Jimi Uitti

Kielimallien soveltaminen suomen kielen oppimisessa

Suomen kielen morfologian osaaminen

Tekniikka ja innovaatiojohtaminen
Diplomityö
Automaatio- ja tietotekniikka

Vaasa 2025

VAASAN YLIOPISTO
Tekniikka ja innovaatiojohtaminen
Tekijä: Jimi Uitti
Tutkielman nimi: Kielimallien soveltaminen suomen kielen oppimisessa: Suomen kielen morfologian osaaminen
Tutkinto: Diplomi-insinööri
Oppiaine: Automaatio- ja tietotekniikka
Työn ohjaaja: Tomi Pasanen
Valmistumisvuosi: 2025 **Sivumäärä:** 84

TIIVISTELMÄ:

Kielimallit ovat erittäin tehokkaita tekoälyjä, jotka pystyvät tekstisyötteen eli kehotteen avulla tuottamaan kaikenlaista. Kielimallin tuotto voi olla mitä tahansa tekstipohjaista, kuten esimerkiksi ohjeita, tietoa tai keskustelua. Tähän kuuluu esimerkiksi kielten opettaminen, jota on englannin kielellä paljon tutkittu. Kielimallien käyttö on pääasiassa keskitetty englanninkieliseen käyttöön. Suomen kielen kanssa ei ole kielimalleja yhtä paljon tutkittu, eikä suomen kieltä ole käytetty yhtä paljoa niiden kouluttamisessa. Tässä tutkimuksessa selvitetään, kuinka hyvin kielimallit ymmärtävät suomen kielen morfologiaa eli taivuttamista. Tätä tutkitaan testaamalla kielimallien kykyä taivuttaa sanoja itsenäisesti sekä lauseissa. Tutkimus keskittyy suomen kielen opettamiseen englanninkielisille, sillä Suomessa asuvilla ulkomaalaisilla on hyvä olla useita eri vaihtoehtoja suomen kielen oppimista varten.

Työ suoritetaan tutkimalla kielimallien suomen kielen morfologian ymmärtämistä sanojen sekä lauseiden kannalta. Työssä testataan pilvipohjaisia suuria kielimalleja sekä paikallisesti käyttöönotettavia pieniä kielimalleja. Sanojen kanssa kielimallit taivuttavat sanoja itsenäisesti perusmuodosta erilaisiin taivutusmuotoihin. Lisäksi myös taivutetaan taivutusmuodoista perusmuotoihin. Lauseiden suhteen kielimallien täytyy korjata lauseita, joissa kaikkien sanojen taivutukset on vaihdettu perusmuotoihin. Kielimalleille annetaan kehotteet erikseen suomeksi ja englanniksi, jotta voidaan kokeilla, vaikuttaako kielimallin kanssa käytetty kieli sen vastauksen laatuun. Kehotteet myös annetaan joko esimerkkivastauksen kanssa tai ilman, jotta voidaan selvittää esimerkin antamisen vaikutuksen kielimallien vastausten laatuun. Kielimallien antamia tuloksia verrataan virallisten lähteiden, kuten esimerkiksi Kielitoimiston sanakirjan kanssa.

Tutkimusten tulosten mukaan suuret kielimallit pystyvät parhaimmillaan taivuttamaan suomenkielisiä sanoja ilman suuria virheitä. Pienet kielimallit onnistuivat heikommin, mutta nekin parhaimmillaan toimivat yksinkertaisten ja helppojen sanojen taivuttamisessa. Perusmuotoihin taivuttaminen oli kielimalleille helpompaa kuin perusmuodoista taivuttaminen, mutta eroavaisuus ei ollut suuri. Lauseidenkin kanssa kielimallit onnistuivat korjaamaan taivutuksia. Englanninkielinen esimerkkivastauksen sisältävä kehotepohja tuotti useimmiten parhaimmat ja luotettavimmat tulokset. Muutkin kehotepohjat toimivat vähintään suurten kielimallien kanssa tyydyttävästi. Kielimallit voivat tuottaa silti virheitä, joten niiden vastauksiin ei voi aina luottaa. Kielimallit ymmärtävät suomen kielen morfologiaa vähintään aloittelijatasolla. Kokeneempien suomen kielen oppijoiden kannattaa kuitenkin käyttää muita lähteitä suomen kielen oppimisessa.

AVAINSANAT: kielimallit, luonnollisen kielen käsittely, kehotteet (tekoäly), suomen kieli, muoto-oppi (kielitiede).

Sisällys

1 Johdanto 6

2 Kielimallien teoria 9

2.1 Suuret kielimallit 9

2.1.1 ChatGPT 11

2.1.2 Gemini 12

2.1.3 Copilot 13

2.2 Pienet kielimallit 13

2.2.1 Llama 15

2.2.2 Gemma 16

2.2.3 DeepSeek 16

2.3 Kielimallien eettisyys ja hallusinaatio 17

3 Kielimallien käyttö kielten oppimisessa 19

3.1 Generoiva käyttö 19

3.2 Keskusteleva käyttö 22

3.3 Arvioiva käyttö 25

3.4 Kääntävä käyttö 27

3.5 Suomen kielen oppiminen 30

4 Tutkimusmenetelmä 33

4.1 Sanojen taivuttamisen osaamisen testaaminen 33

4.2 Lauseiden taivutusten korjaamisen testaaminen 35

4.3 Arviointiprosessi 37

4.4 Testatut suuret kielimallit 38

5 Kielimallien sanojen taivutustaidot 40

5.1 Perusmuodoista taivutusmuotoihin 40

5.1.1 Verbien taivuttaminen 43

5.1.2 Substantiivien taivuttaminen 45

5.1.3 Adjektiivien taivuttaminen 46

5.1.4 Taivuttamisen tulosten yhteenveto 47

5.1.5 Perusmuotojen kääntäminen suomesta englanniksi 48
5.2 Taivutusmuodoista perusmuotoihin 49
5.2.1 Perusmuotojen selvittämisen tulokset 50
5.2.2 Taivutusten muodontunnistus 51
5.2.3 Taivutusten kääntäminen suomesta englanniksi 52
5.3 Kielimallien sanojen taivuttamisessa tekemät virheet 53
5.3.1 Perusmuodosta taivuttamisen virheet 53
5.3.2 Perusmuotoihin taivuttamisen virheet 56
6 Kielimallien lauseiden taivutusvirheiden korjaamistaidot 58
6.1 Helpompien lauseiden korjaaminen 58
6.2 Vaikeampien lauseiden korjaaminen 60
6.3 Tulosten yhteenveto 62
6.4 Kielimallien lauseiden taivutuskorjaamisessa tekemät virheet 63
7 Johtopäätökset ja pohdintaa 67
Lähteet 70
Liitteet 79
Liite 1. Perusmuodosta taivuttamisen kehotepohjat 79
Liite 2. Perusmuotoon taivuttamisen kehotepohjat 82
Liite 3. Lauseiden korjaamisen kehotepohjat 84

Kuviot

Kuvio 1. Transformer-arkkitehtuurin vuokaavio (Jadhav, 2025) 11

Kuvio 2. Verbien perusmuotojen taivuttamisen tulokset 44

Kuvio 3. Substantiivien perusmuotojen taivuttamisen tulokset 46

Kuvio 4. Adjektiivien perusmuotojen taivuttamisen tulokset 47

Kuvio 5. Perusmuotojen taivuttamisen yhteistulokset 48

Kuvio 6. Perusmuotojen kääntämisen yhteistulokset 49

Kuvio 7. Perusmuotoon taivuttamisen tulokset 51

Kuvio 8. Taivutusten muodontunnistuksen tulokset 52

Kuvio 9. Taivutusten englannin kieleen kääntämisen tulokset 53

Kuvio 10. Helppojen lauseiden taivutuskorjaamisen tulokset 60

Kuvio 11. Vaikeiden lauseiden taivutuskorjaamisen tulokset 62

Kuvio 12. Lauseiden taivutuskorjaamisen yhteistulokset 63

Taulukot

Taulukko 1. Sanomalehtikielen yleissanastosta valitut sanat ja niiden yleisyydet 42

Taulukko 2. Perusmuodoista taivutettavat verbit taivutusmuotoineen 42

Taulukko 3. Perusmuodoista taivutettavat substantiivit taivutusmuotoineen 43

Taulukko 4. Perusmuodoista taivutettavat adjektiivit taivutusmuotoineen 43

Taulukko 5. Perusmuotoihin taivutettavat sanat 50

Taulukko 6. Sanan "todeta" taivuttamisen vastaukset ChatGPT-4o-mallilta 54

Taulukko 7. Sanan "tehdä" taivuttamisen vastaukset Llama 3.1 8B -mallilta 55

Taulukko 8. Sanan "riitti" taivuttaminen DeepSeek-r1 14B -mallilla 56

Taulukko 9. Sanan "riitteihin" perusmuotoon taivuttaminen ChatGPT-4o-mallilla 57

Taulukko 10. Taivutuskorjaamisen helpommat lauseet 59

Taulukko 11. Taivutuskorjaamisen vaikeammat lauseet 61

Taulukko 12. Pienten kielimallien taivutuskorjausesimerkki 66

1 Johdanto

Tekoälyt ovat kehittyneet lähiaikoina hyvin nopealla vauhdilla, ja uusia mahdollisuuksia niiden käyttötavoille tulee esille jatkuvasti. Erilaisista tekoälymalleista esimerkiksi suuret kielimallit (englanniksi *Large Language Model*) ovat kehittyneet varsinkin nopeasti. Nämä kielimallit on tehty käsittelemään hyvin erilaisia keskusteluja. Ne voivat esimerkiksi vastata kysymyksiin, tuottaa laadukasta tekstiä, arvioida syötettyä tekstiä, tai jopa kääntää tekstiä eri kielille. Kielimallit on koulutettu pääasiassa englanninkielisillä tekstiaineistoilla. Monet suuriin kielimalleihin liittyvät tutkimukset keskittyvät pääasiassa englanninkieliseen toimintaan, mutta esimerkiksi suomenkielistä toimintaa ei ole tutkittu yhtä paljon.

Suomen kieli on englannin kieleen verrattuna melko harvinainen kieli. Englannin kieltä puhuu maailmalla mahdollisesti jopa 1,5–2 miljardia ihmistä (Crystal, 2008, s. 5). Suomen kieltä puhuvien määrä maailmalla on puolestaan arviolta alle kuusi miljoonaa ihmistä (Kotus, 2024). Jos oletetaan, että tietynkielisten tekstien määrä vastaisi puhujien määrää, niin suomenkielisen tekstin määrä olisi tällöin noin 0,3–0,4 prosenttia englanninkielisen tekstin määrästä. Kielimallien tehokkuuteen ja toimivuuteen vaikuttaa niiden tietokannoissa olevien tietojen määrä. Suomenkielisen tekstin vähäisyys vaikuttaa kielimallien suomenkielisten vastausten laatuun negatiivisesti.

Morfologia eli muoto-oppi tarkoittaa kielitieteessä Tieteen termipankin (2025) mukaan sanojen sisäisen rakenteen osajärjestelmää. Morfologian mukaan sanat koostuvat morfeemeiksi kutsutuista kielen pienimmistä muotoaineksista, joilla on jonkinlainen merkitys. Morfologia on syntaksin eli lauseopin kanssa yksi kieliopillisen järjestelmän peruskomponenteista. Dahl (2008, s. 547) kertoo klassisen morfologisen typologian mukaan kielen jakautuvan rakenteen mukaisesti isoloiviin, agglutinoiviin, flekteeraaviin sekä polysynteettisiin kieliin. Näistä olennaisimmat suomen kielen kanssa ovat hänen mukaansa agglutinoiva sekä flekteeraava kielirakenne. Agglutinoivassa kielessä liitetään affikseja selkeästi sanavartalon loppuun. Flekteeraavassa ne puolestaan sulautuvat yhteen, jolloin sanavartalo muuttuu. Suomen kielessä on kummankin kielirakenteen piirteitä, sillä jotkin

sanat voivat taipua täysin agglutinoivasti tai osittain myös flekteeraavasti. Agglutinoivan sanan pohja ei muutu taivuttaessa lainkaan (esim. **talo** -> **talo**ssa), kun taas flekteeraavan sanan pohja puolestaan muuttuu (esim. ve**si** -> ve**de**ssä). Englannin kieli ei taivuta suomen kieleen verrattuna yhtä monimuotoisesti, sillä esimerkiksi sijainti kerrotaan taivutusten sijasta prepositioiden avulla (Vannest ja muut, 2008, s. 84).

Suurin osa kielimalleista on keskittynyt englanninkielisiin tehtäviin, jolloin ne eivät onnistu yhtä hyvin monikielisissä tilanteissa (Qin ja muut, 2025, s. 1). Suomenkielisen tekstiaineiston määrä on englanninkieliseen verrattuna vähäinen, ja suomen kieli eroaa englannin kielestä suuresti esimerkiksi taivuttamisessa. Kielimalleilla voi näiden takia olla suomen kielen kanssa vaikeuksia. Kuinka hyvin onnistuisi kielimallien soveltaminen suomen kielen oppimisessa näistä vaikeuksista huolimatta? Osaisivatko kielimallit taivuttaa suomenkielisiä sanoja tai korjata taivutusvirheitä lauseissa? Vai tekevätkö tekstiaineistojen puute sekä kielen ainutlaatuisuus tästä kielimalleille mahdottoman tehtävän?

Tutkimuksen päämääränä on selvittää ja testata erilaisia keinoja käyttää suuria kielimalleja suomen kielen taivuttamisen opettamisessa. Tutkimuksen tarkoituksena on tarkentaa kielimallien suomen kielen osaamista morfologian eli taivuttamisen kannalta. Tutkimuksen tavoitteet ovat selvittää:

- Osaisivatko erilaiset suuret kielimallit tällä hetkellä taivuttaa suomenkielisiä sanoja tai korjata taivutuksia luotettavasti?
- Voisivatko aloittelevat englanninkieliset suomen kielen aloittelijat oppia oikein suurten kielimallien antamista taivutuksista?
- Onnistuisivatko pienetkin kielimallit suomen kielen taivuttamisessa?

Kielimallit kehittyvät jatkuvasti, joten viime vuosina kerätyt tulokset voivat olla jo vanhentuneita. Kielimallit ovat keskittyneet pääasiassa englanninkieliseen toimintaan, eivätkä ne mahdollisesti aina osaa harvinaisempia kieliä. Taivutukset ovat olennainen osa suomen kielen kielioppia. Jos kielimallit eivät ymmärrä suomen kielen morfologiaa, niin

ne eivät voi luotettavasti auttaa suomen kielen oppimisessa. Kielimallit voisivat olla yksi hyvä keino ulkomaalaisten suomen kielen oppimisen auttamisessa. Tilastokeskuksen (2025) mukaan vuonna 2024 yhteensä 583 tuhatta Suomen väestöstä oli syntynyt ulkomailla ja 610 tuhannella äidinkieli on vieraskieli. Lisäksi vuosien 2021–2024 välillä vieraskielisten määrä nousi joka vuosi noin 50 tuhannella. Suomessa elämisen helpottamista varten heidän olisi hyvä oppia vähintään suomen kielen perusteet. Tässä tutkimuksessa keskitytään suomen kielen taivutusten perusteiden opettamiseen englanninkielisille. Pienet kielimallit ovat suuriin kielimalleihin verrattuna ympäristöystävällisempiä, mutta niiden pienemmän koon takia ne saattavat tuottaa enemmän virheellisiä vastauksia.

Ensimmäisenä luvussa 2 käydään läpi kielimalleihin liittyvää teoriaa, johon kuuluu suurten sekä pienten kielimallien toiminnan selittäminen. Luvussa 3 avataan aikaisempia tutkimuksia kielimallien käytöstä kieliopissa sekä suomen kielen oppimisesta. Luvussa 4 esitellään tutkimusmenetelmät, joiden mukaan kielimalleja testattiin. Sen jälkeen luvuissa 5 ja 6 kerrotaan testauksista saadut tulokset. Lopuksi luvussa 7 selitetään tutkimuksen johtopäätökset, pohditaan kielimallien käyttökelpoisuutta, sekä annetaan mahdollisia jatkotutkimusaiheita.

2 Kielimallien teoria

Tässä luvussa käsitellään tämän työn keskeiset käsitteet. Ensin suurien kielimallien toiminta, kuten transformer-arkkitehtuuri, esitetään. Sitten kerrotaan muutamasta eri suuresta kielimallista, joita tullaan myös testaamaan tässä työssä. Lisäksi selitetään kehotesuunnittelu esimerkkien avulla. Lopuksi myös käydään läpi erilaisia ongelmia ja haasteita kielimalleihin liittyen, kuten eettisyys ja ympäristöystävällisyys.

2.1 Suuret kielimallit

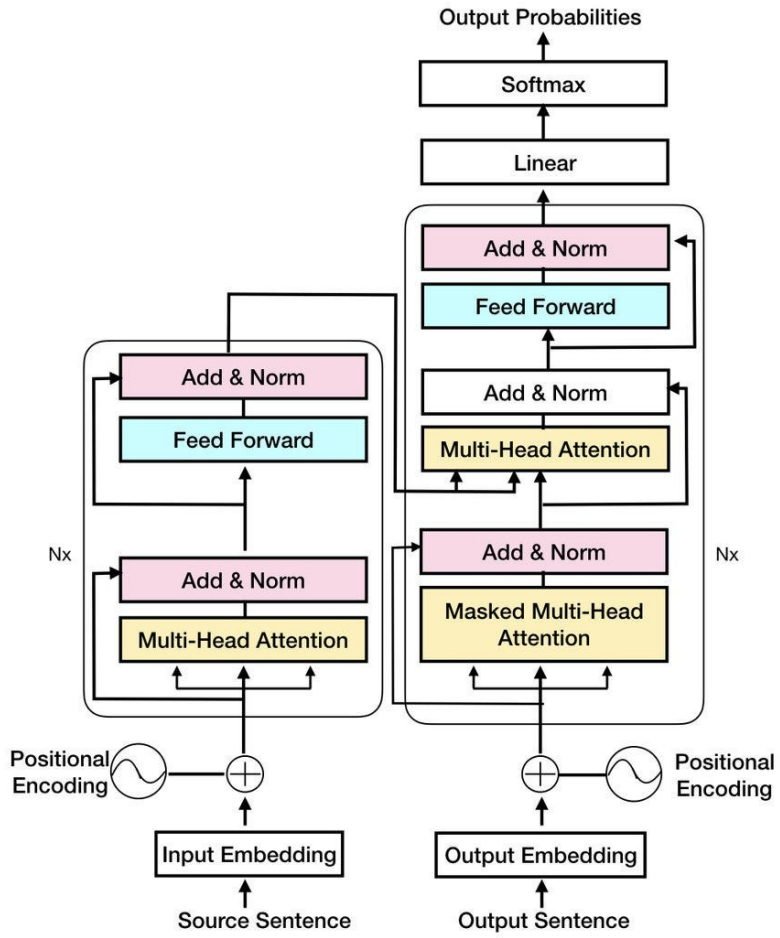
Suuret kielimallit perustuvat luonnollisiin kielimalleihin, jotka ovat ihmisten kielen ymmärtämistä ja tuottamista varten kehitettyjä tekoälymalleja (Haukka, 2024, s. 8). Nämä transformer-pohjaiset kielimallit ovat johtaneet suuriin edistyksiin luonnollisen kielen käsittelyn tutkimuksissa (Greco ja muut, 2024, s. 865). Kielimallit, kuten ChatGPT, ovat Brigantin (2024, s. 1566) mukaan pohjaltaan syväoppivia neuraaliverkkoja. Nämä syväoppivat neuraaliverkot ovat matemaattisia malleja, joiden tarkoitus on yrittää kopioida ihmisten aivojen toimintaa. Briganti selittää, että neuraaliverkko koostuu useista neuroniyksiköistä, jotka ovat yhteydessä toisiinsa painoiksi kutsuttujen yhteyksien avulla. Suuret kielimallit, kuten esimerkiksi GPT-mallit, ovat erityisesti hyviä luonnollisen kielen ymmärtämisessä sekä tuottamisessa (Yenduri ja muut, 2024, s. 2). Kielimallit pystyvät ymmärtämään tekstin tarkoituksen ja tunnistamaan kokonaisuuksia ja suhteita virkkeissä. Ne pystyvät myös kirjoittamaan luovaa sisältöä sekä vastaamaan kysymyksiin selkeästi.

