieman kasvaa ja korkeimmalla taitotasolla C2 yhdyssanojen suhteellinen määrä onkin kaikista korkein. Tämä viittaa edistyneiden kieltenoppijoiden kehittyneeseen leksikaaliseen ja morfologiseen kompetenssiin. Kokonaisuudessaan yhdyssanojen esiintyvyyden S2-opiskelijoiden teksteissä voidaan siis todeta kasvavan ylemmille taitotasolle noustessa, vaikka kehitys ei ole täysin lineaarista. Erityisesti alimman (A1) ja ylimmän (C2) tason välinen ero on selkeä.

5 Yhdyssanojen tyypit ja rakenteet eri taitotasoilla

Tässä luvussa tarkastelen suomea toisena kielenä opiskelevien kirjoittajien teksteissä esiintyvien yhdyssanojen sanaluokkaa. Lisäksi tutkin yhdyssanojen rakenteellista kompleksisuutta, tarkemmin sitä, ovatko ne kaksi- vai moniosaisia. Korpuksen hakuehtoni ei anna tuloksia kaikista mahdollisesti kielenoppijoiden kirjoittamissa teksteissä esiintyvistä virheellisesti erikseen kirjoitetuista yhdyssanoista. En myöskään tässä tutkimuksessa ota huomioon esimerkiksi kirjoitusvirheitä, mikäli yhdyssanojen semanttinen merkitys on edelleen tulkittavissa. Tutkin jokaista CEFR-tasoa (A1-C2) erikseen. Tulokseni perustuvat korpuksesta saamiini sataan ensimmäiseen hakuehdon täyttävään ilmaukseen kullakin taitotasolla. Tuloksiani havainnollistavat kaaviot 2 ja 3 löytyvät luvusta 7, jossa myös analysoin saamiani tuloksia tarkemmin.

5.1 A1- ja A2-tasojen yhdyssanat

A1-taitotason sadasta ensimmäisestä hakutuloksesta substantiiviyhdyssanoja oli 92 kappaletta, esimerkiksi yhdyssanat *lentoasemalla* ja *olohuone*. Adjektiiviyhdyssanoja esiintyi kaksi kappaletta: *kolmevuotiaana* ja *vaikutusvaltaiset*. Yhdysnumeraaleja oli myös kaksi: *kuusikymmentä* ja *kolmekymmentä*. Sadan ensimmäisen tuloksen joukossa oli yksi kielenoppijan muodostama virheellinen yhdyssana, *kilonleipää*. Lisäksi hakuehdon täyttävistä sanoista tutkimustani ajatellen virheellisiä oli kolme kappaletta, sillä ne eivät lainkaan muodostaneet yhdyssanaa: *televisiosta*, *television* ja *television*.

Yhdyssanojen rakennetta tarkastellessa kaksiosaisien yhdyssanojen määrä A1-tasolla oli 90 kappaletta, joista esimerkkejä ovat *yliopistolle* ja *kotimaassasi*. Moniosaisia yhdyssanoja oli kuusi kappaletta, esimerkiksi *Kesälomamatka* ja *linja-autopysäkkiin*. Virheellisiä ilmauksia olivat samat neljä kuin edellä mainitsin.

A2-taitotason sadasta ensimmäisestä hakutuloksesta substantiiviyhdyssanoja oli 73 kappaletta, esimerkiksi yhdyssanat *isoäitini* ja *silmälasit*. Olen laskenut substantiiviyhdyssanoihin myös kaksi hieman epämääräistä rakennetta, *sukupolvetsotaa* ja *päärakkausjuoni*, sillä yhdyssanojen perusosa on kuitenkin tunnistettavissa substantiiviksi. Adjektiiviyhdyssanoja esiintyi 11 kappaletta, esimerkiksi *vaaleanruskea* ja *harvapuheinen*. Yhdysnumeraaleja taas oli 10 kappaletta, kuten yhdyssanat *kahdeksantoista* ja *neljäkymmentäkahdeksan*.

Aineiston kirjoittajien tuottamia selkeästi väärin muodostettuja yhdyssanoja oli kaksi: *espanjakieli* (lausekontekstin perusteella kirjoittaja tarkoittaa sanaa 'espanjan kieltä') sekä *jälkeenjuhlissa*. Lisäksi kolmen hakutuloksen, *laulunsaansa laulunsaanaan* ja *laulunsaanaan*,

merkitys jäi kokonaan epäselväksi. Tämä tekee näiden sanojen luokittelusta sanaluokkiin perustuen mahdotonta.