Kielimallit toimivat syötteiden ja koulutustietojen kanssa Alin ja muiden (2024, s. 1) mukaan jakamalla tiedon tokeneiksi. Yksi tokeni voi kuvastaa osamerkkijonona kirjainta, tavua, tai sanaa. Sanojen ja lauseiden muuttaminen tokeneiksi tehdään tokenisoijan avulla. Rajaraman ja muut (2024, s. 1) kertovat tokenisoijan muuntavan merkkijonon peräkkäisiksi ”token id” -merkkiarvoiksi, jotka kohdistavat eri osamerkkijonoihin. Tokenisoija täytyy ideaalisesti kouluttaa vektorimuuntajan kanssa yhdessä päästä päähän -tarkkuuden

takia. Tästä syystä tokenisoija yleensä koulutetaan osalla samasta tietokannasta kuin itse kielimalli.

Kielimallit perustuvat erilaisiin syväoppimismenetelmiin. Olennaisin näistä on transformer-arkkitehtuuri, jota voidaan myös kutsua vektorimuuntajaksi. Vektorimuuntaja on edistynyt neuraaliverkko, joka on aikaisempiin neuraaliverkkomalleihin verrattuna paljon skaalautuvampi ja tehokkaampi sekventiaalisen tiedon käsittelyssä (Alomar ja muut, 2024, s. 3). Vektorimuuntajan keskeisin osa on eräänlainen ”self-attention”-mekanismi, joka pystyy vertailla ja yhdistellä syötesarjan kaikkia osia keskenään (Haukka, 2024, s. 10). Tämä mekanismi arvioi sanojen tärkeyden ja laskee sopivat syötteen edustajat syntaktisten ja semanttisten yhteyksien perusteella (Greco ja muut, 2024, s. 867). Nämä edustajat sitten painotetaan "attention function" -laskennan avulla vastauksessa annettavien kirjaimien ennustamista varten. Ennustaminen tehdään kielimalleissa autoregressiivisellä eli aikaisempia havaintoja hyödyntävällä tavalla.

Kuviossa 1 on Jadhavin (2025) tekemä vuokaavio, jossa visualisoidaan transformer-arkkitehtuuri eli vektorimuuntaja. Briganti (2023, s. 2–3) selittää omassa paperissaan yksinkertaisesti suuren kielimallin ja siten myös vektorimuuntajan toiminnan. Kielimallien kouluttamisessa neuraaliverkolle, jolla on useita miljardeja parametrejä, annetaan syötteenä erittäin suuri määrä tekstiä. Jotta kielimalli voi käsitellä virkkeitä ja sanoja, ne täytyy ensin tokenisoijan avulla lyhentää pienempiin kirjaimista koostuvin merkkijonoihin. Nämä tokenisoijan jakamat merkkijonot annetaan järjestyksessä takaisinkytkeytyvälle neuraaliverkolle (englanniksi *Recurrent Neural Network*), joka käsittelee merkkijonojen väliset yhteydet. Neuraaliverkon laskemat representaatiot sanoista annetaan dekooderille, joka tuottaa mahdollisia ratkaisuja esittäviä ulostulovektoreita. Vektorimuuntajan ”attention”-mekanismi samalla valitsevasti keskittyy syötteessä oleviin tiettyihin merkkijonoihin, jotka liittyvät eniten senhetkiseen ulostuloon. Tämän ansiosta dekooderi voi tuottaa ulostuloja, jotka liittyvät erityisesti syötteen kontekstiin ja tarkoitukseen. Lopuksi ulostulovektorit johdatetaan lineaarisen kerroksen ja ”softmax”-aktivointifunktion läpi todennäköisyysjakauman tuottamiseksi mahdollisille ratkaisuille.



Kuvio 1. Transformer-arkkitehtuurin vuokaavio (Jadhav, 2025)

2.1.1 ChatGPT

ChatGPT on OpenAI:n tekemiin GPT (Generative Pre-trained Transformer) -malleihin perustuva keskusteleva älykäs robotti, joka voi suurena kielimallina antaa yksityiskohtaisen vastauksen kehotteessa annettuun käskyyn (Wu, 2023, s. 1). ChatGPT on näyttänyt onnistuvansa esimerkiksi monikielisessä konekääntämisessä, koodin virhekorjauksessa ja tarinoiden kirjoittamisessa. Sillä on myös kyky muistaa, mitä käyttäjä on aikaisemmin sille kertonut.

ChatGPT-malleista tämän tutkimuksen aikana oli käytettävissä kolme erilaista mallityyppiä. Ensimmäinen näistä sisältää niiden suurimmat mallit, johonka tutkimuksen aikana kuului ChatGPT-4o. GPT-4o on autoregressiivinen monimuotoinen malli (OpenAI, 2024a, s. 1). Monimuotoisuudella tarkoitetaan mallin kykyä vastaanottaa syötteenä sekä tuottaa tekstin lisäksi myös ääniä, kuvia ja videoita. GPT-4o on OpenAI:n mukaan aikaisempaan malliin GPT-4 Turboon verrattuna muissa kuin englannin kielessä paljon paranneltu. Toiseen mallityyppiin sisältyy suuremmasta mallista paljon pienemmät versiot, kuten ChatGPT-4o-mini. GPT-4o-mini on GPT-4o:sta pohjautuva kevyempi ja halvempi versio, joka on silti yhtä monimuotoinen. (OpenAI, 2024b). Kolmanteen mallityyppiin kuuluu päättelevät mallit, kuten ChatGPT-o4-mini. Nämä mallit osaavat miettiä vastaustaan pidempään, jonka ansiosta niiden on tarkoitus olla älykkäämpiä (OpenAI, 2025). ChatGPT:n ”mini”-malleja pystyi tutkimuksen aikana käyttämään vapaasti ilman käyttörajoituksia.

2.1.2 Gemini

Gemini on Googlen tekemä monimuotoinen tekoälytyökalu, joka koostuu useista suuriin kielimalleihin ja luonnollisen kielen prosessointiin liittyvistä teknologioista (Imran & Almusharraf, 2024, s. 2). Gemini pystyy käsittelemään tekstin lisäksi useita erilaisia syötemuotoja, kuten kuvia, videoita, sekä PDF-tiedostoja. Gemini oli aikaisemmin tunnettu nimellä Bard (Baytak, 2024, s. 2). Googlen Gemini Team (2025, s. 1) kertovat tuoreimmista Gemini-malleista tämän työn aikana, jotka ovat nimetty Gemini 2.5 Pro ja Gemini 2.5 Flash -malleiksi. Kummatkin mallit ovat monimuotoisia pitkiä kontekseja käsitteleviä kielimalleja. Gemini 2.5 Pro on Gemini-malleista älykkäin päättelevä malli. Gemini 2.5 Flash puolestaan on kevyempi malli, joka tasapainottelee laatua, hintaa ja viivettä monimutkaisten tehtävien kanssa. Gemini 2.5 Flash on pienemmästä koostaan riippumatta tehokkaampi kuin viime vuoden Gemini 1.5 Pro -malli (Gemini Team, 2025, s. 12).

2.1.3 Copilot

Copilot on Microsoftin kielimalleja käyttävä tekoälyavustaja, josta on olemassa useita versioita kuluttajille sekä yrityksille (Microsoft Learn, 2025). Yrityksille tarjottava Copilot kutsutaan nimellä Microsoft 365 Copilot. Tämän yrityskeskeisen version oli tarkoitus Kytön (2024, s. 9) mukaan yhdistää kielimallin tehokkuus muiden Microsoft 365 -palvelun sovelluksien kanssa. Microsoft 365 Copilot pystyisi esimerkiksi tekemään Word-dokumentin PowerPoint-esityksestä tai kirjoittamaan Teams-keskustelusta muistiinpanot automaattisesti. Lisäksi Copilot-kielimalleista on kuluttajille ilmaisesti käytettävissä oleva versio nimellä Microsoft Copilot (Microsoft Learn, 2025). Tässä kuluttajaversiossa oli tutkimuksen aikana saatavilla ”Think Deeper”-asetus, jossa kielimalli käyttää pohjanaan OpenAI:n o1-mallia (The Copilot Team, 2025). Kyseinen malli on syväajatteleva, joten sen on tarkoitus pystyä käsittelemään monimutkaisempiakin tehtäviä.

2.2 Pienet kielimallit

Pieni kielimalli (englanniksi *Small Language Model*) on kielimalli, jonka tarkoitus on pitää suuren kielimallin tarkkuus ja/tai mukautuvuus, kun esimerkiksi laitteisto tai tiedon määrä ovat rajoitettuja (Nguyen ja muut, 2024, s. 1). Suuret kielimallit ovat hyvinkin tehokkaita, mutta ne vaativat tehokkuutensa saavuttamiseksi paljon resursseja ja energiaa. Tästä syystä ne toimivat vain suurissa tietokeskuksissa erikoistuneilla laitteistoilla, joihin yhdistetään verkon kautta. Pienet kielimallit käyttävät vähemmän resursseja, joten ne säästävät energiaa (Zhang ja muut, 2025, s. 31). Ne voi mahdollisesti ottaa käyttöön paikallisesti ilman verkkoyhteyttä erilaisilla kuluttajalaitteilla, kuten esimerkiksi tietokoneilla. Hyvin pieniä kielimalleja voi ottaa käyttöön jopa älypuhelimillakin. Tämä on mahdollista niiden pienen koon ansiosta, jolloin ne eivät vaadi erikoistuneita laitteistoja.

Koska pienten kielimallien sisäistämän tiedon määrä on suuriin kielimalleihin verrattuna vähäinen, niin ne eivät osaa yhtä hyvin vastata laajasti kaikkiin kysymyksiin. Ne voivat

kuitenkin erikoistua tiettyihin tehtäviin, jolloin ne voivat kyseisissä tehtävissä toimia laadukkaasti. Lisäksi suurten kielimallien käyttämisellä voi olla haitallisia ympäristövaikutuksia. Esimerkiksi kielimallien kouluttamiseen saatetaan käyttää uusia ja tehokkaita laitteistoja käytettyjen sijaan, joka vaikuttaa hiilijalanjälkeen (Ding ja Shi, 2024, s. 38). Datakeskukset, jotka tarvitaan suurten kielimallien käyttöönottoon, käyttävät usein myös koneiden viilentämiseen ja sähkön tuottamiseen paljon vettä (George ja muut, 2023, s. 92–93). Pienten kielimallien käyttäminen suurten sijasta voi olla parempi vaihtoehto vähintään ympäristön kannalta.

Pienet kielimallit voidaan Zhangin ja muiden (2025, s. 32) mukaan luokitella tyhjästä tai suuresta kielimallista koulutettuihin malleihin. Tyhjästä kouluttamisessa käytetään tehokkaampaa transformer-arkkitehtuuria. Tämä on mahdollista muun muassa eteenpäinsyöttöverkkojen (englanniksi *feed forward network*), tehokkaamman ”attention”-mekanismin tai ”attention”-mekanismista vapaan arkkitehtuurin avulla. Suuresta kielimallista koulutettu pieni kielimalli mahdollistetaan Zhangin ja muiden (2025, s. 33–34) kertoen mallin kompressoinnilla. Kielimalli voidaan kompressoida pieneksi kolmella eri tavalla, jotka ovat kvantisointi, karsiminen, ja tislaaminen. Kvantisoinnissa mallin painojen ja/tai aktivoimisten tarkkuutta vähennetään mallin pienentämiseksi. Karsimisessa kielimallista karsitaan pois turhia painoja tai neuroneja operaation yksinkertaistamiseksi. Tislaamisessa tieto tislataan suuremmasta ”opettaja”-mallista tiiviin ”opiskelija”-mallin kouluttamiseen. Kaikki tässä tutkimuksessa testatut pienet kielimallit kuuluvat suuresta kielimallista koulutettuihin malleihin.

Kielimallien koot perustuvat niissä olevien parametrien määrään. Nykyään suurimmalla osalla malleista parametrien määrä on miljardeissa, mutta suurimmat mallit yltävät jopa biljooniin asti (Katz, 2024). Nalpas (2024) kertoo esimerkkinä Gemma 2B -kielimallin, jossa ”2B” tarkoittaa noin kahden miljardin parametrin kokoa. Tämä parametrien määrä on tapa mitata kielimallien suorituskykyä, sillä suurempi parametrien määrä vastaa useimmiten tehokkaampaa mallia. Kuitenkin tämä parametrien määrä voi vaikuttaa negatiivisesti sen toimintaan liian heikoilla laitteistoilla. Kielimallin toimivuuteen vaikuttaa

laitteen keskusmuisti ja erillisen näytönohjaimen sisältävän laitteen kanssa myös näytönohjaimen muisti. Jos muistia ei ole kielimallille tarpeeksi paljon, niin se toimii erittäin hitaasti ja vastauksia voi joutua odottamaan hyvinkin pitkään.

Pienen kielimallin määritelmää ei vielä ole standardoitu (Subramanian ja muut, 2025, s. 2). Nguyen ja muut (2024, s. 1) toteavat ”suuren” tai ”pienen” kielimallin merkityksen muuttuvan ajan kuluessa. He antoivat esimerkkinä GPT-2-kielimallin, jota oli vuonna 2019 pidetty suurena kielimallina, vaikka sen koko oli vain 1,5 miljardia parametria. Nykyaikaisista pienistä kielimalleista osa ovat jo tätä suurempia. Yksi suurten ja pienten kielimallien eroavaisuuksista on, että suurten käyttö on riippuvainen pilvipohjaisista ohjelmointirajapinnoista, kun taas pienet voi ottaa käyttöön paikallisesti (Srivastava ja muut, 2025, s. 1). Tässä työssä pieni kielimalli on määritetty siten, että se on kielimalli, jonka voi ottaa käyttöön paikallisesti kuluttajalaitteilla ilman verkkoyhteyden tarvetta. Kaikki tässä työssä testatut pienet kielimallit ovat enintään 14 miljardin parametrin kokoisia. Kielimallien testaamiseen käytetyllä kuluttajatietokoneella olisi ollut isompien mallien kanssa vaikeuksia.

2.2.1 Llama

LLaMA-kielimallit ovat Meta AI:n kehittämiä avoimen lähdekoodin suuria kielimalleja (Touvron ja muut, 2023, s. 1). Nämä kielimallit ovat ilmaisia käyttää omalla koneella avoimen lähdekoodin ansiosta. Nämä mallit voivat olla avoimia, koska niiden kouluttamisessa käytettiin ainoastaan avoimien lähteiden materiaaleja ja tekstejä. Tutkimuksen aikana oli tarjolla Llama 3 -sarjan kielimallit, jotka Metan Llama Teamin (2024, s. 2) mukaan keskittyvän monikielisyyteen. Llama-kielimalleista tarjotaan useita erikokoisia malleja parametrien määrien mukaan. Tarjolla on Llama 3.1 -version kanssa esimerkiksi pienin 8 miljardin, suuri 70 miljardin, sekä jättimäinen 405 miljardin parametrien mallit. Llama Team on suositellut käyttämään pienempiä malleja vain englannin kielen kanssa, vaikka niiden kouluttamiseen on käytetty monikielistä tietoa. Näistä malleista kuitenkin vain

8B:n voidaan todeta pieneksi malliksi, joten tässä työssä testataan kyseistä mallia suosituksesta riippumatta.

2.2.2 Gemma

Gemma-kielimallit ovat Googlen tuottamia vapaan lähdekoodin malleja, jotka on suunniteltu yhdessä Gemini-kielimallien kanssa (Gemma Team, 2025, s. 1). Tutkimuksen aikana uusin versio on Gemma 3, josta on olemassa 1, 4, 12 ja 27 miljardin parametrin mallit. Näistä suurin mahdollinen testikoneella käytettävä kielimalli on 12 miljardin parametrin malli, jota tässä työssä testataan. Gemma Team (2025, s. 1) kertoo kaikkien Gemma 3 -mallien olevan koulutettu tiedon tislaamisen avulla, ja ne käyttävät Gemini 2.0 -kielimallien tokenisoijaa. Lisäksi niiden tietokokoelma oli käyty uudelleen läpi keskittyen mallien monikielisyyteen. Uuden koulutuksen jälkeisen lähestymistavan ansiosta kaikkien Gemma 3 -mallien kyvyt, kuten monikielisyyden, kehittyivät entistä enemmän. Näiden ansiosta Gemma 3 on suorituskykytestien mukaan monikielisyydessä aikaisempaa Gemma 2 -mallia entistäkin parempi (Gemma Team, 2025, s. 21). Uudemman Gemma 3:n 12 miljardin parametrin malli on monikielisyydessä samalla tasolla kuin Gemma 2:n suurin 27 miljardin malli.

2.2.3 DeepSeek

DeepSeek-mallit ovat vapaan lähdekoodin suuria ja pieniä kielimalleja (DeepSeek-AI, 2025, s. 1). Niistä tutkimuksen aikana uusimmat mallit ovat DeepSeek-R1 ja DeepSeek-R1-Zero. DeepSeek-R1-Zero-mallin kouluttamisessa käytettiin vain vahvistavaa oppimista, eikä sisällytetty lainkaan ohjattua hienosäätämistä (DeepSeek-AI, 2025, s. 5). Kielimallille ei annettu ohjattua koulutustietoa, vaan sen täytyi keskittyä itsearvioimiseen pelkän vahvistavan oppimisen avulla. Vaikka se pystyi hyvin perustelemaan vastauksensa itsenäisesti, sillä oli ongelmia huonon luettavuuden ja kielten sekoittamisen kanssa (DeepSeek-

AI, 2025, s. 9). DeepSeek-R1 oli suunniteltu käyttämään DeepSeek-R1-Zero-mallin kouluttamiskeinojen lisäksi kylmäkäynnistystietoja. Näiden kylmäkäynnistystietojen tarkoitus on tehdä kielimallin tuottamasta tekstistä luettavampaa pitäen samalla aikaisemman mallin itsenäisesti selvittämät perustelut. DeepSeek-r1 -malleista suurin on kooltansa 605 miljardia parametria, mutta tiedon tislaamisen kautta on olemassa useampia paljon pienempiä vaihtoehtoja. Suurin DeepSeek-kielimalli, joka toimii tutkimuksessa käytetyllä testikoneella, on 14 miljardin parametrin malli.

2.3 Kielimallien eettisyys ja hallusinaatio

Suurien kielimallien käyttämisessä voi tulla esille monenlaisia ongelmia, joista osa ovat disinformaatio ja eettisyys**.** Disinformaation suhteen kielimallit voivat helposti tuottaa valheellista tietoa joko koulutustietojen, mallin kouluttamisen, tai niitä hyödyntävien ihmisten takia (Barman ja muut, 2024, s. 8). Esimerkiksi jos koulutustiedot sisältävät matalalaatuisia ja ennakkoasenteisia tietoja, ne voivat vääristää kielimallin tuloksia mahdollisesti vahingoittaviksi. Tämä on mahdollista hyvin suurissa kielimalleissa, kuten ChatGPT:ssä, sillä niiden kouluttamisessa on käytetty niin paljon tietoa, ettei sitä kaikkea ole voitu vahvistaa laadukkaaksi. Kielimallien koulutustiedot saattavat myös sisältää henkilökohtaisia ja yksityisiä tietoja, mikä on myös osa suurten kielimallien eettisiä ongelmia (Jiao ja muut, 2024, s. 9). Mahdollisia keinoja kielimallien eettisyyden parantamiseen voisivat esimerkiksi olla koulutustietojen laadun parantaminen, yksityisten tietojen unohtaminen, ja hallusinaatioiden proaktiivinen tunnistus (Jiao ja muut, 2024, s. 17–18).

Kielimallien kanssa hallusinaatio tarkoittaa niiden itsevarmasti kertomia todenkuuloisia valheita (Kalai ja muut, 2025, s. 1). Kielimalli saattaa luulla virheellisen vastauksensa olevan totta, sillä se kuulostaa sen mielestä todennäköiseltä. Yksi syy hallusinaatioiden olemassaoloon on se, että kielimallien kouluttamisessa niitä rangaistaan epävarmoista vastauksista (Kalai ja muut, 2025, s. 4). Kielimallien halutaan antavan vastauksia käyttäjien kysymyksiin. Kielimallien itsevarmuus johtuu siis siitä, että ne haluavat koulutuksensa

takia välttää epävarmojen vastausten antamista. Kielimallien arviointimetodeissa hyväksyttävien epävarmojen vastauksien palkitseminen voisi auttaa hallusinaatioiden vähentämisessä (Kalai ja muut, 2025, s. 16). Kielimalleja voi avustaa antamaan laadukkaampia vastauksia ja siten vähentää niiden hallusinointia esimerkiksi kehotesuunnittelun avulla.

Kehotesuunnittelu (englanniksi *prompt engineering*) on menetelmä, jolla pyritään ohjaamaan suurien kielimallien vastauksia haluttuun suuntaan suunnitellun kehotesyötteen avulla (Chen ja muut, 2024, s. 2). Liian yksinkertainen kehote voi antaa kielimallille liikaa vapautta päätöksissään, jolloin mallin vastaus voi olla hyvin laaja (Chen ja muut, 2024, s. 4). Antamalla selkeät ohjeet hyvin suunnitellun kehotteen avulla voidaan vähentää epäselkeyttä ja tarkentaa mallin vastausta (Giray, 2024, s. 3). Muita kehotesuunnittelukeinoja ovat esimerkiksi roolin ja yhden tai useamman esimerkkivastauksen antaminen kielimallille (Chen ja muut, 2024, s. 5–6). Kehotteeseen voi myös lisätä kontekstia aiheen selventämistä varten (Giray, 2024, s. 3). Kielenkäytön palautteen kanssa Huang (2023, s. 75) on suositellut kertomaan kielimallille tarkasti, miten sen pitäisi antaa tekstistä palautetta. Kehotteessa kannattaa esimerkiksi kertoa, että haluaa pelkästään palautetta ilman uudelleenkirjoittamista. Muuten kielimalli mahdollisesti antaa heti vastauksen, jonka käyttäjä voi heti kopioida ilman, että hänen tarvitsisi itse korjata tekstiä. Lisäksi on hyvä kertoa, mitä tasan tarkalleen kielimallin pitäisi kielenkäytön suhteen arvioida, kuten kirjoitusvirheitä tai selkeyttä.

3 Kielimallien käyttö kielten oppimisessa

Suurien kielimallien käyttö kielten oppimisessa keskittyy ainakin neljään eri osa-alueeseen, jotka ovat:

- Generoivat: Tuotetaan ihan mitä vaan, kuten tekstiä, kuvia, aikatauluja jne.
- Arvioivat: Arvioivat annetun kehotteen tekstin laatua eri näkökulmista
- Keskustelevat: Keskustelukaveri, jonka kanssa voi keskustella opittavalla kielellä
- Kääntävät: Ottavat yhdellä kielellä kirjoitetun tekstin ja kääntävät sen toiselle.

Tässä kappaleessa käydään läpi erilaisia aikaisempia tutkimuksia, joiden kielimallien käyttötavoissa ja metodeissa ne keskittyvät vähintään yhteen näistä neljästä osa-alueesta. Lisäksi käydään läpi suomen kielen oppimiseen liittyviä tutkimuksia. Eri osa-alueita voidaan käyttää samanaikaisesti eri töissä. Tällöin kategoria on tutkimukselle valittu sen mukaisesti, mikä niistä oli tutkimuksen suhteen tärkein.

3.1 Generoiva käyttö

Generoivassa käytössä kielimallit tuottavat käyttäjälle jotakin, kuten vaikka jonkinlaisen suunnitelman tai tehtävän. Tähän kategoriaan sisältyy kaikki kielen oppimiseen liittyvät tutkimukset, jotka tekevät jotakin muuta kuin arviointia, keskustelua, tai kääntämistä. Tähän kuuluu esimerkiksi pelkkä tekstin tai koodin generointi, jossa käyttäjä voi pyytää kielimallia tekemään esseekappaleita tai auttaa ratkaisemaan ohjelmointiongelmia (Laato ja muut, 2023, s. 228–229). Suuret kielimallit voivat myös Yun ja muiden (2024, s. 4) mukaan tuottaa esimerkiksi opetusmateriaaleja. Kielimallit voivat myös muuttaa annetun tiedon johonkin toiseen haluttuun muotoon. Ne voivat esimerkiksi tehdä kirjasta opettajan ja opiskelijoiden välisiä dialogeja tai opetusdioista jäsenneltyjä opetusohjelmia. Kielimallit eivät kuitenkaan onnistuneet Yun ja muiden tutkimuksessa tekemään tarpeeksi monipuolisia ja syvällisiä tehtäviä. Näihin tehtäviin kuuluvat esimerkiksi kulttuurisesti monipuoliset lukumateriaalit tai dynaamiset ymmärtämisharjoitukset.

Esimerkiksi yhdessä tutkimuksessa tutkitaan ChatGPT:n käyttöä englannin kielen opetusmateriaalin tuottamiseen (Koraishi, 2023). tässä tutkimuksessa pyydettiin ChatGPT:tä tuottamaan useita eri asioita. Yhtenä esimerkkinä tutkimuksessa mallia pyydettiin tekemään tuntisuunnitelma tietyn englannin kielen B2-tasolla (Koraishi, 2023, s. 62). ChatGPT antoi kehotteen mukaisesti tuntisuunnitelman liittyen naapurustoihin ja yhteisöihin. Tuntisuunnitelma sisälsi esimerkiksi erilaisia paritöitä, joissa opiskelijat keskustelisivat keskenään sanojen tarkoituksista ja myöhemmin tekisivät lyhyen dialogin liittyen yhteisöongelmiin. Suunnitelma sisälsi myös muutamia mahdollisia kotitehtäväesimerkkejä opiskelijoille. Hu ja muut (2024, s. 1446) olivat myös tutkineet matematiikan opetussuunnitelman tuottamista GPT-4-mallilla. Suurilla kielimalleilla olisi Hun ja muiden (2024, s. 1456) kertoen suuri potentiaali opetussuunnittelun kanssa. Kielimallit osoittavat heidän mukaansa useissa arviointiulottuvuuksissa ja erityisissä tietomoduuleissa korkeaa suorituskykyä.

Bacherikov (2024, s. 6) oli omassa kandidaatin tutkimuksessaan käyttänyt ChatGPT:tä tuottamaan tehtäviä suomen kielen oppimista varten. Kielimalli antaisi myös arvion käyttäjän vastauksesta tehtävään. Bacherikov (2024, s. 13) myös lisäsi kielimallia käyttävälle sovellukselleen kyvyn selvittää minkä tahansa taivutetun sanan perusmuoto. Tämän kyvyn tarkoitus on auttaa mahdollisten vaikeiden taivutusten ymmärtämistä. Perusmuodon hakemiseen hän käytti kielimallin sijasta suomen kielen prosessointiin keskittyvää Voikko-kirjastoa. Hänen sovelluksensa voi myös antaa mille tahansa suomen kielen sanalle Sanakirja.fi:stä selityksen. Pyytäessään ChatGPT:tä tuottamaan tehtäviä ilman hienosäätöä, Bacherikovilla (2024, s. 35–36) tuli mallin käytössä ongelmia esille. Jos kielimallia pyysi esimerkiksi antamaan satunnaisen aiheen, se päätti usein antaa samoja aiheita. ChatGPT:n arvioinnit vastauksista eivät myöskään olleet tarpeeksi laadukkaita. Hienosäätämällä GPT-mallia saatiin usein laadukkaampia arviointeja verrattuna normaaliin (Bacherikov, 2024, s. 47). Aiheet kuitenkin satunnaistettiin kielimallin ulkopuolella, sillä ChatGPT oli tutkimuksen mielestä huono satunnaistamaan (Bacherikov, 2024, s. 48).

Nikolova-Stoupak ja muut (2024, s. 153–154) olivat tutkimuksessaan testanneet kielimallien avulla sanojen ympärille kontekstin luomista monivalinta- tai täydennystehtäviä varten. Kielimallin on tarkoitus tuottaa virkkeitä erilaisten sanojen ympärille, joista sitten kyseiset sanat otetaan pois. Käyttäjän pitäisi pystyä selvittämään piilotettu sana kielimallien tuottamien virkkeiden ja vinkkien avulla. Vinkkien avulla täydennystehtävissä annetaan pari ensimmäistä kirjainta sanasta, ja monivalinnoissa sana annetaan vaihtoehtona muiden väärien vastauksien kanssa. Testattuina kielimalleina olivat pieni ”Mistral-7B-Instruct”-malli ja suuri ”Gemini 1.0 Pro”-malli (Nikolova-Stoupak ja muut, 2024, s. 155). Kielimalleille oli myös suunniteltu kehotteet ilman esimerkkejä sekä viiden esimerkin kanssa (Nikolova-Stoupak ja muut, 2024, s. 158). Esimerkit olivat tehneet ammattilais-opettajat. Automaattisen arvioinnin kanssa, jossa kielimallien tuottoja verrataan tiettyyn viitteeseen, kummatkin mallit tuottivat esimerkkien kanssa parempia tuloksia (Nikolova-Stoupak ja muut, 2024, s. 159). Mistral-malli esimerkkien kanssa tuotti tässä parhaimmat tulokset. Opettajien mielestä kielimallien tuottojen arvioinnissa kielimallit tuottivat parhaimmat tulokset ilman esimerkkejä (Nikolova-Stoupak ja muut, 2024, s. 160). Geminin malli ilman esimerkkejä kehotteessa tuotti heidän mielestänsä parhaimmat tulokset. Esimerkkienkin kanssa kielimallit tuottivat silti lähestulkoon yhtä hyvät tulokset kuin ilman esimerkkejä (Nikolova-Stoupak ja muut, 2024, s. 162).

Luukkonen ja muut (2023, s. 1–2) olivat kehittäneet oman GPT-pohjaisen suomen kieleen keskittyneen kielimallin. GPT-mallin lisäksi he kouluttivat 176 miljardin parametrin kokoista BLOOM-mallia suomenkielisten tietojen kanssa. Heidän työnsä tarkoituksena oli selvittää, kuinka hyvän kielimallin saisi tehtyä suomen kielelle, kun esimerkiksi Wikipediassa vain alle 1 % tekstistä on suomeksi. Kielimallin kouluttamiseen käytettyihin tekstilähteisiin kuului esimerkiksi Ylen suomenkielisiä uutisia vuosien 2011 ja 2020 väliltä, suomenkielinen Wikipedia, sekä useita eri Common Crawl -lähteitä (Luukkonen ja muut, 2023, s. 2–3). Kielimallien testaamisessa luukkonen ja muut (2023, s. 6) käyttivät heidän tekemää FIN-bench-tietojoukkoa. FIN-bench perustuu BIG-bench-tietojoukkoon, jonka tarkoitus on mitata kielimallien useiden eri toimintojen toimivuutta. FIN-bench-tietojoukolla testattiin esimerkiksi analogioiden, aritmetiikan, väärinkäsitysten sekä tunteiden

ymmärtämistä. Testauksissa kielimalleille annettujen esimerkkien määriä vaihdeltiin nollan ja kolmen välillä. Testattuihin kielimalleihin kuului Luukkosen ja muiden (2023, s. 7) tekemän TurkuNLP:n lisäksi FinnishNLP-malli ja Hatanpään tekemä malli. Kaikkien arviointiosuuksien kokonaisuudessa TurkuNLP oli tehokkain. Kielimalleilla oli Luukkosen ja muiden (2023, s. 16) tuottamien tuloksien mukaan vaikeuksia kertolaskujen ja monilukuisten jakolaskujen kanssa. Syyn ja seurauksen ymmärtämisessä kielimallit onnistuivat noin puolet ajasta. Tunteiden, yleistiedon, ja aikomusten tunnistamisessa onnistui pääasiassa ainoastaan yli miljardin parametrin TurkuNLP-mallit. Väärinkäsitysten, epäselkeiden virkkeiden ja parafraasien tunnistamisessa parametrien määrä ei suuresti vaikuttanut tulokseen.

3.2 Keskusteleva käyttö

Keskustelevassa käytössä kielimallit keskittyvät vain keskustelemiseen käyttäjän kanssa esimerkiksi keskustelukaverina. Ye ja muut (2024, s. 7–8) antavat esimerkkeinä kielimallien kyvyn simuloida luokkahuonetta sekä antaa dynaamista ja henkilökohtaista ohjausta. Simuloidun luokkahuoneen avulla opettajat tai tutkijat voivat esimerkiksi testata pedagogisia strategioita ja hiota opettamistekniikoita ilman fyysisen luokkahuoneen järjestämistä. Henkilökohtaisessa ohjauksessa puolestaan kielimallit voivat antaa laajasti yksilöllistä opastusta. Ye ja muut (2024, s. 8) kuitenkin huomauttavat, että luokkahuoneen simuloimisen rakenteiden vertailuarvoja ei ole yhdenmukaistettu. Lisäksi myös kielikohtaisten ohjausstrategioiden, esimerkiksi reaaliaikaisten keskusteluharjoitusten, tehokkuutta ei vielä ole täysin tutkittu.

Park ja muut (2024, s. 1) olivat tutkineet kielimallien tehokkuutta opettavaisen englanninkielisen keskustelun tuottamisessa. Ilman keskustelemisen harjoittelemista aidoissa tilanteissa opiskelijat saattavat olla epävalmiita jokapäiväisiin keskusteluihin, vaikka osaisivatkin yleiskielen hyvin (Park ja muut, 2024, s. 2). Tutkimuksessa käytettiin suunnittelu- ja kehitystutkimusmenetelmää, jossa systemaattisesti tutkitaan suunnittelun, käyttöön-

oton ja arvioinnin prosesseja (Park ja muut, 2024, s. 8). Tämän metodologian avulla tutkimuksessa saatiin tuotettua empiirisiin todisteisiin perustuvia ohjeistavia sekä ei-ohjeistavia todisteita. Park ja muut (2024, s. 9) olivat ensin analysoineet aiheeseen liittyvät tarpeet, joiden avulla he perustivat kielimallien kehittämisen suunnittelun periaatteet. He olivat tutkineet huippuluokan kielimalleja ja suunnitelleet tietojoukon kielimallien hienosäätämistä varten. He myös kehittivät arviointikriteerit ja kehotepohjat testaamista varten. Prototyyppimallin käytön palautteen keräämiseksi he tekivät haastatteluita kolmen eri opettajan kanssa. Parkin ja muiden (2024, s. 23) mukaan hyvin hienosäädetty ja kehotesuunniteltu ChatGPT:tä pienempi malli pystyy myös tuottamaan opettavaista keskustelua ilman virheitä. Esimerkiksi kielimalli ”NeuralHermes-2.5-Mistral-7B-AWQ” tuotti hyvän tuloksen melko nopeasti ilman virheitä.

Alsafari ja muut (2024, s. 2) Tekivät omassa tutkimuksessaan apuopettajana toimivan chatbotin, jossa käytettiin ”Retrieval-Augmented Generation” (RAG) -tekniikkaa. Tämän tekniikan on tarkoitus rikastuttaa kielimallien tietoa verkkoaluekohtaisilla lisätiedoilla. Tutkimuksen tarkoituksena oli selvittää, kuinka tätä RAG-tekniikkaa käyttävät systeemit eroavat tyypillisistä chatboteista. He halusivat samalla selvittää varsinkin näiden RAG-tekniikkaa käyttävien kielimallipohjaisten chatbottien kyvyn opettaa toiseen asteen opiskelijoita. He arvioivat tämän chatbotin automaattisesti toisen mallin avulla sekä manuaalisesti ihmisten antamien arvioiden mukaan (Alsafari ja muut, 2024, s. 10–11). Chatbotin keskustelukyky arvosteltiin sen oikeassa olemisen, relevanssin, tarkkuuden, selkeyden, ajatuksen syvyyden ja yleisen auttavuuden mukaan. Lopullisena arviona yhdestä viiteen chatbotille annettiin 4.76, eli se toimisi aika hyvin apuopettajana, vaikka parannettavaakin olisi.

Tyen ja muut (2024, s. 3) käyttivät omassa tutkimuksessaan chatbottina BlenderBot-nimistä kielimallia. He selittivät valintansa siten, että tämä kielimalli ei ole etukäteen hienosäädetty olemaan ohjeistava. Se on myös hienosäädetty keskusteluun keskittyvällä tietojoukolla, jolloin se voi olla kunnon keskustelija virtuaalisen apulaisen sijasta. Chatbotin käyttäminen nosti oppijoiden itseluottamusta, vaikka yksi sessio ei ehkä ehtinyt

opettaa paljoa (Tyen ja muut, 2024, s. 4). Kielimallisovellukseen oli myös sisälletty toiminto, jossa sanaa klikkaamalla näytettiin sen sanakirjaselitys (Tyen ja muut, 2024, s. 7). Yli puolet tutkimukseen osallistuneista eivät kuitenkaan käyttäneet tätä toimintoa. Ne, jotka käyttivät sitä, klikkasivat sanoja vain alle kymmenesosassa chatbotin viesteistä. Mahdollisina syinä ajateltiin, että vaikeita sanoja saattoi olla hyvin harvoin, osallistujat halusivat mieluiten keskittyä keskusteluun, ja sanakirjaselitykset saattoivat olla heidän mielestänsä hyödyttömiä. Lisäksi voi olla, että he unohtivat tämän toiminnon olemassaolon.

Suurten kielimallien kanssa pystyisi myös tehdä keskusteluja esimerkiksi virtuaalisessa todellisuudessa. Tällöin aitoja keskustelutilanteita voidaan helpommin opettaa opiskelijoille. Pan ja muut (2024, s. 5) olivat yhdistäneet keskusteluun keskittyvän kielimallin sosiaalisen virtuaalitodellisuustilan kanssa. He käyttivät kielimallinsa pohjana GPT-4-mallia, ja virtuaalitodellisuustiloina käytettiin VRChat-sovelluksen tarjoamia tiloja. Kielimalli myös käytti OpenAI:n ”Tekstistä Puheeksi” -palvelua, jolloin sen kanssa pystyi keskustelemaan puheen avulla. Panin ja muiden (2024, s. 3) ideana oli tukea englannin kieltä opiskelevia oikean elämän tilanteisiin perustuvien roolileikkikeskustelujen avulla. Opiskelijat oppisivat tällöin interaktiivisesti kielimallin ja virtuaalitodellisuuden kanssa englannin kieltä. Panin ja muiden (2024, s. 8) mukaan tutkimukseen osallistuneet totesivat kielimallin kanssa keskustelun tuntuvan siltä niin kuin puhuisi todelliselle opettajalle. Keskustelut kulkivat sulavasti, ja osallistujat myös kehuivat kielimallin kykyä vastata heidän henkilökohtaisiin kokemuksiinsa liittyvillä tavoilla. Esimerkiksi yksi osallistuja ilmoitti kielimallille olevansa Indiasta, jolloin se antoi tervehdyksensä hindun kielellä. Kielimalli oli tulosten mukaan englanniksi onnistunut kopioimaan ihmisopettajan käyttäytymistapoja.

3.3 Arvioiva käyttö

Arvioivassa käytössä kielimallit arvioivat niille annetun tekstin laatua mahdollisesti eri näkökulmista. Kielimallit voivat esimerkiksi automaattisesti arvioida opiskelijoiden esseiden tulokset ja jäljittää opiskelijoiden tietyn aiheen ymmärtämisen kehitystä (Yu ja muut, 2024, s. 5). Ne voivat myös korjata ja selittää kielioppivirheitä sekä antaa palautetta (Yu ja muut, 2024, s. 6). Näitä voidaan myös yhdistää toistensa kanssa. Kielimallit voivat esimerkiksi antaa esseelle arvosanan ja samalla selittää esseessä parannettavia kohtia. Yun ja muiden (2024, s. 6) mukaan kielimallien arviointi ei kuitenkaan ole täydellinen. Kielimallien tuottamat arviot voivat erota asiantuntevien ohjaajien arvioista. Kielimallien empatian puute saattaa mahdollisesti antaa opiskelijoita masentavia arvioita. Lisäksi ei ole vielä onnistuttu määrittämään sellaista automaattista palautetta, joka antaisi parhaimmat oppimistulokset. Lisäksi arvioinnin ja palautteen yhdistämisessä niiden linjaus on heikko. Vaikka arviointi olisikin kattava, niin palautteen tarkkuus saattaa silti olla puutteellinen.