Yhdyssanojen rakennetta tarkastellessa A2-tasolla kaksiosaisien yhdyssanojen määrä oli 81 kappaletta, esimerkiksi yhdyssanat *elokuva* ja *kerrostalossa*. Moniosaisia yhdyssanoja oli 14 kappaletta, esimerkiksi *ex-poikaystävä* ja *omakotitalossa*. Olen laskenut moniosaisiin yhdyssanoihin myös aiemmin mainitsemani *sukupolvetsotaa* ja *päärakkausjuoni*. Täysin virheellisiä ja epäselviä ilmauksia en luokitellut lainkaan myöskään rakenteen perusteella.

5.2 B1- ja B2-tasojen yhdyssanat

B1-taitotasolla sadasta ensimmäisestä hakutuloksesta substantiiviyhdyssanoja oli 90 kappaletta, esimerkiksi sanat *viikonloppu* ja *lihapullat*. Olen laskenut substantiiviyhdyssanaksi myös hieman epätyypillisen ilmauksen *pikkutyttö*, joka voidaan kuitenkin luokitella substantiiviksi sen perusosan perusteella. Adjektiiviyhdyssanoja oli kuusi kappaletta, esimerkiksi yhdyssanat *ylimielinen* ja *viisivuottiaana*. Yhdysnumeraaleja oli kolme kappaletta: *yksitoista*, *yksitoista* ja *Kaksikymmentäviisi*. Yksi hakutuloksista oli selkeästi virheellinen: *'serranosta' kinkusta*, sillä ilmaus ei muodosta yhdyssanaa.

Rakenteeltaan B1-tason yhdyssanat olivat pääosin kaksiosaisia, sillä niitä oli yhteensä 95 kappaletta, esimerkiksi *hammaslääkärissä* ja *hautausmaalla*. Moniosaisia yhdyssanoja esiintyi tällä taitotasolla vain neljä, esimerkiksi *yliopistokavereiden* ja *valokuvauskerhossa*. Virheellistä hakutulosta *'serranosta' kinkusta* en sijoittanut kumpaankaan näistä ryhmistä.

B2-tason sadasta ensimmäisestä hakutuloksesta substantiiviyhdyssanoja oli 90 kappaletta, esimerkiksi yhdyssanat *olkapäille* ja *omakuvia*. Olen laskenut substantiiviyhdyssanaksi myös hieman epätyypillisen *Aikaihminenä*-ilmauksen. Epätyypillisyydestään huolimatta yhdyssana voidaan sen perusosaa tarkastelemalla tulkita substantiiviyhdyssanaksi. Adjektiiviyhdyssanoja oli seitsemän, esimerkiksi yhdyssanat *lyhythihainen* ja *keski-ikäisiä*. Yhdysnumeraaleja oli kolme: *viisi-sataa*, *kahdentoista* ja *neljäkymmentäviisi*. Ilmauksessa *viisi-sataa* on virheellinen kirjoitusasu, mutta sen merkitys vastaa yhdyssanamuotoa.

Yhdyssanojen rakenteen perusteella jaettuna kaksiosaisien yhdyssanojen määrä B2-tasolla oli 96 kappaletta, joista esimerkkejä ovat *pohjatietoja* ja *sadekuuron*. Moniosaisia yhdyssanoja esiintyi neljä kappaletta, esimerkiksi *puutarhanhoito* ja *maataloustyöhön*.

5.3 C1- ja C2-tasojen yhdyssanat

C1-taitotason sadasta ensimmäisestä hakutuloksesta 81 kappaletta oli substantiiviyhdyssanoja, esimerkiksi yhdyssanat *ulkopuolinen* ja *verenpaine*. Adjektiiviyhdyssanoja taas oli 18 kappaletta, esimerkiksi *jääkylmällä* ja *helläsydäminen*. Yhdysnumeraaleja oli vain yksi, *kaksitoista*.

Kaksiosaisten yhdyssanojen määrä C1-tasolla oli 93 kappaletta. Näistä esimerkkejä ovat *ai-neenvaihduntaa* ja *sähkösaunan*. Moniosaisia yhdyssanoja oli seitsemän, esimerkiksi yhdyssanat *polkupyöräosasto* ja *sotaelokuvista*.

C2-tason sadasta ensimmäisestä hakutuloksesta substantiiviyhdyssanoja oli 90 kappaletta, esimerkiksi yhdyssanat *tavaramerkki* ja *Murtovarkaus*. Näiden joukkoon olen laskenut myös hakutuloksen *elektroniikka yritys*. Vaikka ilmaus on kirjoitettu aineistossa virheellisesti erikseen, sen merkitys vastaa yhdyssanamuotoa. Adjektiiviyhdyssanoja esiintyi kahdeksan kappaletta, esimerkiksi *kevytmielinen* ja *pitkäaikaiset*. Yhdysnumeraaleja oli vain yksi: *viisikymmentä*. Hakutuloksista yksi oli myös kokonaan virheellinen: *periaatteet*. Sana ei ole yhdyssana, joten en ole luokitellut sitä mihinkään sanaluokkakategoriaan.

Rakenteen perusteella jaettuna kaksiosaisten yhdyssanojen määrä C2-tasolla oli 87 kappaletta. Esimerkkejä näistä ovat yhdyssanat *aikakauden* ja *keskiluokan*. Moniosaisia yhdyssanoja esiintyi tällä taitotasolla 12, esimerkiksi *terveydenelintarvikkeiden* ja *maitojauheyritys*. Virheellinen hakutulos *periaatteet* on jälleen jätetty kokonaan kategorioimatta.

6 Yleisimmät yhdyssanat eri taitotasoilla

Tässä luvussa tarkastelen S2-oppijoiden kirjoittamissa teksteissä esiintyviä kymmentä yleisintä yhdyssanaa eri CEFR-taitotasoilla (A1-C2). Yleisimpien yhdyssanojen analysointi täydentää edellistä lukua, sillä se tarjoaa tarkempaa näkökulmaa siihen, kuinka monipuolisesti yhdyssanoja käytetään eri taitotasoilla. Tässä luvussa tarkasteluni kohteena ovat yhdyssanojen frekvenssit, eli niiden esiintymismäärät, sekä niiden semanttiset ja temaattiset piirteet. Olen jättänyt huomioimatta taitotasojen A1-, A2-, B1- ja B2 hakutuloksissa esiintyneen sanan *televisio*, sillä se ei ole yhdyssana. Samasta syystä C1-tasolla analyysistä on jätetty pois hakutuloksen antama sana *esimerkki*. Saamani tuloksia havainnollistava kaavio 4 löytyy luvusta 7, jossa myös analysoin tuloksia tarkemmin.

6.1 A1- ja A2-tasojen yleisimmät yhdyssanat

A1-taitotason koko korpusaineistossa esiintyi yhteensä 381 yhdyssanaa. Kymmenen eniten esiintyvää yhdyssanaa, joita oli yhteensä 78 kappaletta, kattoi 20,5 % koko aineiston yhdyssanoista. Yleisin sana *yliopisto* esiintyi 22 kertaa ja kattoi 5,8 % koko aineiston yhdyssanoista.

TAULUKKO 1. A1-tason yleisimmät yhdyssanat.

Yleisimmät A1-tasolla esiintyvät yhdyssanat	Osuus kaikista tasolla esiintyvistä yhdyssanoista (%)
yliopisto	5,8
olohuone	2,1
elokuva	2,1
aamupala	1,8
kirjoituspöytä	1,6
kirjahylly	1,6
kaksitoista	1,6
syntymäpäivä	1,3
nojatuoli	1,3
lempiruoka	1,3

A2-taitotason koko aineistossa esiintyi yhteensä 5010 yhdyssanaa. Kymmenen yleisintä yhdyssanaa muodostivat 959 kappaleella 19,2 % koko aineistosta. Yleisin sana oli *yliopisto* (232 kappaletta), joka kattoi 4,6 % kaikista aineiston yhdyssanoista.

TAULUKKO 2. A2-tason yleisimmät yhdyssanat.

Yleisimmät A2-tasolla esiintyvät yhdyssanat	Osuus kaikista tasolla esiintyvistä yhdyssanoista (%)
yliopisto	4,6
olohuone	2,2
elokuva	2,0
kylpyhuone	1,9
makuuhuone	1,7
kirjoituspöytä	1,6
vaatekaappi	1,4
kirjahylly	1,3
kaksitoista	1,3
viikonloppu	1,2

6.2 B1- ja B2-tasojen yleisimmät yhdyssanat

B1-tason aineistossa esiintyi yhteensä 20 662 yhdyssanaa, joista kymmenen yleisintä yhdyssanaa (2 982 kappaletta) kattoi 14,4 %. Yleisin yhdyssana oli jälleen *yliopisto*, joka esiintyi 741 kertaa, muodostaen 3,6 % kaikista yhdyssanoista.

TAULUKKO 3. B1-tason yleisimmät yhdyssanat.