Arvioivaan käyttöön kuuluu myös kielimallien itsearviointi, jossa malli joko arvioi itseään tai jokin muu malli arvioi toista mallia. An ja muut (2023, s. 2) olivat esimerkiksi kokeilleet lähettää kielimallien vialliset vastaukset toiselle kielimallille, joka korjasi aikaisemman vastauksen virheet. Tätä prosessia kutsutaan ajatteluketjuksi (englanniksi *Chain of Thought*), jossa kielimallia ohjeistetaan kehotteessa miettimään annettua vastausta syvemmin. He keräsivät ensin useilta eri kielimalleilta erilaisia virheellisiä vastauksia. Nämä vastaukset annettiin sitten GPT-4-kielimallille, joka sitten yritti korjata virheelliset vastaukset oikeiksi. Korjauspohjaisen arvioinnin ansiosta kielimallien vastauksien laadut olivat myös parantuneet (An ja muut, 2023, s. 7–8). Nämä korjaukset voisi sisällyttää kielimalleihin hienosäätämisen avulla. Tällöin kielimallit joutuisivat ajattelemaan antamaansa vastausta tarkemmin, joka puolestaan parantaa vastausten laatua.

Meyerin ja muiden (2024, s. 3–4) tutkimuksessa kielimalli antoi opiskelijoiden kirjoittamasta englanninkielisestä tekstistä palautetta. Tutkimuksessa keskityttiin selvittämään, kuinka hyvin kielimallien antama palaute pystyisi auttamaan opiskelijoita. Tähän kuului

opiskelijoiden korjaamiskykyjen ja uusien samanlaisten tehtävien suorittamisen kehittyminen sekä motivaation kasvattaminen. Yli 400 16-vuotiasta opiskelijaa olivat Meyerin ja muiden (2024, s. 4) tutkimuksessa 20 minuutin aikana tehneet englanninkielisen argumentoivan kirjoitustehtävän. Testiryhmässä puolet opiskelijoista korjasivat omat tekstinsä kielimallin palautteen mukaisesti. Loput kontrolliryhmässä korjasivat ilman mitään palautetta. Opiskelijoiden tekemien tekstien laatu arvioitiin Meyerin ja muiden (2024, s. 5) itse kouluttamalla koneoppialgoritmilla. Koulutustiedot sisälsivät 2 420:n opiskelijan tekstejä sekä niille asiantuntijoiden antamia arvosanoja. Opiskelijat, jotka saivat apua kielimallilta, korjasivat tekstinsä paremmin ja olivat motivoituneempia verrattuna opiskelijoihin, jotka eivät saaneet apua (Meyer ja muut, 2024, s. 6).

Kielimallit pystyvät myös tehokkaasti arvioimaan, onko kirjoittajan käyttämä kieli hänen äidinkielensä vai vieras kieli. Cong (2024, s. 1) tutki erityisesti hienosäädettyjä kielimalleja, jotka keskittyivät arvioimaan sanan yllättävyyden. Yllättävyydellä hän tarkoitti tietyn lauseen järkevyyttä esimerkiksi kieliopin suhteen. Cong antoi esimerkkinä englanninkieliset lauseet “The keys to the cabinet *are* on the table” ja “The keys to the cabinet *is* on the table.” Jälkimmäisessä lauseessa on monikkovirhe, joka nostaa sen yllättävyyttä. Tällaiset virheet ovat merkkejä siitä, että englannin kieli ei olisi heidän äidinkielensä. Cong (2024, s. 9) totesi, että kun vieraan kielen oppijan taito sen kielen kanssa kasvaa, niin samalla myös oppijan kirjoittamien lauseiden yllättävyysarvo laskee. Kielimallien avulla pystyisi yksinkertaisesti arvioimaan kirjoittajan käyttämän kielen luonnollisuuden laatua.

Chang ja Ginter (2024, s. 1–2) käyttivät ChatGPT:tä suomen kielen opiskelijoiden lyhyiden vastauksien arvioimisessa. He käyttivät ChatGPT:n versioita GPT-3.5 ja GPT-4, ja kehotteesta oli tehty versiot esimerkin kanssa sekä ilman esimerkkiä (Chang & Ginter, 2024, s. 3). Kehotteessa annetaan kielimallille rooli, jonka mukaan sen pitää jäljitellä opettajien arviointia. Mahdollisessa esimerkissä annetaan esimerkki epäonnistuneesta sekä onnistuneesta vastauksesta. Vastauksien pituudet myös vaihtelivat kurssien mukaisesti noin 50 ja 600 kirjaimen välillä. GPT-4-malli esimerkin kanssa sai Changin ja Ginterin (2024, s. 5) parhaimmat tulokset verrattuna esimerkittömään GPT-4- sekä GPT-3.5-malliin. GPT-4

saa esimerkin kanssa lähes aina yli 60 % vastauksista vähintään yhden arvosanapisteen vaille todellisesta arvosanasta. Arvosanat on annettu eri kurssien aiheisiin liittyen eikä pelkästään kirjoituslaadun mukaan. Chang ja Ginter (2024, s. 8) antavat mahdollisena keinona parantaa tuloksia esimerkiksi kehotteen paremmin suunnittelemisen. Kehotteeseen voisi esimerkiksi lisätä useamman esimerkin tai parantamalla kysymysten selkeyttä. Kielimallille voisi myös antaa arvosanakriteereitä ja esimerkkivastauksia.

Creutz (2024, s. 1) oli tutkinut, kuinka hyvin Claude-, GPT-3.5- ja GPT-4-kielimallit onnistuvat aloittelijatason haastavien suomen kielen oppijoiden tekstien korjaamisessa. Hän halusi selvittää, kuinka hyvin laadukkaimmat kielimallit pystyisivät uudelleenmuotoilemaan aloittelijatason tekstejä oikeisiin muotoihin. Kyseessä ei ollut tiettyjen virheiden tunnistaminen, vaan sujuvien ja luonnollisten korjausten tekeminen. Kielimallien täytyisi siis osata muotoilla aloittelijan tekstit mahdollisimman laadukkaiksi. Testaamisessa Creutz (2024, s. 2) sääteli kielimallien satunnaisuuden määrää. Jokainen kokoonpano pyöritettiin kahdesti satunnaisuuden takia, eivätkä edelliset suoritukset vaikuttaneet tuloksiin. Kokoonpanoja oli yhteensä 36, ja tekstejä oli 25 kappaletta. GPT-4 sai parhaimmat tulokset matalalla satunnaisuudella, jossa ensimmäisellä suorituksella korjattiin oikein 175 virkettä 210 virkkeestä (Creutz, 2024, s. 3). GPT-malleilla oli myös taipumusta ylikorjata sujuvuuden takia, mutta niiden tuotot olivat laadukkaita ja sisälsivät hyvin vähän kirjoitusvirheitä (Creutz, 2024, s. 5).

3.4 Kääntävä käyttö

Kääntävässä käytössä kielimallit kääntävät yhdellä kielellä kirjoitetun tekstin toiselle kielelle, esimerkiksi englannin kielestä suomen kielelle. Esimerkiksi Heikkilän tutkimuksessa (2024) hän tutki ChatGPT:n käännösten laatua suomen ja englannin kielen kanssa. Hän keräsi 10 suomenkielistä artikkelia ja 10 englanninkielistä artikkelia, joista puolet olivat tiedeaihealueen artikkeleita ja loput puolet olivat paikallisuutisartikkeleita (Heikkilä, 2024, s. 17). Artikkelit syötettiin ChatGPT:lle kerran samalla pyytäen kehotteessa käännöstä (Heikkilä, 2024, s. 16). Tulokset kuitenkin näyttävät, että käännöksien laatu ei ole

tarvittavan hyvä. ChatGPT kääntää sanat suoraan, jonka takia jotkin lauseet eivät kuulosta luonnollisilta. Kun kääntää englannista suomeksi, niin virheitä tapahtuu vieläkin enemmän (Heikkilä, 2024, s. 24).

Shahriar ja muut (2024, s.10) olivat tutkineet GPT-4o:n useiden eri kykyjen tehokkuutta. Yksi näistä tutkituista kyvyistä oli tekstin kääntäminen englannista kuudelle eri kielelle. Kieliin kuuluivat espanja, arabia, hindu, ranska, portugali ja venäjä. Näihin kieliin keskittymisen oli tarkoitus antaa monipuoliset tulokset kielimallin kääntämiskyvyn arvioinnille. Jokaiselle kielelle oli kerätty tietojoukot, joista jokaisesta valittiin satunnaisesti 500 eri tietopistettä. Kielimallin kääntämiä sekä ihmisten kääntämiä virkkeitä verrattiin käyttämällä toista luonnollisen kielen prosessoijamallia (BERT) (Shahriar ja muut, 2024, s. 11). Malli muutti virkkeet vektorimuotoon, jolloin virkkeitä pystyttiin vertaamaan matemaattisesti. Shahriar ja muut (2024, s. 11–12) löysivät GPT-4o -kielimallin kääntämiskyvyn lähestyvän ihmisten tekevien käännösten laatua. Espanjan kielen kanssa tarkkuus oli 88 %, joka oli testatuista kielistä parhain tarkkuus. Ranskan kielellä oli 75 %:n tarkkuuden kanssa kielistä huonoin tarkkuus.

Yuan ja muut (2024, s. 3) testasivat tutkimuksessaan keinoja parantaa LLaMA-kielimallin monikielisyyttä kouluttamalla mallin pelkästään englannin ja jonkin muun kielen kanssa. He tekivät 101 eri versiota kielimallista, joissa jokaisessa malli hienosäädettiin kouluttamalla englannin kielen kanssa toisella eri kielellä. Yhdessä näistä malleista toisena kielenä oli suomen kieli. Näiden mallien toimivuutta testattiin kääntämällä Flores-101-tietojoukko kaikille 101:lle eri kielelle (Yuan ja muut, 2024, s. 4). Tuloksien avulla kielet jaettiin neljään eri kvadranttiin riippuen niiden kyvyistä auttaa joko eri kielien, koulutettujen kielien (eli vain englanti ja valittu kieli), kummankin tai ei kummankaan laadun parantamisessa (Yuan ja muut, 2024, s. 5). Suomen kieli oli esimerkiksi yksi kielistä, joka aina paransi vain muiden kielien laatua, mutta ei parantanut englannin tai suomen kielen laatua normaaliin verrattuna (Yuan ja muut, 2024, s. 6).

Jos kielimallit ymmärtävät suomen kieltä hyvin, niin ne pystyisivät sen avulla helpommin kääntämään esimerkiksi monimutkaisia taivutuksia. Moisio ja muut (2024, s. 1) olivat omassa tutkimuksessaan tutkinut, kuinka hyvin useat eri kielimallit pystyivät erottelemaan sanojen taivutusmuodon. Testattuihin malleihin kuuluivat esimerkiksi GPT-4-turbo, GPT-3.5-turbo, ja Llama2-70B. Omissa kehotteissaan Moisio ja muut (2024, s. 3) antoivat kielimalleille kehotteessaan joko yksi, viisi, kymmenen, tai ei yhtään esimerkkitaivutusmuotoa. Kielimalleille annettiin esimerkeissä taivutettu sana, sekä sen perusmuoto, luku (yksikkö vai monikko), sijamuoto, sekä omitusliite. Esimerkkien jälkeen annetaan kielimalleille vain taivutusmuoto ja perusmuoto sanasta. Kielimallien täytyi sanasta ratkaista taivutusmuodon luku, sijamuoto ja omistusliite. Moision ja muiden (2024, s. 3–4) testaamista malleista vain GPT-4-turbo oli onnistunut tässä, eikä mikään muu malli onnistunut tehtävässä lainkaan hyvin. Kielimallit onnistuivat ratkaisemaan taivutusmuodon luvun ja omistusliitteen melko tarkasti. Sijamuodon kanssa oli kuitenkin kielimalleilla eniten vaikeuksia. GPT-4-turbo-kielimallilla oli yhteisesti viiden ja kymmenen esimerkin kanssa noin 80 %:n tarkkuus oikean taivutusmuodon selvittämisessä. Moisio ja muut (2024, s. 5) eivät tämän epätarkkuuden takia suosittele kielimallien käyttöä harvinaisten ja monimutkaisten sanamuotojen analysoimisen kanssa.

Hauhio ja Friberg (2024, s. 103–104) käyttivät suomen ja englannin kielen kääntämisessä erilaisia tekniikoita. He tekivät erillisen mallin, joka pystyy tunnistamaan fraaseja ja tuottavat niiden sekä terminologian kanssa listan rajoituksista. Jokaiselle termille on useita eri rajoituksia liittyen niiden eri taivutusmuotoihin. Testaamisessa he käyttivät takaisinkääntämistä, jossa rajoitetut termit käännetään takaisin alkuperäiselle kielelle toisen kielimallin avulla. Esimerkkinä takaisinkääntämisestä Hauhio ja Friberg (2024, s. 104) antoivat suomenkielisen lauseen ”Hosat ovat käteviä työkaluja.” Kielimalli oli kääntänyt tässä lauseessa olevan sanan ”hosa” englanninkieliseksi termiksi ”fire swatter”. Kun englanninkielinen lause takaisinkäännettiin suomen kielelle, niin ”hosa” oli nyt lauseessa muuttunut sanaksi ”paloswatteri”. He sitten vertasivat takaisinkäännettyä sanaa alkuperäiseen termiin ja arvioivat, kumpi oli lähempänä haluttua termiä. Jos takaisinkäännetty termi oli parempi, niin he korvasivat lähdekielen virkkeestä otetun alkuperäisen termin

sillä. Hauhio ja Friberg (2024, s. 109) arvioivat, että pelkästään tämä takaisinkääntäminen pystyisi suuresti parantamaan käännöksien laatua.

3.5 Suomen kielen oppiminen

Tässä kappaleessa käydään läpi suomen kielen oppimiseen liittyviä aikaisempia tutkimuksia. Tarkoituksena on tutkia erilaisia tapoja ja keinoja opettaa suomen kieltä ilman kielimalleja.

Lesonen ja muut (2020, s. 527) tutkivat, kuinka suomen kieltä opiskelevat ulkomaalaiset oppivat kielirakenteet verrattuna niihin, joilla suomen kieli on äidinkieli. He halusivat selvittää, voisiko ulkomaalaiset oppia samalla kohdepohjaisella oppimisella, jolla äidinkieliset oppivat oman kielensä. Kohdepohjaisessa oppimisessa kielirakenteet kehittyvät sanastollisesti erityisistä ilmaisuista tuottoisampiin ja käsitteellisiin kaavoihin. Käyttöpohjaisessa oppimisessa oppijat sen sijaan kehittyvät kielellisesti täsmällisistä lausekaavoista (esim. ”Haluan matkustaa Saksaan”) (Lesonen ja muut, 2020, s. 528). He oppivat näistä kaavoista tiettyjä kieliopillisia kuvioita (esim. kuinka voi ”haluan”-sanalla kertoa, mitä haluaa tehdä). Tutkimuksessa katsottiin neljän henkilön kielirakenteiden oppimista keskittyen sanoihin ”haluta” ja ”tykätä” yhdeksän kuukauden aikana (Lesonen ja muut, 2020, s. 527). Lesonen ja muut (2020, s. 560) selvittivät, että ulkomaalaiset voivat oppia kohdepohjaisesti. Tämä voi kuitenkin johtua muiden kielijärjestelmien kautta opituista kuvioista. Loppujen lopuksi tutkitut oppijat siirtyivät enemmän käyttöpohjaiseen oppimiseen, jossa he siirtyivät tietyistä kuvioista kaavamaisiin kuvioihin.

Kimppa ja muut (2019, s. 75) Yrittivät omassa tutkimuksessaan selvittää, kuinka suomen kieltä oppivat ulkomaalaiset saavat ja käsittelevät suomen kielen morfologiaa. Monilla vieraan kielen oppijoilla on usein vaikeuksia laajasti morfologiaa käyttävien kielten kieliopillisten näkökohtien kanssa. He usein esimerkiksi jättävät morfologisia ominaisuuksia pois tai käyttävät niitä epäsystemaattisesti. Kimppa ja muut (2019, s. 78) testasivat tutkimuksessa 15 suomalaista ja 30 saksalaista. Saksalaisista puolet olivat suomen kielen

aloittelijoita ja puolet edistyneitä, yliopiston opiskelijoita. Tutkimus tehtiin antamalla osallistujille neljä eri sanaa erilaisten morfologisten liitteiden kanssa ja mittaamalla heidän aivojen sähköistä toimintaa EEG:n avulla (Kimppa ja muut, 2019, s. 78–79). Edistyneet suomen kielen opiskelijat osasivat Kimpan ja muiden (2019, s. 87) mukaan jo aluksi vähän taivuttamista ja muista sanoista johdattelemista. Aloittelijat puolestaan näyttivät luottaneen enemmän johdannaisten täysien muotojen aktivoimiseen. Neuraaliset erilaisuudet suomea äidinkielenä puhuvien ja vieraana kielenä puhuvien välillä näyttävät pienenevän, kun kielitaito kehittyy (Kimppa ja muut, 2019, s. 88).

Salmela (2024, s. 14) oli tutkinut laajasti suomen kielen vieraana kielenä oppimista. Yhtenä syynä tutkimuksen tekemiseen annettiin ulkomaalaisten määrän nousu suomessa. Esimerkiksi Suomessa oli vuoden 2023 lopussa lähes 100 000 venäläistä, 50 000 virolaista, ja 40 000 arabialaista. Jotkin työt saattavat olla itsenäisiä, jolloin näillä ulkomaalaisilla ei ole paljoa tilanteita, joissa he voisivat harjoitella suomen kieltä. Salmelan (2024, s. 13) päämääränä oli tutkimuksessa selvittää kolme eri asiaa. Ensimmäiseksi tutkittiin, kuinka sanaston tuntemusta voisi arvioida verkkotyökalun avulla nopeasti ja luotettavasti. Toiseksi katsottiin, kuinka suomea vieraana kielenä puhuvat käsittelevät morfologisesti monimutkaisia sanoja. Kolmanneksi tehtiin morfologisuuden opetusohjelma, jonka mahdollista käyttöä suomen kielen morfologian hankkimisessa arvioitiin. Suomea vieraana kielenä opiskelevilla oli osittaisesti läpinäkyvien taivutuksien (esim. ”ilta” taivutetaan muotoon ”illalla”) kanssa eniten ongelmia (Salmela, 2024, s. 38). Opetusohjelman kanssa ohjelmaa kokeilleella testiryhmällä kehittyi morfologisen kirjoittamisen tuotantotaito suuresti verrattuna ohjelmaa käyttämättömään kontrolliryhmään (Salmela, 2024, s. 42).

Heikkola ja muut (2024, s. 118) olivat tutkineet, kuinka suomen kieltä äidinkielenä puhuvat suomalaiset saattavat ulkomaalaisen kanssa vaihtaa keskustelun kielen englanniksi. Suomalaiset saattavat pitää kielen vaihtoa kohteliaisuutena. Suomalainen osaa englannin kielen usein tarpeeksi hyvin, että keskustelu käy mahdollisesti sulavammin ja hel-

pommin englanniksi ulkomaalaisen kanssa. Ulkomaalainen ei kuitenkaan voi tällöin harjoitella suomen kielen käyttöä, eikä kielen oppiminen edisty. Heikkola ja muut (2024, s. 128) olivat tehneet tutkimuksessaan kyselyn, johonka vastasi 342 suomen kieltä puhuvaa suomalaista. Kyselyn vastauksien mukaan vain 4,5 % heistä ei koskaan vaihtanut ulkomaalaisen kanssa keskustellessaan suomen kielestä englanniksi. 55,3 % vaihtoivat englannin kieleen joskus, 20,9 % vaihtoi usein, 19 % vaihtoi harvoin ja vain yksi vastaaja vaihtoi aina. Useimmat vastasivat samaa mieltä väittämään, jossa vastaaja haluaisi auttaa ulkomaalaista puhujaa keskustelemaan helpommin kielen vaihtamisen avulla. Monet olivat myös samaa mieltä väittämän kanssa, jonka mukaan he vaihtavat kielen halutakseen päästä asiassa eteenpäin nopeammin. Jos ei suomen kieltä pysty julkisilla paikoilla usein harjoittelemaan, niin kielimallin kanssa voisi mahdollisesti päästä kielen harjoittelemisessa alkuun.