Yleisimmät B1-tasolla esiintyvät yhdyssanat	Osuus kaikista tasolla esiintyvistä yhdyssanoista (%)
yliopisto	3,6

elokuva	2,3
olohuone	1,5
kylpyhuone	1,4
kirjoituspöytä	1,1
isoäiti	1,0
kaksitoista	0,9
vaatekaappi	0,9
makuuhuone	0,9
kirjahylly	0,9

B2-taitotason aineistossa esiintyi kaikkiaan 15 157 yhdyssanaa. Kymmenen yleisintä yhdyssanaa muodostivat 1 549 kappaleella 10,2 % aineiston yhdyssanoista. Yleisin yhdyssana oli *elokuva*, joka kattoi 2,3 % kaikista yhdyssanoista esiintyen 353 kertaa.

TAULUKKO 4. B2-tason yleisimmät yhdyssanat.

Yleisimmät B2-tasolla esiintyvät yhdyssanat	Osuus kaikista tasolla esiintyvistä yhdyssanoista (%)
elokuva	2,3
yliopisto	1,9
päähenkilö	1,4
äidinkieli	1,0
mielipide	0,7
viikonloppu	0,6
työntekijä	0,6
isäpuoli	0,6
työpaikka	0,6
yhteiskunta	0,5

6.3 C1- ja C2-tasojen yleisimmät yhdyssanat

C1-taitotason aineistossa esiintyi kokonaisuudessaan 5 148 yhdyssanaa. Kymmenen yleisintä yhdyssanaa kattoi 10,6 % 544 kappaleella. Yleisin yhdyssana *äidinkieli* esiintyi 149 kertaa ja kattoi 2,9 % koko aineiston yhdyssanoista.

TAULUKKO 5. C1-tason yleisimmät yhdyssanat.

Yleisimmät C1-tasolla esiintyvät yhdyssanat	Osuus kaikista tasolla esiintyvistä yhdyssanoista (%)
äidinkieli	2,9
elokuva	1,6
yliopisto	0,9
lähdekieli	0,8
kohdekieli	0,8
kielioppi	0,8
päähenkilö	0,8
mielipide	0,7
puhekieli	0,7
oppijankieli	0,6

C2-tason aineistossa esiintyi yhteensä 1 141 yhdyssanaa. Kymmenen yleisintä muodostivat 173 kappaleella 15,2 % kaikista aineiston yhdyssanoista. Yleisin yhdyssana *äidinkieli* esiintyi 45 kertaa ja kattoi 3,9 % kaikista yhdyssanoista.

TAULUKKO 6. B2-tason yleisimmät yhdyssanat.

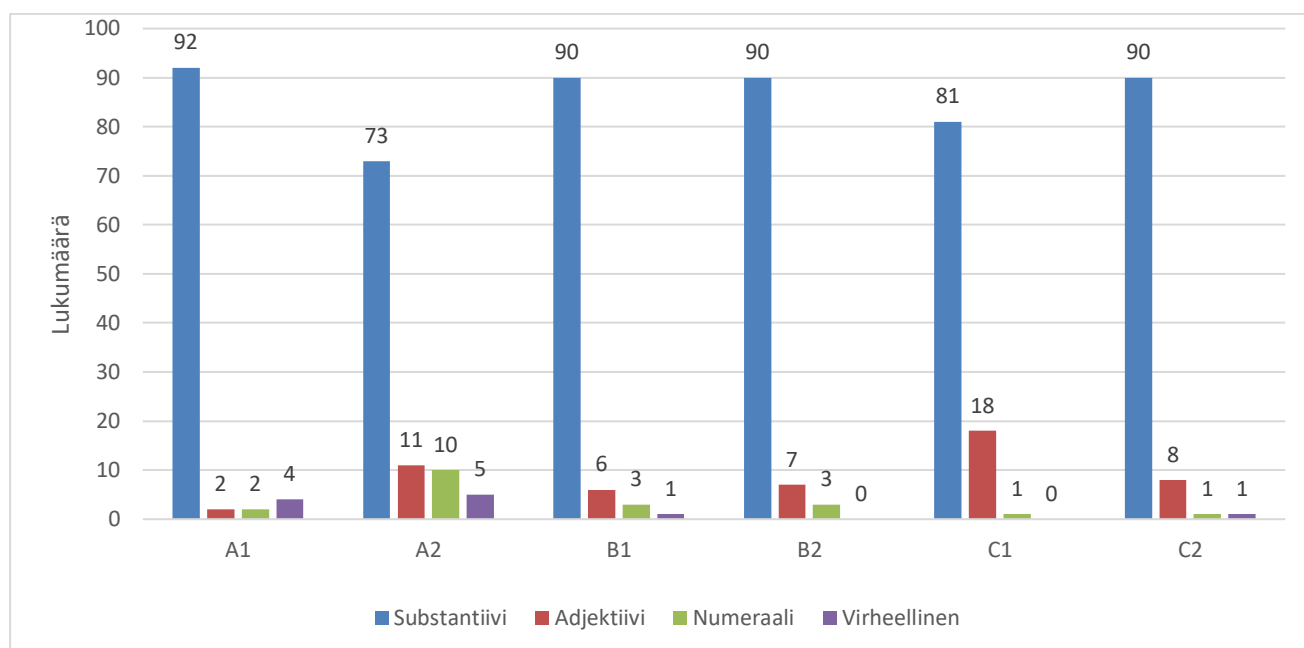
Yleisimmät C2-tasolla esiintyvät yhdyssanat	Osuus kaikista tasolla esiintyvistä yhdyssanoista (%)
äidinkieli	3,9
isäpuoli	2,5
kielitaito	1,8
tekstilaji	1,5