4 Tutkimusmenetelmä

Tässä tutkimuksessa on tarkoitus tutkia suurten kielimallien soveltamista suomen kielen taivutuksien opettamisessa yleisen kieliopin ja sanojen oppimisen näkökulmasta englanninkielisille. Sanojen taivuttamisessa tutkitaan kielimallien kykyä taivuttaa substantiiveja, verbejä ja adjektiiveja perusmuodoista taivutusmuotoihin ja taivutusmuodoista perusmuotoihin. Lisäksi testataan kielimallien taitoja kääntää nämä suomenkieliset sanat englanniksi. Sanojen taivuttamisen kanssa tutkitaan samalla lauseissa olevien sanojen taivutusten korjaamista. Korjaus tehdään lauseille, joissa kaikki sanat ovat perusmuodoissaan ilman mitään taivuttamista. Sanojen ja lauseiden testauksen tulokset käydään läpi verraten virallisiin aineistoihin. Tulokset poimitaan kielimallien antamista vastauksista kuvioihin ja taulukkoihin. Kuvioissa näytetään kielimallien vastauksien keskiarvoja, ja taulukoissa on kielimallien antamat vastaukset tietyille sanoille tai lauseille.

4.1 Sanojen taivuttamisen osaamisen testaaminen

Kielimallien kielioppitaitojen testaamisessa keskityttäisiin tällöin sanojen taivuttamisen osaamiseen. Jotta saadaan testattua tarpeeksi laajasti kielimallien taivuttamisen osaamista sanojen kannalta, kielimalleja tullaan testaamaan kahdesta eri näkökulmasta. Ensimmäisessä testataan perusmuodoista taivuttamista, ja toisessa testataan perusmuotoihin taivuttamista.

Perusmuodosta taivuttamisessa testataan, voiko kielimallit antaa erilaisten sanojen eri taivutusmuodot verbeille, substantiiveille sekä adjektiiveille. Perusmuodoilla tarkoitetaan verbien kanssa 1. infinitiiviä ja substantiivien sekä adjektiivien kanssa nominatiivin yksikköä. Kaikkien sanojen kanssa testataan Kielitoimiston sanakirjan (2024) antamia taivutuksia. Verbien kanssa sanakirjan antamiin taivutuksiin kuuluvat indikatiivin 1. persoonan preesens ja 3. persoonan imperfekti, konditionaalin, potentiaalin, sekä imperatiivin 3. persoonan preesens, 2. partisiippi sekä passiivin imperfekti. Substantiivien sijamuo-

doista keskitytään nominatiiviin, genetiiviin, partitiiviin ja illatiiviin yksikkö- ja monikkomuotoihin. Nämä sijamuodot ovat inessiivin ja elatiivin kanssa yleisimpiä sijamuotoja (Kettunen, 2005, s. 16). Vähintään näiden sijamuotojen osaaminen olisi aloittelijoille hyödyllistä. Adjektiivien suhteen testattaisiin substantiivin kanssa samojen sijamuotojen lisäksi komparatiivi- ja superlatiivimuotoja. Kielimallien täytyisi pystyä verbien, substantiivien ja adjektiivien perusmuodoista taivuttaa vähintään edellä mainittuihin taivutusmuotoihin ilman virheitä. Kielimalleja pyydetään myös samalla antamaan sanoille englanninkieliset käännökset. Tulosten vertaamisessa katsotaan, täsmääkö kielimallien antamat taivutukset sekä käännökset tietojoukoissa olevien vastauksien kanssa. Kielimallille annetaan sanat kehotteessa esimerkin kanssa sekä ilman esimerkkiä. Tällöin voidaan mitata, kuinka paljon esimerkin antaminen mahdollisesti auttaisi tässä tilanteessa.

Perusmuotoon taivuttamisessa testataan, voivatko kielimallit antaa taivutetulle sanalle oikean perusmuodon. Bacherikov (2024) oli omassa kielimallisovelluksessaan antanut taivutetuille sanoille perusmuodot, mutta hän käytti tähän erillistä kirjastoa kielimallien sijasta. Koska mahdollisia taivutettavia muotoja on useita, niin valittujen sanojen ja taivutuksien määrä on perusmuodosta taivuttamiseen verrattuna pienempi. Työmäärä olisi muuten ihan liian suuri, joten sanojen ja taivutuksien määrä täytyy rajata. Jotta saadaan silti monipuolisia tuloksia, niin perusmuotoon taivutettaviin muotoihin kuuluvat helppoja sekä vaikeitakin taivutuksia. Verbien kanssa testataan 3. persoonan indikatiivin imperfektistä ja potentiaalin preesensistä perusmuotoon taivuttamista. Substantiivien ja adjektiivien suhteen testataan genetiivin yksiköstä sekä illatiivin monikosta taivuttamista. Lisäksi adjektiivien kanssa kokeillaan myös superlatiivimuodosta taivuttamista.

Kummassakin testivaiheessa ja jokaisen sanan kanssa kielimallien kanssa käyty keskustelu nollataan. Kielimallit eivät tällöin muista aikaisempien keskustelujen vastauksia, jotka voivat vaikuttaa niiden antamiin tuloksiin. Kielimalleille annetut sanat on valittu ottamalla Suomen sanomalehtikielen taajuussanastosta (CSC – IT Center for Science, 2004) useita yleisimpiä sanoja, jotka taipuvat eri tavalla toisistaan. Ulkomaalaisten olisi hyvä osata yleisimpiä sanoja, joten kielimallienkin olisi hyvä osata taivuttaa niitä oikein.

Sanojen taivutuksia vertailtiin Kielitoimiston sanakirjan (2024) kanssa, jotta kaikki sanat taipuisivat vähän eri tavalla monipuolisuuden saavuttamiseksi.

Kielimalleille annetut kehotteet perustuvat kehotepohjiin, joihin laitetaan taivutettavat sanat. Tällöin saadaan helposti tehtyä valmiita kehotteita, jotka ohjeistavat kielimalleja antamaan halutut taivutukset. Kehotepohjat tehtiin verbeille, substantiiveille ja adjektiiveille suomeksi sekä englanniksi. Kehotepohjien avulla voidaan samalla testata, kuinka kielimallin antamien taivutuksien laatu voi mahdollisesti muuttua kehotteen kielen takia. Lisäksi tehtiin esimerkkivastauspohjia, jotka voi laittaa tai olla laittamatta kehotteiden loppuun. Esimerkkivastausten on tarkoitus avustaa kielimalleja vastauksen tuottamisessa. Esimerkkinä käytettäisiin minkä tahansa toisen sanan taivutuksia. Esimerkin ei ole tarkoitus antaa kielimallille suoraa vastausta, vaan selventää kielimallille, kuinka vastaus haluttaisiin. Esimerkiksi sanan ”vuosi” sekä esimerkkisanan ”asia” kanssa kehote esimerkkivastauksen kanssa on:

”Haluaisin selvittää substantiiville "vuosi" joitain taivutuksia tietyille sijamuodoille yksikössä sekä monikossa. Halutut sijamuodot ovat: nominatiivi, genetiivi, partitiivi, illatiivi. Anna lisäksi nominatiivin yksikölle enintään kolme englanninkielistä käännöstä, jos useita käännöksiä on olemassa. Esimerkkivastaus substantiivin "asia" kanssa: Yksikkö: Nominatiivi = asia; Genetiivi = asian; Partitiivi = asiaa; Illatiivi = asiaan; Monikko: Nominatiivi = asiat; Genetiivi = asioiden; Partitiivi = asioita; Illatiivi = asioihin; Englanniksi = thing, case, matter.”

Kielimallien perusmuodosta taivuttamisen testaamiseen käytetyt kehotepohjat löytyvät liitteestä 1, ja perusmuotoon taivuttamisen kehotepohjat ovat liitteessä 2.

4.2 Lauseiden taivutusten korjaamisen testaaminen

Sanojen taivuttamisen lisäksi on tarkoituksena myös testata arviointipohjaista käyttöä. Tähän kuuluisi virheellisiä taivutuksia sisältävien lauseiden korjaaminen. Tarkoituksena on kokeilla, pystyisivätkö kielimallit mahdollisesti auttamaan suomen kielen oppijoita

korjaamaan omia taivutusvirheitä. Esimerkiksi englannin kielessä ei sanoja taivuteta lauseissa samalla tavalla kuin suomen kielessä. Kielimalleja testataan pyytämällä niitä korjaamaan taivutus lauseissa, joissa kaikki sanat ovat perusmuodoissa. Jos on esimerkiksi lause ”Mikä sinun nimesi on”, niin muuttamalla kaikki sanat perusmuotoon lauseesta tulee ”Mikä sinä nimi olla”. Kielimallien on tarkoitus arvioida tämän lauseen taivutuksia ja korjata lause arviointiensa mukaisesti mahdollisimman lähelle alkuperäistä lausetta.

Lauseet valitaan kahdesta eri lähteestä, jotka eroavat toisistaan lauseiden vaikeuden kanssa. Ensimmäinen, helpommista lauseista koostuva lähde, on kokoelma yleisimpiä ohjelmille annettuja suomenkielisiä tekstityksiä OpenSubtitles-palvelussa (orgtre, 2025). Osat näistä tekstityksistä sisältävät lauseita, joita mahdollisesti käytetään usein myös jokapäiväisissä keskusteluissa. Tästä lähteestä lauseet valittiin yleisyyden, sekä muutaman kriteerin kanssa. Kriteereihin kuului, että lauseen täytyy sisältää vähintään kolme sanaa sekä vähintään yksi taivutettu sana. Lauseiden täytyi myös taipua vähän eri tavoilla, jotta testi olisi mahdollisimman laaja. Tämän laajuuden vuoksi myös otettiin kokoelmasta vähemmän yleisempiä lauseita.

Toinen lähde, joista lauseita valittiin, koostui vaikeammista lauseista, joita löytyi Suomen puupankki -korpuksesta (Helsingin yliopisto, 2023). Suomen puupankki koostuu esimerkkivirkkeistä, jotka on otettu teoksesta nimeltä Iso suomen kielioppi. Lauseet valittiin sen mukaisesti, että niissä täytyi olla vähintään kymmenen sanaa. Näillä lauseilla pystytään testaamaan kielimallien suomen kielen taivuttamisen osaamisen äärirajoja, sillä ne ovat pitkiä ja monimutkaisia. Näiden lauseiden sanojen taivutusten perusmuodosta korjaaminen alkuperäiseen on haastavaa. Jotkin lauseista sisältävät esimerkiksi verbien kanssa preesensmuodon lisäksi imperfekti-, perfekti-, tai pluskvamperfektimuodoissa olevia sanoja. Koska oikeaa taivutusmuotoa ei pysty pelkästään sanoista aina erottamaan, kielimallien ei tarvitse täysin täsmätä alkuperäisten lauseiden kanssa.

Kehotteiden kanssa käytetään samaa suunnitelmaa sanojen taivuttamisen kanssa, eli lauseidenkin kanssa testataan erilaisia kehotepohjia. Kehotepohjiin kuuluu tässäkin siten

suomenkieliset ja englanninkieliset pohjat esimerkkivastauksien kanssa. Tällöin saadaan neljä erilaista kehotetta, joihin lauseet voidaan laittaa ja sitten syöttää kielimalleille. Helpompien sekä vaikeampien lauseiden kanssa on käytetty samoja kehotepohjia. Lauseiden kanssa käytetyt kehotepohjat löytyvät liitteestä 3.

4.3 Arviointiprosessi

Tässä luvussa kerrotaan sanojen ja lauseiden arvioinnin kriteerit. Sanojen taivuttamisessa perusmuodosta sekä perusmuotoon taivuttamisessa sanojen arvioimisen kanssa on samat kriteerit. Jos kielimallin antama taivutusmuoto tai perusmuoto täsmää Kielitoimiston sanakirjan (2024) kanssa, niin vastauksesta annetaan piste. Jos kielimalli tekee vastauksessaan yhdenkin virheen, esimerkiksi sanassa on kirjoitusvirhe, niin siitä ei anneta pistettä. Välillä kielimallit sattuivat antamaan vastauksessaan yhden sijasta kaksi eri vaihtoehtoa. Näissä tilanteissa kielimalli saa pisteen, jos ensimmäinen on oikein. Ensimmäisen vastauksen voi olettaa olevan tärkeämpi vastaus, sillä käyttäjä useimmiten näkee sen ensin. Jos ensimmäinen on väärin, mutta toinen on oikein, niin vastauksesta annetaan silloin puoli pistettä.

Perusmuotoon taivuttamisen suhteen vastaukselle annetaan pisteet samoilla kriteereillä kuin perusmuodosta taivuttamisen kanssa. Tässä lisäksi oli testattu myös muodontunnistamista, eli kielimallin täytyi antaa sille taivutetun sanan muodon nimi. Kielimalli saa muodon nimeämisestä vastaukselleen pisteen, kun se täsmää Kielitoimiston sanakirjan (2024) sanalle antaman muodon nimen kanssa. Jos verbien kanssa päämuoto, esim. indikatiivi tai potentiaali, tai substantiivien ja adjektiivien kanssa pelkkä sijamuoto on oikein, niin silloin vastauksesta saa puoli pistettä. Adjektiivien superlatiivimuotojen suhteen puoli pistettä annetaan sijamuodon sijaan superlatiivin nimeämisestä. Muissa tapauksissa kielimalli ei saa vastauksestaan pisteitä.

Perusmuotojen sekä taivutusmuotojen suhteen käännösten täytyi tarkoittaa lähestulkoon samaa kuin Sanakirja.org (2011) -sivuston sanalle antamien käännösten kanssa. Jos

kielimallin sanalle antamat käännökset eivät sisällä samaa kontekstia, tai niissä on kirjoitusvirhe, niin ne ovat virheellisiä. Perusmuodoista taivuttamisen kanssa kielimalleja pyydettiin antamaan enintään kolme käännöstä, sillä sanoille oli olemassa useita eri käännöksiä. Taivutusmuotojen kanssa puolestaan pyydettiin vaan antamaan yksi käännös, koska sanojen kääntäminen ei ollut tässä tärkein tehtävä. Kielimalleja pyydetään taivutuksien kanssa antamaan vain sanan perusmuoto sekä taivutusmuodon nimi. Kolmen käännöksen kanssa kielimalli antaisi tällöin enemmän vastauksia kääntämisessä kuin taivuttamisessa. Joillekin taivutusmuodoille voi olla vaikea antaa selkeää käännöstä. Taivutusten käännösten kanssa kielimallit saavat tästä syystä pisteen oikeasta käännöksestä sanan taivutukselle tai sen perusmuodolle. Jos käännös on annettu väärälle taivutukselle, niin siitä annetaan vain puoli pistettä.

Lauseiden taivutusten korjaamisessa käytetään sanojen taivuttamiseen verrattuna eri kriteerejä, sillä lauseet ovat pidempiä ja niissä voidaan tehdä useita eri korjauksia. Lauseiden korjaamista arvioidaan nollan ja neljän pisteen väliltä. Kielimallin vastaukselle annetaan neljä pistettä, kun se täsmää alkuperäisen kanssa täysin. Kolme pistettä tulee, kun vastaus sisältää alkuperäisen kanssa saman kontekstin, mutta ei ole täysin samalla tavalla kirjoitettu. Kaksi pistettä tarkoittaa, että konteksti ei ole enää täysin sama alkuperäisen kanssa, ja lauseen rakenne ei kuulosta täysin oikealta. Yksi piste merkitsee kontekstin olevan muuttunut hyvin erilaiseksi sekä lauseen olevan jokseenkin virheellisesti kirjoitettu. Nolla pistettä on virheellinen tulos, jossa lause on kirjoitettu virheellisesti ja konteksti on muuttunut täysin tai se on mahdoton selvittää. Joidenkin lauseiden korjaaminen alkuperäiseen muotoon voi olla hyvinkin vaikeaa, sillä taivuttamattomassa lauseessa ei kerrota aikamuotoja. Kolmen pisteen yläpuolella oleminen on kielimalleilta hyvä suoritus.

4.4 Testatut suuret kielimallit

Testatut suuret kielimallit kuuluvat kuuteen eri kielimalliperheeseen. Näistä ensimmäiset kolme, jotka kuuluvat suuriin kielimalleihin, ovat ChatGPT, Gemini ja Copilot. Loput

kolme ovat pieniin kielimalleihin kuuluvat Llama, Gemma ja DeepSeek. ChatGPT-malleista testataan kahta mallia, joista toinen on tehokkaampi malli ja toinen on pienempi päättelevä malli. Muiden mallien kanssa testataan aina päättelevää mallia, jos on sellainen saatavilla.

Koska jokainen testausvaihe tehtiin eri aikoina, kielimalleja jouduttiin eri testausvaiheiden välillä vaihtamaan. Joskus uudempi versio korvaa vanhemman version, jonka jälkeen vanhempi ei ole enää mahdollisesti saatavilla. Tästä syystä eri taivutustehtävien vastauksia ei voi verrata keskenään. Sanojen perusmuodosta taivuttamisessa kielimalleina olivat ChatGPT-4o, ChatGPT o3-mini, Google Gemini 2.0 Flash Thinking, Microsoft Copilot Think Deeper, Llama 3.1 8B, Gemma 3 12B ja DeepSeek-r1 14B. Perusmuotoon taivuttamisessa ChatGPT-o3-mini päivittyi versioon o4-mini, ja Google Gemini 2.0 Flash Thinking päivittyi versioon Gemini 2.5 Flash Thinking. Näitä kahta mallia lukuun ottamatta perusmuotoon taivuttaminen sisälsi perusmuodosta taivuttamisen kanssa samat kielimallit. Lauseiden taivutusten korjaamisessa kaikki paitsi Gemini 2.5 Flash -malli olivat perusmuotoon taivuttamisen kanssa samat. Gemini 2.5 Flash -mallin Thinking-versio oli kokeellinen, ja se ei ollut saatavilla lauseiden taivutusten tutkimisen aikana.

5 Kielimallien sanojen taivutustaidot

Tässä kappaleessa esitellään kielimallien sanojen taivuttamisen tulokset. Kielimalleja oli ensin pyydetty taivuttamaan perusmuodoissa olevia sanoja useisiin eri taivutusmuotoihin. Lisäksi kielimalleja oli erikseen pyydetty antamaan taivutetuille sanoille niiden perusmuodot. Kummassakin vaiheessa kielimalleja kehotettiin antamaan sanoille myös englanninkieliset käännökset. Tulosten esittelyn jälkeen joitain kielimallien antamia virheellisiä vastauksia katsotaan syvemmin läpi. Tarkoituksena on ymmärtää kielimallien tekemiä taivutusvirheitä paremmin.

Sanojen taivutustaitojen testaamisessa kielimallien antamat vastaukset on arvioitu tiukasti. Jos kielimalli antaa kysytylle muodolle väärän muodon, tai vastauksessa on kirjoitusvirhe, vastaus tulkitaan virheelliseksi. Vastaus on täysin oikein vain, kun kielimalli antaa sanalle oikean muodon siten, että sen voi käyttäjä olettaa selkeästi oikeaksi. Kielimallit saavat antaa useampia vastauksia taivutusmuodoille, jolloin niiden ensimmäinen vastaus vaikuttaa tuloksiin. Kielimallit antoivat välillä taivutuksille yhden vastauksen sijasta kaksi vastausta. Ensimmäisen kielimallin antaman vastauksen voi olettaa olevan tärkeämpi vastaus. Jos kielimalli siten antaa kahden vastauksen kanssa virheellisen taivutuksen ensimmäisenä, mutta edes jälkimmäinen vastauksista on oikein, niin vastaus on puoliksi oikein.

5.1 Perusmuodoista taivutusmuotoihin

Tässä luvussa käydään läpi kielimalleilta sanojen perusmuodoista taivuttamisessa saadut tulokset. Alaluvuissa annetaan tulokset erikseen verbien, substantiivien, sekä adjektiivien osilta, sekä myös yhteistulokset ja sanojen kääntämisen tulokset.

Tässä osiossa testattiin neljää suurta kielimallia, jotka ovat ChatGPT-4o, ChatGPT-o3-mini, Gemini 2.0 Flash Thinking ja Microsoft Copilot ”Think Deeper” -asetuksella. Näistä malleista kaikki paitsi ChatGPT-4o käyttivät syväajattelua vastauksen tuottamisessa. Lisäksi

testattiin kolmea pientä kielimallia, johon kuuluivat Llama 3 8B, Gemma 3 12B ja DeepSeek-r1 14B. Näistä vain DeepSeek-malli oli käyttänyt syväajattelua vastauksen miettimisessä. Tämän vaiheen taivutustulokset ovat otettu 21.3. – 6.4.2025 välisenä aikana.

Sanoja tähän vaiheeseen on valittu yhteensä 22 sanaa, joista 7 ovat verbejä, 10 ovat substantiiveja, ja 6 ovat adjektiiveja. Valitut sanat sekä niiden yleisyydet Suomen sanomalehtikielen taajuussanastossa (CSC – IT Center for Science, 2004) näkyvät taulukossa 1. Taajuussanasto sisältää kokonaisuudessaan 9996 sanaa järjestyksessä yleisimmästä harvinaisempaan. Taulukossa 1 esitetään myös sanojen yleisyydet numerona, joka vastaa sanan sijaintia taajuussanastossa. Esimerkiksi sana ”olla” on sanaston ensimmäinen ja yleisin sana, joten sen yleisyys on silloin 1. Sanojen testatut taivutusmuodot näkyvät verbien suhteen taulukossa 2, substantiivien suhteen taulukossa 3, sekä adjektiivien suhteen taulukossa 4. Näissä kolmessa taulukossa perusmuodoissa olevien sanojen taustat on tummennettu erottaakseen ne taivutusmuodoistansa.

Taulukko 1. Sanomalehtikielen yleissanastosta valitut sanat ja niiden yleisyydet

	Valitut sanat	Yleisyys
Verbit	Olla	1
	Saada	10
	Tehdä	16
	Pitää	18
	Todeta	115
	Sopeuttaa	9927
	Ripotella	9979
Substantiivit	Vuosi	7
	Maa	40
	Ihminen	48
	Mies	51
	Työ	58
	Alue	75
	Notkahdus	9956
	Luokitus	9976
	Riitti	9978
	Tuliaiset	9982
Adjektiivit	Uusi	20
	Hyvä	23
	Suuri	29
	Oma	37
	Tukala	9944
	Karmea	9950

Taulukko 2. Perusmuodoista taivutettavat verbit taivutusmuotoineen

1. infinitiivi	1. per. ind. prees.	3. per. ind. imperf.	3. per. kond. prees.	3. per. pot. prees.	3. per. imperat. prees	2. partis.	passiivi
olla	olen	oli	olisi	lienee	olkoon	ollut	oltiin
saada	saan	sai	saisi	saanee	saakoon	saanut	saatiin
tehdä	teen	teki	tekisi	tehnee	tehköön	tehnyt	tehtiin
pitää	pidän	piti	pitäisi	pitänee	pitäköön	pitänyt	pidettiin
todeta	totean	totesi	toteaisi	todennee	todetkoon	todennut	todettiin
sopeuttaa	sopeutan	sopeutti	sopeuttaisi	sopeuttanee	sopeuttakoon	sopeuttanut	sopeutettiin
ripotella	ripottelen	ripotteli	ripottelisi	ripotellee	ripotelkoon	ripotellut	ripoteltiin

Taulukko 3. Perusmuodoista taivutettavat substantiivit taivutusmuotoineen

yks.nom	yks. gen	yks. part.	yks. ill	mon. nom.	mon. gen.	mon. part.	mon. ill.
vuosi	vuoden	vuotta	vuoteen	vuodet	vuosien	vuosia	vuosiin
maa	maan	maata	maahan	maat	maiden	maita	maihin
ihminen	ihmisen	ihmistä	ihmiseen	ihmiset	ihmisten	ihmisiä	ihmisiin
mies	miehen	miestä	mieheen	miehet	miesten	miehiä	miehiin
työ	työn	työtä	työhön	työt	töiden	töitä	töihin
alue	alueen	aluetta	alueeseen	alueet	alueiden	alueita	alueisiin
notkahdus	notkahduksen	notkahdusta	notkahdukseen	notkahdukset	notkahdusten	notkahduksia	notkahduksiin
luokitus	luokituksen	luokitusta	luokitukseen	luokitukset	luokitusten	luokituksia	luokituksiin
tuliaiset	tuliaisen	tuliaista	tuliaiseen	tuliaiset	tuliaisten	tuliaisia	tuliaisiin
riitti	riitin	riittiä	riittiin	riitit	riittien	riittejä	riitteihin

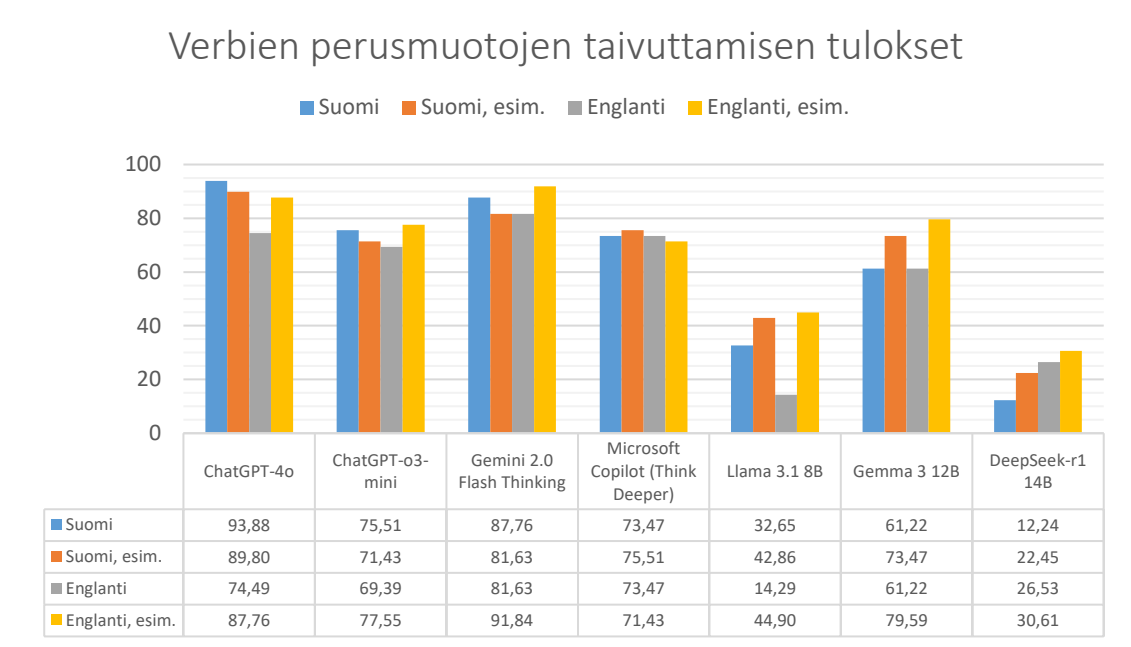
Taulukko 4. Perusmuodoista taivutettavat adjektiivit taivutusmuotoineen

yks. nom	yks. gen	yks. part.	yks. ill	mon. nom.	mon. gen.	mon. part.	mon. ill.	komp.	super.
uusi	uuden	uutta	uuteen	uudet	uusien	uusia	uusiin	uudempi	uusin
hyvä	hyvän	hyvää	hyvään	hyvät	hyvien	hyviä	hyviin	parempi	paras
suuri	suuren	suurta	suureen	suuret	suurten	suuria	suuriin	suurempi	suurin
oma	oman	omaa	omaan	omat	omien	omia	omiin	omempi	omin
tukala	tukalan	tukalaa	tukalaan	tukalat	tukalien	tukalia	tukaliin	tukalampi	tukalin
karmea	karmean	karmeaa	karmeaan	karmeat	karmeiden	karmeita	karmeisiin	karmeampi	karmein

5.1.1 Verbien taivuttaminen

Verbien taivuttaminen pyydettyihin muotoihin ei täydellisesti onnistunut. Tämä johtuu siitä, että esimerkiksi potentiaalimuodot, joista on kielimalleilta pyydetty antamaan sanoille kolmannen persoonan preesens, ovat harvinaisia. Suuret kielimallit, kuten ChatGPT-mallit, onnistuivat indikatiivimuotoon taivuttamisessa, mutta potentiaali- ja imperatiivimuotojen kanssa oli useimmiten vaikeuksia. Verbeistä potentiaali- ja imperatiivimuodot ovat harvinaisia, sillä niitä ei käytetä keskusteluissa usein. Verbien taivuttamisessa täydellisiä tuloksia voisi verrata harvinaisten muotojen takia melko kokeneeseen suomen kielen osaajaan.

Kuviossa 2 näkyy kielimallien verbien taivuttamisen tulosten keskiarvo jokaiselle neljälle eri kehotteelle. Suuret kielimallit onnistuivat verbien kanssa jokseenkin hyvin. Parhaimmat tulokset saivat ChatGPT-4o -kielimalli suomenkielisellä kehotteella ilman esimerkkiä ja Gemini 2.0 Flash Thinking -kielimalli englanninkielisellä kehotteella esimerkkivastauksen kanssa. Kummatkin saivat kyseisillä kehotteilla annettua oikean vastauksen yli 90 % ajasta. Pienet kielimallit puolestaan eivät onnistuneet yhtä hyvin. Vain Gemma 3 oli antanut onnistuneesti vastauksia englanninkielisellä kehotteella esimerkkivastauksen kanssa. Llama 3.1 ja DeepSeek-r1 antoivat sen verran usein virheellisiä tuloksia, ettei niitä voisi käyttää luotettavasti verbien perusmuotojen taivuttamiseen.



	ChatGPT-4o	ChatGPT-o3-mini	Gemini 2.0 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	93,88	75,51	87,76	73,47	32,65	61,22	12,24
Suomi, esim.	89,80	71,43	81,63	75,51	42,86	73,47	22,45
Englanti	74,49	69,39	81,63	73,47	14,29	61,22	26,53
Englanti, esim.	87,76	77,55	91,84	71,43	44,90	79,59	30,61

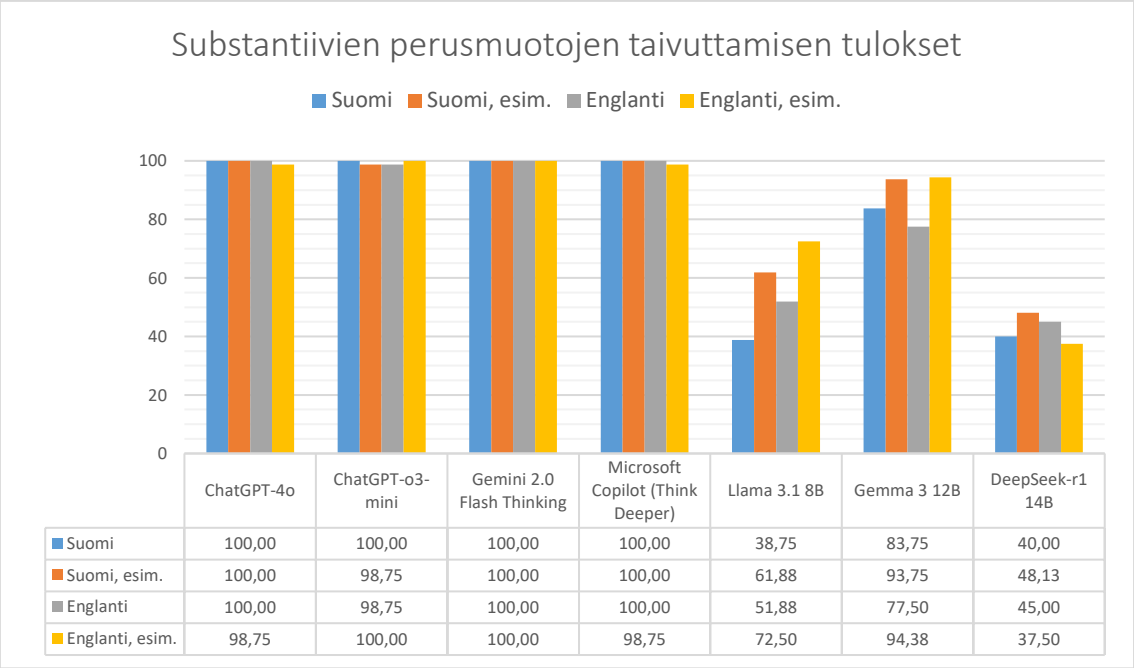
Kuvio 2. Verbien perusmuotojen taivuttamisen tulokset

On olemassa useita eri syitä sille, miksi suomenkieliset sekä esimerkittömät kehotteet antavat välillä parempia tuloksia joidenkin mallien kanssa. Satunnaisuus voi olla yksi syistä, sillä tulokset oli testattu jokaisen sanan ja kehotteen kohdalta vain kerran. Kielimalli saattaa myös ymmärtää suomenkielisen kehotteen paremmin, sillä se ei ehkä aina hahmota tehtävää yhtä hyvin englanniksi. Jos esimerkki ei ole tarpeeksi selkeä, niin kielimalli saattaa ymmärtää sen väärin. Kehotteessa annettu esimerkki voi myös johdattaa kielimallia harhaan, jos siinä käytetty sana taipuu eri tavalla kuin mallille annettu sana.

Lisäksi kielimallien koulutuksessa käytetyt aineistot saattoivat sisältää suomen kielen taivutuksiin liittyviä virheitä. Nämä aineistot perustuvat ihmisten kirjoittamiin teksteihin, ja niiden suuren määrän vuoksi osassa niistä saattoi olla virheellistä tietoa. Nämä kaikki syyt mahdollisesti vaikuttavat siihen, miksi esimerkittömät kehotteet tuottivat välillä parempia tuloksia kuin esimerkkivastauksen sisältävät kehotteet.

5.1.2 Substantiivien taivuttaminen

Kuten kuviosta 3 huomaa, Suuret kielimallit onnistuvat substantiivien kanssa ilman mitään suurempia ongelmia. Kaikki näistä malleista ovat tehneet virheen taivuttamisessa enintään vain kerran. Nämä virheet olivat myös johtuneet pääasiassa kirjoitusvirheistä, joista kerrotaan tarkemmin luvussa 5.3. Tässä on kuitenkin testattu pääasiassa yleisimpiä sijamuotoja, joten tuloksia voi verrata aloittelijan taitoihin. Pienistä kielimalleista Gemma 3 onnistuu myös substantiivien kanssa lähestulkoon yhtä hyvin kuin suuret kielimallit. Llama 3.1 onnistui melko hyvin englanninkielisen ja esimerkkivastauksen sisältävän kehotteen avulla. DeepSeek-r1 onnistui parhaimmillaan vain alle puolet ajasta. Llama 3.1 tuotti parhaimmat tulokset esimerkkivastauksen kanssa, ja se hyötyi myös englanninkielisestä kehotteesta. Gemma 3 puolestaan tuotti parhaiten vastauksia vain esimerkkikehotteiden kanssa, eikä kielellä näyttänyt tässä olevan paljoa vaikutusta. DeepSeek-r1 -mallin tulosten mukaan se onnistui parhaiten suomeksi esimerkkivastauksen kanssa.

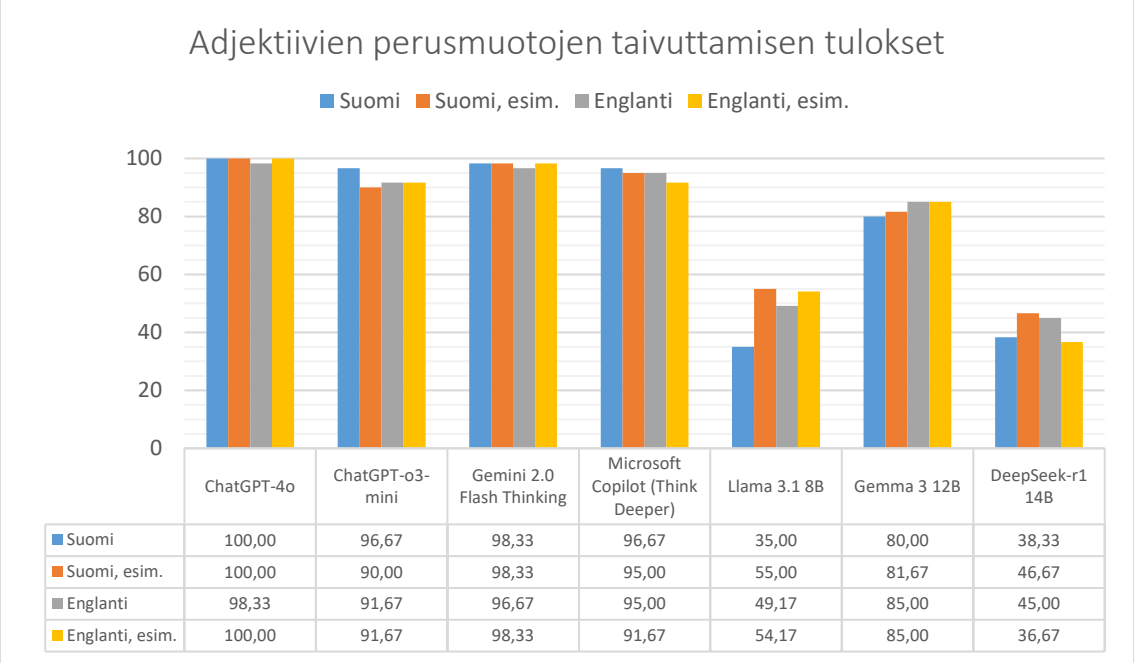


	ChatGPT-4o	ChatGPT-o3-mini	Gemini 2.0 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	100,00	100,00	100,00	100,00	38,75	83,75	40,00
Suomi, esim.	100,00	98,75	100,00	100,00	61,88	93,75	48,13
Englanti	100,00	98,75	100,00	100,00	51,88	77,50	45,00
Englanti, esim.	98,75	100,00	100,00	98,75	72,50	94,38	37,50

Kuvio 3. Substantiivien perusmuotojen taivuttamisen tulokset

5.1.3 Adjektiivien taivuttaminen

Adjektiivien kanssa kielimalleilla oli kuvion 4 mukaan taivuttaminen onnistunut vähän substantiiveja heikommin. Suuret kielimallit olivat kuitenkin silti onnistuneet taivuttamisessa hyvin. Adjektiivien kanssa on myös sijamuotojen lisäksi pyydetty komparatiivi- ja superlatiivimuotoja. Kielimallit olivat joutuneet antamaan paljon vastauksia, jotka mahdollisesti vaikuttavat tulokseen negatiivisesti. Tuloksia voisi siten verrata vähän kokeneemman aloittelijan osaamiseen. Pienten kielimallien vastausten laatu heikentyi substantiiveihin verrattuna, mutta Gemma 3 -mallilla englanninkielisen esimerkkivastauksettoman kehotteen kanssa laatu oli noussut. Kehotteen kielellä tai esimerkkivastauksen antamisella ei ollut suurten kielimallien kanssa paljoa vaikutusta. Suuret kielimallit olivat onnistuneet tuottamaan pääasiassa samanlaatuisia tuloksia kehotteesta riippumatta. Pienten kielimallien kanssa Gemma 3 hyötyi englanninkielisestä kehotteesta, kun taas Llama 3.1 hyötyi eniten esimerkistä. DeepSeek-r1 -malli puolestaan hyötyi esimerkistä vain suomeksi. Kehotteiden vaikutus DeepSeek-r1 -mallin toimintaan adjektiivien kanssa näyttää olevan substantiivien kanssa samanlainen.