kielenopetus	1,4
näkökulma	0,9
ulkonäkö	0,8
oppikirja	0,8
avioero	0,8
kaksikielinen	0,7

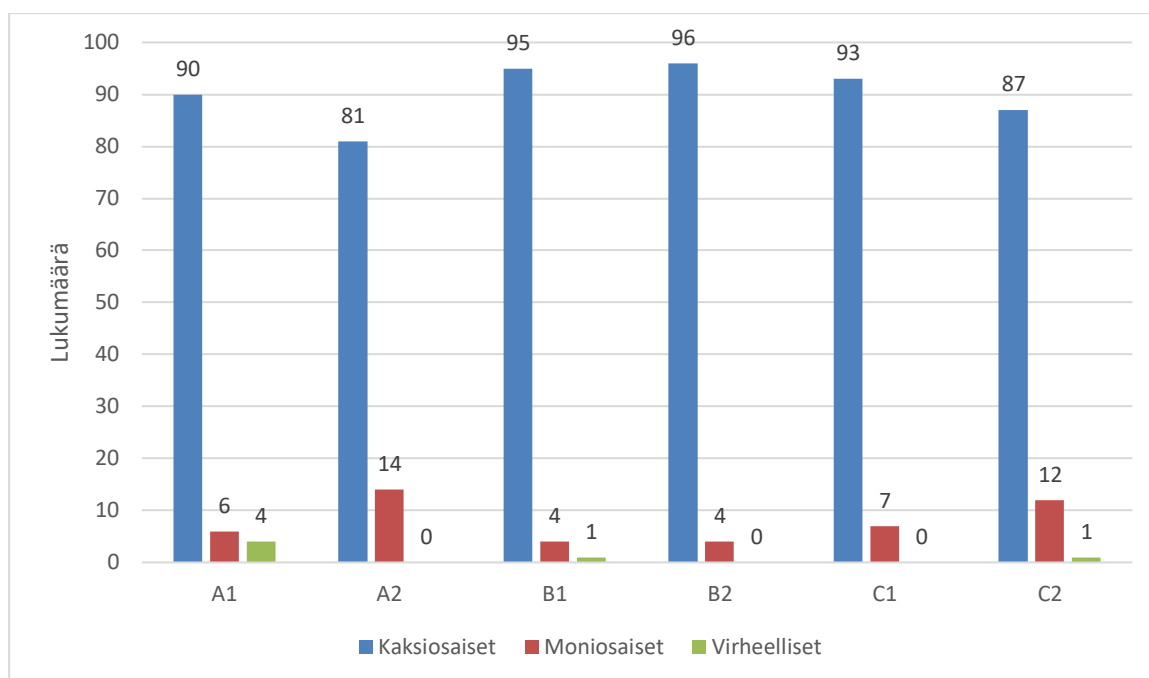
7 Yhteenveto ja päätelmät

Tässä tutkielmassa tarkastelin yhdyssanojen käyttöä eri CEFR-taitotasolla olevien suomen kielenä -oppijoiden kirjoittamissa teksteissä. Tavoitteenani oli selvittää, kuinka paljon yhdyssanoja S2-oppijoiden kirjoittamissa teksteissä esiintyy, millaisia ne ovat ja miten niiden käyttö vaihtelee taitotasoin. Aineistoni koostui Kansainvälisen oppijansuomen korpuksen teksteistä, joita analysoin sekä kvantitatiivisesti että kvalitatiivisesti.

7.1 Päätelmät tutkimustuloksista



KUVIO 2. Yhdyssanojen sanaluokat taitotasoin.