	ChatGPT-4o	ChatGPT-o3-mini	Gemini 2.0 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	100,00	96,67	98,33	96,67	35,00	80,00	38,33
Suomi, esim.	100,00	90,00	98,33	95,00	55,00	81,67	46,67
Englanti	98,33	91,67	96,67	95,00	49,17	85,00	45,00
Englanti, esim.	100,00	91,67	98,33	91,67	54,17	85,00	36,67

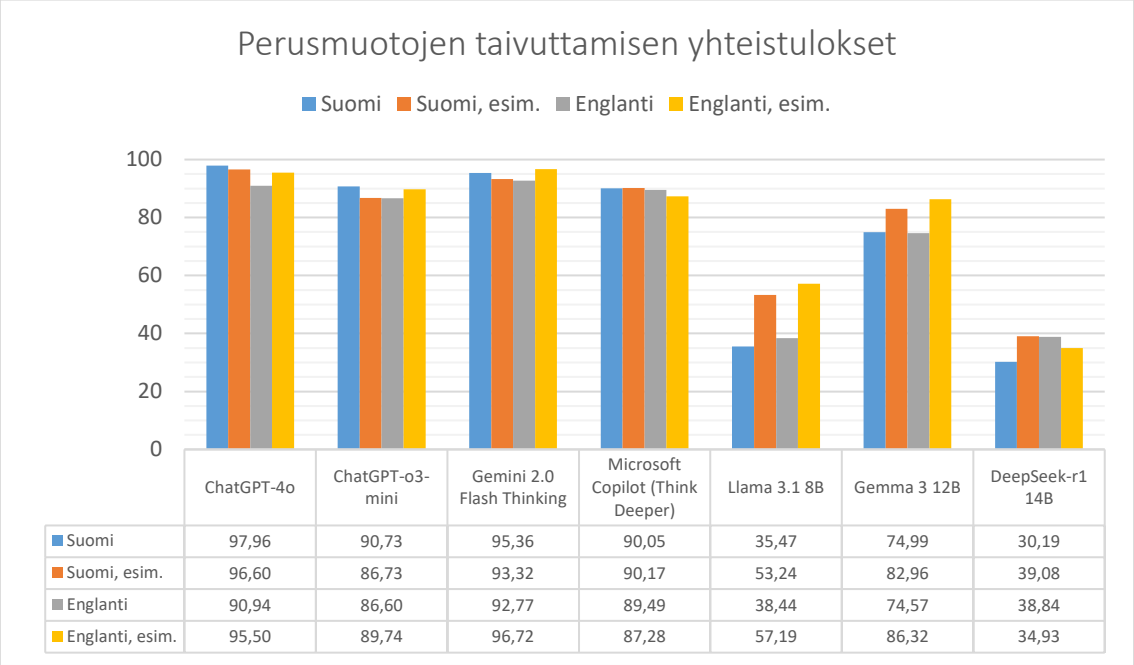
Kuvio 4. Adjektiivien perusmuotojen taivuttamisen tulokset

5.1.4 Taivuttamisen tulosten yhteenveto

Kuviossa 5 näkyy kielimalleilla sanojen perusmuodoista taivuttamisen onnistumisen lopputulos jokaisen kielimallin ja kehotteen kanssa prosentteina. Suurimmilta osin kielimalleista suuret kielimallit onnistuivat taivuttamisessa. Näistä malleista parhaiten oli suoriutunut ChatGPT-4o -malli. Seuraavaksi parhain oli Gemini 2.0 Flash Thinking, joka suoriutui parhaiten englanniksi esimerkkivastauksen kanssa.

Kaikki suurista kielimalleista onnistuivat melko hyvin, mutta pienet kielimallit eivät onnistuneet yhtä hyvin. Pienistä kielimalleista vain Gemma 3 12B -malli oli onnistunut hyvin, varsinkin kun sille oli annettu kehote englanniksi esimerkkivastauksen kanssa. Llama 3.1 8B -malli ei englanniksi ymmärtänyt kehotetta lähestulkoon yhtään, mutta esimerkkivastausten kanssa tai suomeksi se välillä onnistui oikeiden vastauksien antamisessa. DeepSeek-r1 14B -malli puolestaan onnistui heikoiten, varsinkin suomenkielisellä kehotteella

ilman esimerkkivastausta. Llama 3.1 saattaa olla liian pieni malli ymmärtääkseen suomen kieltä tarpeeksi hyvin. DeepSeek puolestaan ei ole keskitetty kielten osaamiseen, jolloin se mahdollisesti vaatisi vielä isomman mallin suomen kielen osaamiseen. Gemma 3 on optimoitu monikielisyyttä varten, joten se onnistuu muihin pieniin kielimalleihin verrattuna paljon paremmin.



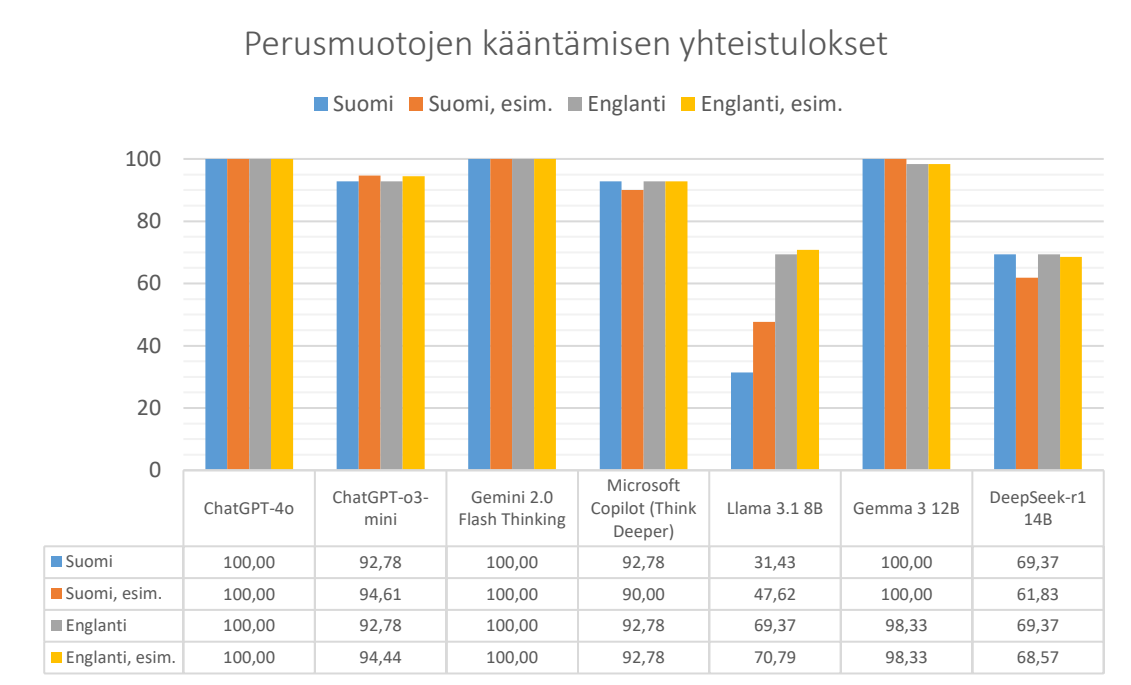
	ChatGPT-4o	ChatGPT-o3-mini	Gemini 2.0 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	97,96	90,73	95,36	90,05	35,47	74,99	30,19
Suomi, esim.	96,60	86,73	93,32	90,17	53,24	82,96	39,08
Englanti	90,94	86,60	92,77	89,49	38,44	74,57	38,84
Englanti, esim.	95,50	89,74	96,72	87,28	57,19	86,32	34,93

Kuvio 5. Perusmuotojen taivuttamisen yhteistulokset

5.1.5 Perusmuotojen kääntäminen suomesta englanniksi

Kielimalleja oli myös pyydetty antamaan sanoille käännökset englanniksi. Käännetyillä sanoilla täytyi olla alkuperäisten sanojen kanssa sama konteksti, jotta ne saisivat pisteen. Käännökset vahvistettiin Sanakirja.org (2011) -sivuston käännöksien mukaisesti. Kuvio 6 näyttää sanojen kääntämisen tulosten keskiarvot jokaiselle neljälle kehotteelle ja kielimallille. Kuvion mukaisesti lähestulkoon kaikki mallit onnistuivat kääntämään sanat vähintään samaa tarkoittaviin englanninkielisiin sanoihin. Llama-mallilla oli eniten ongelmia suomenkielisen kehotteen kanssa. Kehotteen kielellä tai esimerkillä ei muiden mallien kanssa ollut suurta vaikutusta käännösten laatuihin. Niiden eroavaisuudet olivat sen

verran pienet, että ne voivat johtua pelkästään satunnaisuudesta. DeepSeek osasi antaa yleisimmille sanoille oikeat käännökset, mutta vaikeampien sanojen kanssa se ei onnistunut yhtä usein. Todennäköisesti DeepSeek-mallin vähäinen keskittyminen monikielisyyteen on vaikuttanut negatiivisesti myös sen kykyyn kääntää sanoja.



	ChatGPT-4o	ChatGPT-o3-mini	Gemini 2.0 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	100,00	92,78	100,00	92,78	31,43	100,00	69,37
Suomi, esim.	100,00	94,61	100,00	90,00	47,62	100,00	61,83
Englanti	100,00	92,78	100,00	92,78	69,37	98,33	69,37
Englanti, esim.	100,00	94,44	100,00	92,78	70,79	98,33	68,57

Kuvio 6. Perusmuotojen kääntämisen yhteistulokset

5.2 Taivutusmuodoista perusmuotoihin

Tässä luvussa testataan, kuinka hyvin kielimallit pystyvät antamaan taivutetun sanan perusmuodon. Perusmuotoihin taivuttamista testataan vain kahden eri taivutusmuodon kanssa. Taulukossa 5 näkyy kaikki sanat ja taivutukset, joita tämän luvun testeissä käytettiin. Malleilta pyydetään sanojen takaisin perusmuotoon taivuttamisen kanssa myös taivutettujen sanojen taivutusmuodot, jotka ovat myös taulukossa 5.

Taulukko 5. Perusmuotoihin taivutettavat sanat

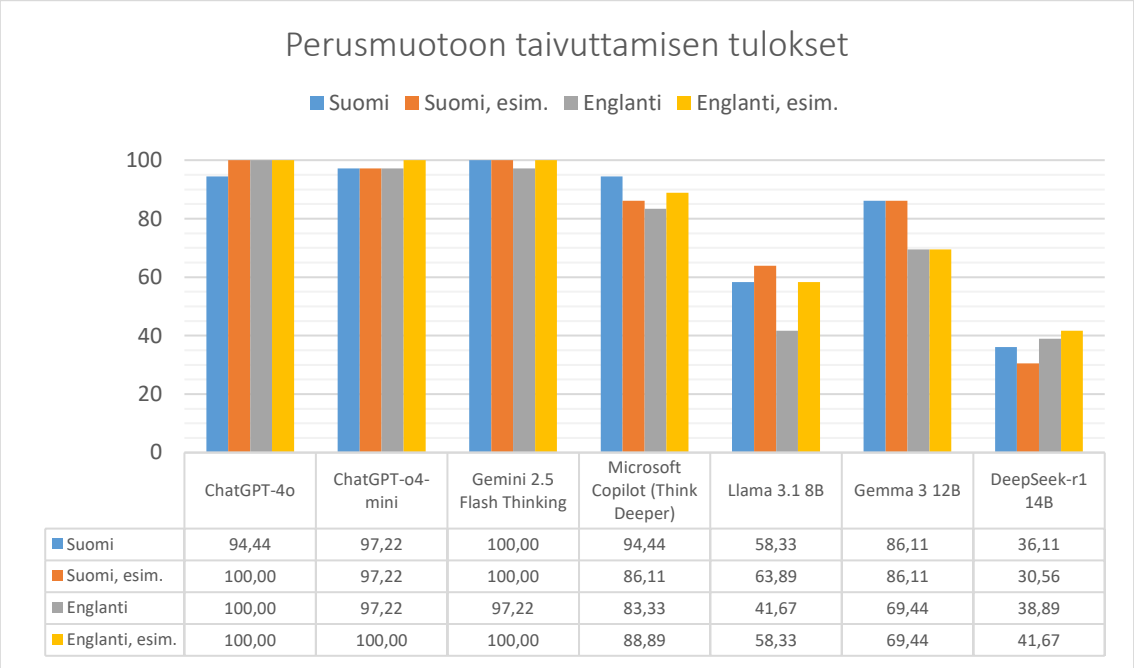
	Valittu sana	Muoto	Perusmuoto
Verbit	oli	3. per. ind. imperf.	olla
	lienee	3. per. pot. prees.	
	totesi	3. per. ind. imperf.	todeta
	todennee	3. per. pot. prees.	
	ripotteli	3. per. ind. imperf.	ripotella
	ripotellee	3. per. pot. prees.	
Substantiivit	vuoden	gen. yks.	vuosi
	vuosiin	ill. mon.	
	notkahduksen	gen. yks.	notkahdus
	notkahduksiin	ill. mon.	
	riitin	gen. yks.	riitti
	riitteihin	ill. mon.	
Adjektiivit	hyvän	gen. yks.	hyvä
	hyviin	ill. mon.	
	paras	super. nom. yks.	
	tukalan	gen. yks.	tukala
	tukaliin	ill. mon.	
	tukalin	super. nom. yks.	

Testatut suuret kielimallit ovat tässä testissä ChatGPT-4o, ChatGPT-o4-mini, Gemini 2.5 Flash Thinking (preview) ja Microsoft Copilot Think Deeper. Näistä malleista kaikki paitsi ChatGPT-4o käyttivät tehtävässä syväajattelua. Pienien kielimallien suhteen testattuina malleina olivat Llama 3.1 8B, Gemma 3 12B ja DeepSeek-r1 14B. Tämän vaiheen testit tehtiin 14.5 – 20.5.2025 -välisenä aikana. Koska ensimmäisen vaiheen testeistä oli kulunut yli kuukausi, Pari kielimallia oli ehtinyt päivittyä uudempaan versioon. ChatGPT-o3-mini päivittyi versioon o4-mini, ja Gemini 2.0 Flash Thinking päivittyi versioon 2.5.

5.2.1 Perusmuotojen selvittämisen tulokset

Kuviossa 7 on keskiarvot sanojen perusmuotoihin taivuttamisen tuloksille kaikilla malleilla ja kehotteilla. Kokonaisuudessaan suuret kielimallit onnistuivat tässä melko hyvin. ChatGPT-4o ja o4-mini -kielimallit antoivat virheellisiä vastauksia vain muutamia kertoja, ja Gemini 2.5 Flash Thinking tuotti vähiten virheitä. Copilot Think Deeper puolestaan tuotti heikoimmat tulokset suurista kielimalleista. Pienistä kielimalleista Gemma 3 toimi

parhaiten suomeksi. Englanninkielisellä kehotteella sen vastausten laatu heikentyi. Esimerkkivastausten olemassaolo kehotteessa näytti hyödyttävän kaikkia malleja englanniksi, mutta ne näyttivät auttaneen eniten Llama 3.1 -mallia. Tulosten mukaan perusmuotoon taivuttaessakin esimerkkivastaukset näyttävät auttavan kielimalleja. DeepSeek-r1 ei suomeksi hyötynyt esimerkistä lainkaan. Tämä voi kyseisen mallin kanssa johtua suomen kielen osaamisen puutteesta.



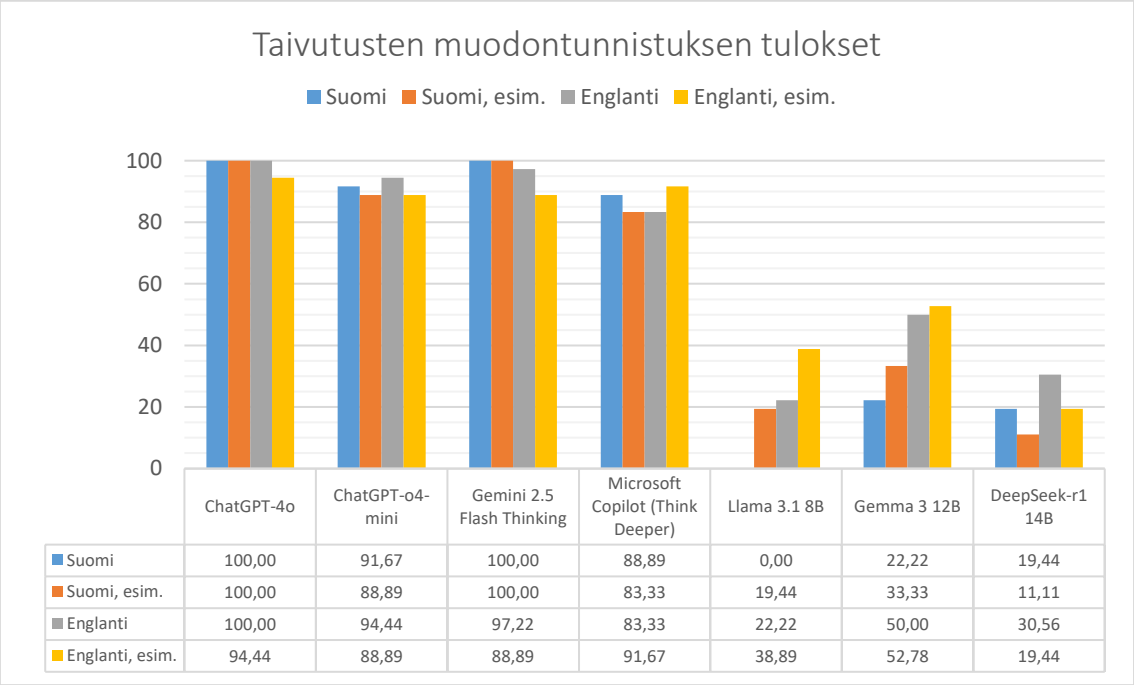
	ChatGPT-4o	ChatGPT-o4-mini	Gemini 2.5 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	94,44	97,22	100,00	94,44	58,33	86,11	36,11
Suomi, esim.	100,00	97,22	100,00	86,11	63,89	86,11	30,56
Englanti	100,00	97,22	97,22	83,33	41,67	69,44	38,89
Englanti, esim.	100,00	100,00	100,00	88,89	58,33	69,44	41,67

Kuvio 7. Perusmuotoon taivuttamisen tulokset

5.2.2 Taivutusten muodontunnistus

Taivutuksien muotojen tunnistamisessa vain suuret kielimallit olivat osanneet vastata suurimmilta osin oikein. Gemini 2.5 Flash Thinking toimi parhaiten suomenkielisellä kehotteella, mutta muuten kehotteen kieli ei vaikuttanut suurten kielimallien vastauksiin paljon. Esimerkkivastaukset eivät tässä testissä näyttäneet vaikuttavan muodontunnistamisen laatuun suurien kielimallien kanssa Copilot-mallia lukuun ottamatta. Pienet kielimallit eivät onnistuneet muodon tunnistamisessa. Llama 3.1 -malli ei suomeksi saanut ilman esimerkkivastausta annettua yhtään oikeaa muotoa, sillä se unohti välillä antaa

vastauksessaan myös taivutuksen muodon. Gemma 3 onnistui englanninkielisellä kehotteella noin puolet ajasta muodontunnistamisessa.



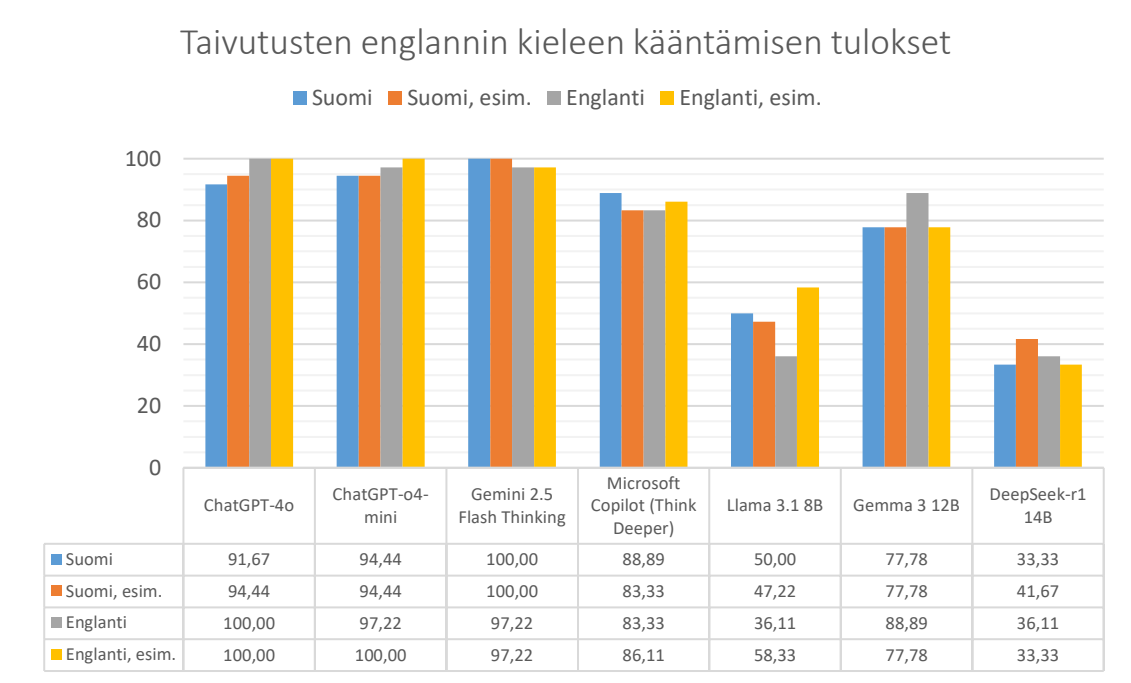
	ChatGPT-4o	ChatGPT-o4-mini	Gemini 2.5 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	100,00	91,67	100,00	88,89	0,00	22,22	19,44
Suomi, esim.	100,00	88,89	100,00	83,33	19,44	33,33	11,11
Englanti	100,00	94,44	97,22	83,33	22,22	50,00	30,56
Englanti, esim.	94,44	88,89	88,89	91,67	38,89	52,78	19,44

Kuvio 8. Taivutusten muodontunnistuksen tulokset

5.2.3 Taivutusten kääntäminen suomesta englanniksi

Kielimallien taivutettujen sanojen kääntämisen keskiarvotulokset näkyvät kuviossa 9. Kielimallin antama käännös on tästä syystä arvioitu oikeaksi, kun sen konteksti täsmää kielimallille annetun taivutetun sanan tai sen perusmuodon kanssa. Tämän takia suuret kielimallit onnistuivat parhaimmillaan yli 80 % ajasta taivutusten kääntämisessä. Pienistä kielimalleista vain Gemma 3 oli onnistunut aina yli puolet ajasta kääntämisessä. Esimerkit eivät tässä auttaneet kaikkia kielimalleja. Varsinkin Gemma 3 -malli tuotti englanniksi esimerkin kanssa paljon huonommat tulokset esimerkittömään verrattuna. Tämä voi johtua pelkästään satunnaisuudesta, mutta esimerkki saattoi lisäksi tehdä kääntämistehtävän epäselkeämmäksi. Esimerkin painotus taivuttamistehtävään voi myös vaikuttaa negatiivisesti kääntämisen laatuun. Kuitenkin esimerkiksi ChatGPT-mallit olivat hyötyneet

esimerkkivastauksista. Englanninkielinen kehote näytti tässä tehtävässä tuottaneen parhaimmat tulokset.