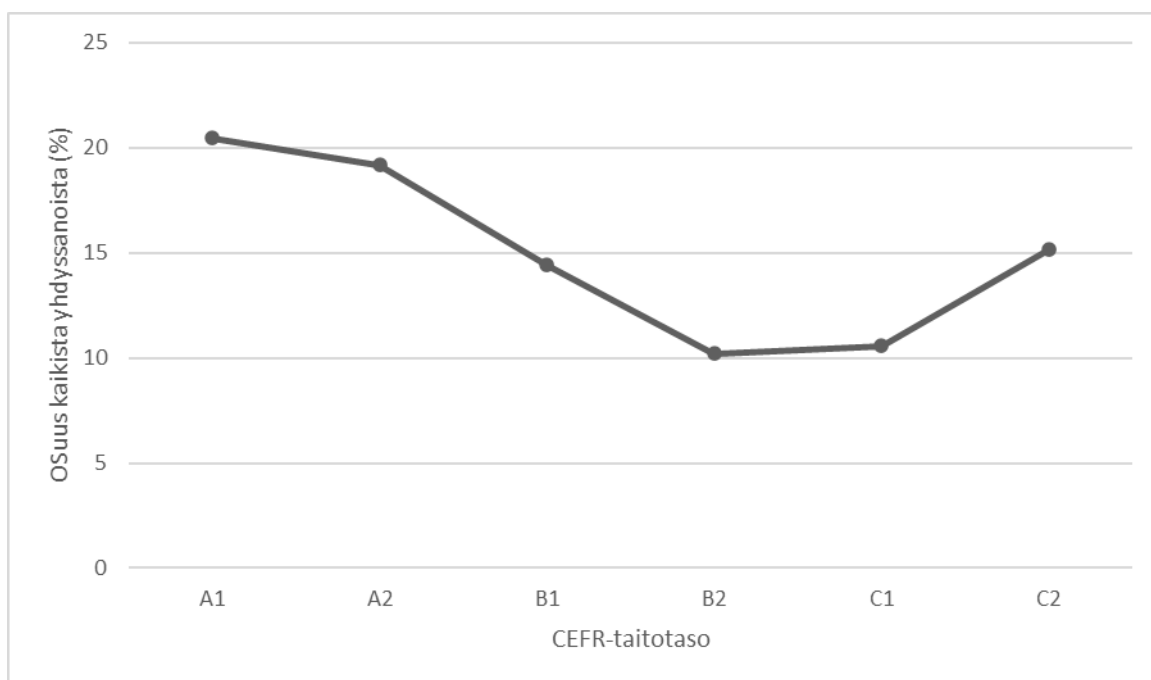
KUVIO 3. Yhdyssanojen rakenteet taitotasoin.

Luvussa 5 esittämieni tulosten perusteella S2-opiskelijoiden teksteissä esiintyvien yhdyssanojen sanaluokissa ja rakenteellisessa kompleksisuudessa on vaihtelevuutta CEFR-tason mukaan. Kaikilla tasoilla kuitenkin substantiiviyhdyssanat kattavat selkeästi suurimman osan tutkituista yhdyssanoista. Substantiiviyhdyssanat muodostavatkin suomen kielen suurimman yhdyssanatyypin (ISK 2004, 388). Tämän perusteella substantiiviyhdyssanojen hallitsemista voidaan pitää erityisen tärkeänä jo oppimisen varhaisessa vaiheessa.

Adjektiiviyhdyssanojen määrä vaihteli paljon taitotason mukaan. A1- ja A2-tasoilla adjektiiviyhdyssanoja oli yhteensä 13 kappaletta. Myös B1- ja B2-tasoilla adjektiiviyhdyssanoja oli yhteensä 13 kappaletta. C1- ja C2-tasoilla adjektiiviyhdyssanojen yhteenlaskettu määrä nousi 26:een. Näin tarkasteltuna adjektiiviyhdyssanojen määrän kasvu ylemmillä taitotasoin voi kuvastaa oppijoiden kasvavaa kykyä tuottaa kuvailevia ja tarkentavia ilmauksia. Yhdysnumeraaleja esiintyi tasaisen vähän kullakin taitotasolla. Ainoastaan A2-tasolla yhdysnumeraaleja oli poikkeavasti 10 kappaletta, mutta pidän tätä sattumana aineiston rajauksen osalta. Yhdysnumeraalien käytön vähäisyys voi johtua esimerkiksi siitä, että ne on mahdollista korvata käyttämällä numeroita kirjoitetun muodon sijaan.

Rakenteellisesti kaksiosaiset yhdyssanat olivat moniosaisia yhdyssanoja suurempi ryhmä kaikilla taitotasoin. Moniosaisen yhdyssanojen käyttö kuitenkin tuplaantui A1-tasolla esiintyvistä kuudesta kappaleesta C2-tason 12 kappaleeseen. Tämä kehitys kuvastaa oppijoiden

kasvavaa kykyä tuottaa rakenteellisesti monimutkaisempia ilmauksia. Virheellisiä tai epäselviä ilmauksia esiintyi lähinnä A- ja B-tasoilla. C-tasoilla yhdyssanojen tuottaminen oli selkeästi tarkempaa.



KUVIO 4. Kymmenen yleisimmän yhdyssanan osuus (%) kaikista yhdyssanoista CEFR-taitotasolla.

Luvussa 6 esitetyistä tuloksista voidaan havaita, että sekä A1- että A2-tasolla kymmenen eniten esiintyvää yhdyssanaa kattoi noin viidesosan koko aineiston yhdyssanoista. Yleisimmät yhdyssanat A1- ja A2-taitotasolla viittasivat pääosin arkisiin ja konkreettisiin ilmiöihin, kuten huonekaluihin, ruokailuun ja asumiseen. Kaikki yleisimmät yhdyssanat olivat kaksiosaisia. Yksinkertaisten yhdyssanojen käytön alemmilla taitotasolla voidaan tulkita kuvastavan oppijoiden rajoittunutta sanavarastoa.

B1- ja B2-tasoilla yhdyssanojen käyttö hajautui laajemmalle. Kymmenen yleisintä yhdyssanaa kattoivat pienemmän osuuden aineistojen kokonaisyhdyssanamäärästä ja yleisimpien yhdyssanojen joukkoon mahtui alempiin taitotasoihin verrattuna hieman abstraktimpia yhdyssanoja, esimerkiksi *mielipide* ja *äidinkieli*.

Korkeampien taitotasojen C1 ja C2 yhdyssanat painoutuivat metakielellisiin, eli kieltä kuvaileviin, ja akateemisiin käsitteisiin. Päättelen tämän kuvastavan edistyneiden oppijoiden reflektiivistä kielenkäyttöä. C1-tasolla yhdyssanojen käyttö hajaantui edelleen alempiin taitotasoihin verrattuna. C2-tasolla kymmenen yleisimmän yhdyssanan muodostama prosenttiosuus

jälleen kasvoi, mutta tulkiten sen liittyvän huomattavasti suppeampaan aineistoon verrattuna esimerkiksi tasoon C1.

Saamani tulokset osoittavat, että yhdyssanojen käyttö lisääntyy taitotason noustessa. Yhdyssanojen määrän kehitys ei kuitenkaan ole suurta, eikä se etene lineaarisesti taitotasoin. Selkeän kehityksen yhdyssanojen määrässä huomaa kuitenkin hyvin verratessa vain alinta (A1) ja ylintä (C2) taitotasoa keskenään. Tämä tukee tutkimukseni alkuperäistä hypoteesia siitä, että yhdyssanojen käyttö lisääntyy korkeammilla taitotasolla olevien oppijoiden teksteissä. S2-oppijoiden muun leksikaalisen ja morfologisen kompetenssin kehittyminen heijastuu siis myös yhdyssanojen käyttöön.

Yhdyssanojen sanaluokkajakauma oli samankaltainen kaikilla tasoilla, sillä substantiiviyhdys-sanat muodostivat selkeästi suurimman ryhmän. Adjektiiviyhdyssanojen määrän kasvu erityisesti ylimmillä C1- ja C2-tasoilla viittaa oppijoiden kasvaneeseen kykyyn tuottaa kuvailevampaa ja tarkentavampaa kieltä. Myös rakenteellisesti tarkasteltuna yhdyssanoissa ei juurikaan ollut vaihtelevuutta, sillä ne olivat kaikilla taitotasolla suurimmaksi osaksi kaksiosaisia. Moniosaisen yhdyssanojen määrä kuitenkin kasvoi, ja jälleen erityisesti alinta ja ylintä taitotasoa verrattaessa kasvu on selkeä.

Kymmenen yleisimmän yhdyssanan tarkastelu kultakin taitotasolta osoitti, että alemmilla taitotasolla yhdyssanat koostuivat pääosin arkisista ja konkreettisia merkityksiä sisältävistä sanoista, kun ylemmillä tasoilla esiintyi enemmän abstrakteja yhdyssanoja. Tämä havainto tukee sekä omaa hypoteesiaani että aiempaa tutkimusta, jonka mukaan rakenteellinen ja leksikaalinen kompleksisuus korreloivat positiivisesti kielitaidon tason kanssa (ks. Bulte – Roothoof 2020, luku 5). Yleisimpien yhdyssanojen osuuden pieneneminen verrattuna kaikkiin tason aineistossa esiintyneisiin yhdyssanoihin viittaa muun muassa oppijoiden sanavaraston laajenemiseen.

7.2 Lopuksi

Tutkimukseni sisältää tiettyjä rajoituksia. Eri taitotasojen aineistojen koko vaihteli paljon. Analyysini perustui korpuksen hakuehtoihin, jotka eivät tavoittaneet kaikkia mahdollisia virheelliseen muotoon tai erikseen kirjoitettuja yhdyssanoja. Lisäksi analyysini rajautui vain saamaan ensimmäiseen hakutulokseen kullakin taitotasolla, mikä vaikuttaa tulosten yleistettävyyteen.

Mahdollisessa jatkotutkimuksessa yhdyssanojen käyttöä S2-oppijoiden teksteissä voisi tarkastella esimerkiksi juuri laajemmalla aineistolla. Tutkimusta olisi mahdollista syventää esimerkiksi ottamalla huomioon aineiston tekstien tekstilajeja tai keräämällä pitkittäisaineistoa yksittäisten oppijoiden yhdyssanojen käytön kehityksestä. Lisäksi yhdyssanojen käytössä esiintyvien virheiden ja virhetyyppien kehittymisen tutkiminen voisi tarjota hyödyllistä tietoa suomena toisena kielenä opiskelevien oppimisprosessista.

Tämä tutkimus osoittaa, että yhdyssanat tarjoavat informatiivisen ja käyttökelpoisen näkökulman suomena toisena kielenä -oppijoiden kielenkehityksen tarkasteluun. Yhdyssanojen määrän, rakenteen ja semanttisen luonteen muutokset heijastuvat S2-oppijoiden kehittyvään kielitaitoon ja ovat todiste yhdyssanojen keskeisestä roolista suomen kielessä ja sen opetuksessa, sekä kielitaidon arvioinnissa.

Lähteet

Aineslähteet:

City Digital Group 2025. *ICLFI – Kansainvälinen oppijansuomen korpus* [Tekstikorpus].

Kielipankki. Viitattu 26.1.2026. Saatavissa: <https://www.kieli-pankki.fi/korp/#?cqp=%5B%5D>.

Tutkimuskirjallisuus:

s. Bachman, Lyle 1990. *Fundamental Considerations in Language Testing*. Oxford University Press

Bulte, Bram – Roothoof, Hanne 2020. Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System*, 91, 102246. DOI: <https://doi.org/10.1016/j.system.2020.102246>.

Council of Europe 2026. The Common European Framework of Reference for Languages (CEFR). <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>

Eronen, Riitta – Maamies, Sari – Räikkälä, Anneli 1996. Yhdyssanat. *Kielikello* 4/1996. Saatavissa: <https://kielikello.fi/yhdyssanat/>.

Hongisto, Paula 2023. Tekstitaidot: Kielenhuolto - Oikeinkirjoitus: yhdyssanat (PDF). *Kielibuusti.fi*. Viitattu 5.1.2026. Saatavissa: <https://www.kielibuusti.fi/sites/default/files/2023-07/Yhdyssanat.pdf>.

Hakulinen, Auli - Vilkuna, Maria - Korhonen, Riitta - Koivisto, Vesa - Heinonen, Tarja Riitta - Alho, Irja 2004: *Iso suomen kielioppi*.

Masonen, Virpi 2013. Leksikaalinen kompetenssi osana kielitaitoa. *Lähivõrdlusi. Lähivertailuja* 23, 187–209. DOI: <https://doi.org/10.5128/LV23.08>.

Penttinen, Kati 2010. Voisitko apua? Suomi toisena kielenä -oppijoiden sananmuodostustaitojen jäljillä. Pro gradu -tutkielma. Jyväskylän yliopisto. URN: <https://jyx.jyu.fi/bitstreams/73cc9162-c7f8-4a1a-b68b-6066cfbec01a/download>.

Rusava, Vilja 2025. Aikuisten S2-oppijoiden suomen kielen morfologinen prosessointi ja siihen liittyvät morfofonologiset ja -syntaktiset haasteet sekä L1-kielten vaikutukset. Kandidaatintutkielma. Turun yliopisto. https://www.utupub.fi/bitstream/handle/10024/182419/Vilja_Rusava_Kandidaatintutkielma.pdf?sequence=1&isAllowed=y.

Liitteet

Korpus-työkalun tarkka hakulause:

Räätälöity CQP-haku:

```
[lemma = ".+#+." &_.text_cefrlevel = "A1"]
```