	ChatGPT-4o	ChatGPT-o4-mini	Gemini 2.5 Flash Thinking	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	91,67	94,44	100,00	88,89	50,00	77,78	33,33
Suomi, esim.	94,44	94,44	100,00	83,33	47,22	77,78	41,67
Englanti	100,00	97,22	97,22	83,33	36,11	88,89	36,11
Englanti, esim.	100,00	100,00	97,22	86,11	58,33	77,78	33,33

Kuvio 9. Taivutusten englannin kieleen kääntämisen tulokset

5.3 Kielimallien sanojen taivuttamisessa tekemät virheet

Tämän luvun tarkoitus on selventää, miten kielimallit saivat sanojen taivuttamisen tulokset. Joitain suurten sekä pienten kielimallien antamia virheellisiä taivutuksia käydään läpi syvemmin, jotta saadaan selville eri kielimallien tekemiä virheitä liittyen sanojen taivuttamiseen.

5.3.1 Perusmuodosta taivuttamisen virheet

Suuret kielimallit tekivät virheellisissä vastauksissaan pääasiassa kirjoitusvirheitä, mutta joskus muotokin menee väärin. Taulukossa 6 näkyy esimerkkinä ChatGPT-4o-kielimallin vastaukset verbin ”todeta” taivuttamiselle ja kääntämiselle. Taulukossa sanat, joiden

taustat on värjätty, ovat virheellisiä eivätkä siten täsmää alkuperäisen kanssa. Taulukosta huomaa, että esimerkiksi 1. persoonan indikatiivin preesensissä ChatGPT-4o antoi suomenkielisen esimerkkikehotteen kanssa sanan ”totean” sijasta ”toteam”. Se oli vahingossa kirjoittanut sanan viimeisen kirjaimen väärin. Muita esimerkkejä kirjoitusvirheistä tässä taulukossa ovat "totennee”, ”totenee” ja ”totekoon”. Nämä ovat melko lähellä alkuperäistä sanakirjan antamaa muotoa, mutta eroavat vain muutamalla kirjaimella. Lisäksi välillä mallit taulukon mukaisesti olivat myös kirjoittaneet väärän muodon ilman muita virheitä. Esimerkkinä tälläisistä virheistä ovat taulukossa 6 ”todettaisiin”, ”todettaneen”, ja ”todettakoon”.

Taulukko 6. Sanan "todeta" taivuttamisen vastaukset ChatGPT-4o-mallilta

		ChatGPT-4o			
Taivutusmuodot	Alkuperäinen	Suomi	Suo-esim.	Englanti	Eng-esim.
1. infinitiivi	todeta	todeta	todeta	todeta	todeta
1. per. ind. prees.	totean	totean	toteam	totean	totean
3. per. ind. imperf.	totesi	totesi	totesi	totesi	totesi
3. per. kond. prees.	toteaisi	toteaisi	toteaisi	todettaisiin	toteaisi
3. per. pot. prees.	todennee	todettaneen	totennee	todettaneen	totenee
3. per. imperat. prees.	todetkoon	todetkoon	todetkoon	todettakoon	totekoon
2. partis.	todennut/todettu	todettu	todennut	todettu	todennut
pass. imperf.	todettiin	todettiin	todettiin	todettiin	todettiin
Englanniksi	state	to state	state	to state	state
	note	to ascertain	ascertain	to note	ascertain
	notice	to determine	determine	to ascertain	observe

Paikalliset mallit, varsinkin Llama ja DeepSeek, tekivät paljon erilaisia virheitä. Taulukossa 7 annetaan esimerkkinä virheistä Llama 3.1 -mallin antamat vastaukset verbin ”tehdä” perusmuodosta taivuttamiselle. Kaikki taulukossa olevat sanat, joiden taustat on värjätty, ovat virheellisiä vastauksia. Taulukosta huomaa, että suomenkielisten kehotteiden kanssa malli oli tämän sanan kanssa unohtanut antaa kehotteen ohjeiden mukaisesti käännökset englanniksi. Tämä virhe tapahtui välillä englanniksikin, mutta paljon harvemmin. Mahdollisena syynä voi olla, että Llama 3.1 ja DeepSeek eivät ymmärtäneet suomenkielistä kehotetta tarpeeksi hyvin. Suomen kielen osaamisen puutteen takia nämä kielimallit eivät ehkä tajunneet, että niiden pitäisi myös antaa sanoista englanninkieliset käännökset. Toisena syynä voi myös olla, että taivuttaminen on mainittu kehotteessa

useampaan kertaan ja tarkemmilla ohjeilla. Kääntäminen sen sijaan on mainittu kehotteessa lisäohjeena. Kielimallit saattoivat tällöin keskittyä taivuttamistehtävään ja unohtivat, että niiden pitäisi myös antaa käännökset.

Taulukko 7. Sanan "tehdä" taivuttamisen vastaukset Llama 3.1 8B -mallilta

		Llama 3.1 8B			
Taivutusmuodot	Alkuperäinen	Suomi	Suo-esim.	Englanti	Eng-esim.
1. infinitiivi	tehdä	tehdä	tehdä	tehdä	tehdä
1. per. ind. prees.	teen	tekee	teen	tehnän	teen
3. per. ind. imperf.	teki	teki	teki	teki	teki
3. per. kond. prees.	tekisi	olisi tehnyt	tekisi	tehtäisi	tekisi
3. per. pot. prees.	tehnee	voisi tehdä	tekissee	tehtävän	teknään
3. per. imperat. prees.	tehköön	tee	teen	tehdi	tee!
2. partis.	tehnyt	tehnyt	tehnyt	tehdekää	tehnyt
pass. imperf.	tehtiin	tehtiin	tehtiin	tehtäköön	tehtiin
Englanniksi	do perform make			to make to do to produce	do make perform

DeepSeek-r1-kielimallilla oli muihin malleihin verrattuna eniten ongelmia taivuttamisessa. Useimmiten tämä johtui siitä, ettei se aina tiennyt, mitä annettu suomenkielinen sana tarkoitti. Taulukossa 8 näytetään DeepSeek-r1-mallin antamat vastaukset substantiivin ”riitti” perusmuodosta taivuttamiseen. Kaikki sanat, joiden taustat on taulukossa värjätty punertavaksi, ovat virheellisiä vastauksia. Keltaiseksi värjätty tausta tarkoittaa, että vastauksesta annettiin puoli pistettä. Taulukosta näkee esimerkiksi käännöksien kautta, että DeepSeek-kielimalli ei minkään kehotteen kanssa tiennyt, mitä riitti oikeasti tarkoitti. Llama 3.1-mallia lukuun ottamatta kaikki kielimallit tiesivät, että riitti vastaa englanniksi sanoja ”ritual”, ”rite” ja ”ceremony”. Llama 3.1 oli myös luullut kerran riitin tarkoittavan rytmiä. DeepSeek-mallin mielestä riitti tarkoitti englanninkielisen käännöksen mukaan aina sanaa ”reitti”. Voi olla, että kielimalli kuvitteli kääntämisvaiheessa ensimmäisen i-kirjaimen olevan kirjoitusvirhe. Englanninkielisen esimerkkivastauskehotteen kanssa näkyy myös välillä toistuva virhe, jossa se jättää antamatta kaikkia pyydettyjä taivutuksia tai käännöksiä. Substantiivin ”riitti” kanssa taulukon mukaisesti DeepSeek-malli oli unohtanut antaa partitiivin ja illatiivin yksikkö- ja monikkomuodot, sekä englanninkielisen käännöksen. Syy puutteellisiin vastauksiin voi olla, että ajatteluvaiheessaan

se mietti vastaustaan liian pitkään. DeepSeek ei ollut varma vastauksestaan, joten se oli miettinyt englanninkielisen esimerkkivastauskehotteen kanssa vastausta pitkään. Pienet kielimallit voivat käsitellä vain tietyn määrän kontekstia, ja vastaustaan miettiessään se saattoi mahdollisesti unohtaa joitain pyydettyjä asioita.

Taulukko 8. Sanan "riitti" taivuttaminen DeepSeek-r1 14B -mallilla

		DeepSeek-r1 14B			
Sijamuodot	Alkuperäinen	Suomi	Suo-esim.	Englanti	Eng-esim.
nom. yks.	riitti	riitti	riitti	riitti	riitta
gen. yks.	riitin	riitin	riitteen/riitin	riitin	riittaa
part. yks.	riittiä	riitta	riitta	riitta	
ill. yks.	riittiin	riitteen	riitteen	riittii	
nom. mon.	riitit	teillä	riitteet	riitit	riitut
gen. mon.	riittien	teidän	riitojen	riitten	riitojen
part. mon.	riittejä	teillä	riiteita	riittja	
ill. mon.	riitteihin	teeseen	riitteen	riittihin	
Englanniksi	rite	path way route	path road route	road street path	

5.3.2 Perusmuotoihin taivuttamisen virheet

Perusmuotoihin taivuttaessa kielimallit onnistuivat tulosten mukaan selkeästi paremmin kuin perusmuodoista taivuttaessa. Perusmuotoihin taivuttaminen sisälsi vähemmän vastattavaa perusmuodosta taivuttamiseen verrattuna, joten tehtävä oli yksinkertaisempi ja selkeämpi. Virheitä silti tapahtui vastauksissa kaikilta malleilta.

Joskus kielimallit olivat antaneet joillekin taivutetuille sanoille väärän, mutta lähes samalla lailla kirjoitetun, sanan. Taulukossa 9 on ChatGPT-4o -kielimallin antamat vastaukset substantiivin ”riitteihin” perusmuotoon taivuttamisessa. ChatGPT-4o-malli oli jokaisessa kohdassa vastannut, että sanan ”riitteihin” perusmuoto on ”riite”, vaikka haluttu vastaus pitäisi olla ”riitti”. Tämä johtuu siitä, että kummankin sanan illatiivin monikot ovat

Kielitoimiston sanakirjan (2024) mukaan homonyymejä keskenään. Riite on riittiä harvinaisempi sana, sillä sitä ei löydy edes Suomen sanomalehtikielen taajuussanastosta (CSC – IT Center for Science, 2004). ChatGPT-4o oli silti päättänyt, että taivutettu sana ”riitteihin” perustui sanan ”riitti” sijaan sanaan ”riite”. Tämän saman virheen oli myös tuottanut Copilot-kielimalli englanninkielisellä kehotteella. Copilotin syväajatteluversio perustuu OpenAI:n o1 -malliin (The Copilot Team, 2025). Kummatkin o1 ja ChatGPT-4o ovat OpenAI:n suurimpia malleja. o1-mallin kouluttamisessa käytetty tietokanta voi mahdollisesti olla ChatGPT-4o-mallin kanssa samanlainen. Pienempi ChatGPT-o4-mini oli kuitenkin silti antanut sanalle perusmuodoksi ”riitti”, joten sen tietokanta näyttää olevan erilainen. Vain suurimpien OpenAI:n tekemien mallien tietokannat näyttävät suosivan riitettä riitin sijaan.

Taulukko 9. Sanan "riitteihin" perusmuotoon taivuttaminen ChatGPT-4o-mallilla

Taivutettu sana = riitteihin		ChatGPT-4o			
	Kehote	Suomi	Suo-esim.	Englanti	Eng-esim.
Sanan perusmuoto	riitti	riite	riite	riite	riite
Taivutusmuoto	ill. mon.	ill. mon.	ill. mon.	ill. mon.	ill. mon.
Käännös		frost	hoarfrosts	into the frosts	into the frost patterns

6 Kielimallien lauseiden taivutusvirheiden korjaamistaidot

Tässä luvussa esitellään kielimallien suomenkielisten lauseiden taivutusvirheiden korjaamisen testaustulokset. Testaaminen tehtiin antamalla kielimalleille lauseita, joidenka sanat ovat kaikki perusmuodoissaan, eli sanoja ei ole taivutettu lainkaan. Lauseet on jaettu kahteen kategoriaan. Ensimmäinen kategoria koostuu helpommista lauseista. Näiden valinta perustuu OpenSubtitles-tekstitysten (orgtre, 2025) yleisimpiin ja vähemmän yleisimpiin lauseisiin. Toinen kategoria koostuu vaikeammista lauseista. Nämä perustuvat Suomen puupankki-korpuksesta (Helsingin yliopisto, 2023) löytyviin lauseisiin. Tarkoituksena on sanojen testaamisen tavoin testata kielimallien osaamista laajasti painottaen samalla yleisempiin lauseisiin. Lisäksi tulosten jälkeen joitain kielimallien virheellisiä korjauksia käydään tarkemmin läpi, jotta voidaan tietää syyt taivutuskorjaamisessa tehdyille virheille.

Testattuihin malleihin tässä tutkimuksessa kuuluvat ChatGPT-4o, ChatGPT-o4-mini, Gemini 2.5 Flash, Microsoft Copilot, Llama 3.1 8B, Gemma 3 12B ja DeepSeek-r1 14B. Testitulosten keräämisen aikana Gemini 2.5 Flash Thinking ei ollut enää saatavilla. Tässä tehtävässä kyseinen kielimalli jouduttiin tästä syystä vaihtamaan Gemini 2.5 Flash -malliin.

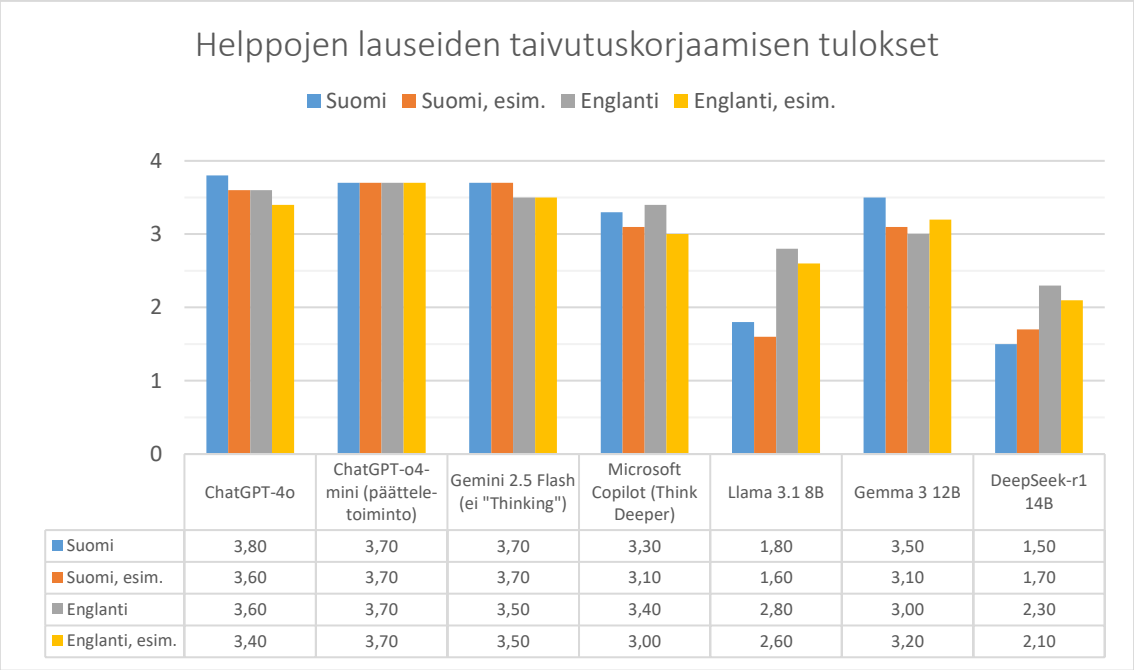
6.1 Helpompien lauseiden korjaaminen

Taulukossa 10 annetaan tässä alaluvussa testatut helpommat lauseet, niiden alkuperäiset muodot sekä niiden yleisyys OpenSubtitles-korpuksessa (orgtre, 2025). Koska perusmuotoisista lauseista olisi vaikea muuten erottaa persoonamuotoja, joihinkin lauseisiin on niiden perusmuodoissa lisätty persoonapronomini. Nämä lisätyt persoonapronominit perusmuotolauseille näkyvät taulukossa suluissa.

Taulukko 10. Taivutuskorjaamisen helpommat lauseet

Alkuperäinen lause	Perusmuotolause	Yleisyys
Mitä sinä teet?	Mitä sinä tehdä?	36
Onko kaikki hyvin?	Olla kaikki hyvä?	91
Kaikki on hyvin.	Kaikki on hyvä.	95
Hän on kuollut.	Hän olla kuollut.	162
Olen todella pahoillani.	(minä) Olla todella pahoilla.	349
Mitä haluat minun tekevän?	Mitä (sinä) haluta minä tehdä?	1107
Otan tuon kohteliaisuutena.	Ottaa tuo kohteliaisuus.	9786
Tuhannet kiitokset.	Tuhat kiitos.	9879
Mikä sinun nimesi olikaan?	Mikä sinä nimi olla?	9918
Onko se liikaa pyydetty?	Olla se liikaa pyytää?	9939

Helppojen lauseiden korjaaminen alkuperäiseen muotoon ei suurimmilta osin pitäisi olla mahdotonta. Kielimallit onnistuivat siten kuvion 10 mukaisesti saavuttamaan hyvän arvosanan taivutuskorjaamisessa. Kuviossa näkyvät tulokset on esitetty prosenttien sijaan pistearvoina. Nämä arvot perustuvat lauseiden korjausten arvioinnissa käytettyyn pisteytykseen, joka esiteltiin luvussa 4.3. Neljän pisteen eli täydellisen korjauksen saavuttaminen voi olla kielimalleille haastava tehtävä esimerkiksi aikamuotojen takia. Kolme pistettä on siten myös tyydyttävä suoritus, sillä kyseisen arvosanan saanut lause on hyvin lähellä alkuperäistä lausetta. Pääosin kaikki paitsi Llama ja DeepSeek saivat keskiarvoksi vähintään arvosanan 3. Vähintään 3,5 pisteen arvosanalla voisi mallin olettaa antavan usein hyviä vastauksia. ChatGPT-mallit ja Gemini 2.5 Flash onnistuivat usein hyvien tulosten saamisessa. Esimerkit eivät tässä näyttäneet auttavan paljoa. Voi olla, että vain yhden yksinkertaisen lauseen korjaaminen ei selventänyt aihetta mallille tarpeeksi.



	ChatGPT-4o	ChatGPT-o4-mini (päättele-toiminto)	Gemini 2.5 Flash (ei "Thinking")	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	3,80	3,70	3,70	3,30	1,80	3,50	1,50
Suomi, esim.	3,60	3,70	3,70	3,10	1,60	3,10	1,70
Englanti	3,60	3,70	3,50	3,40	2,80	3,00	2,30
Englanti, esim.	3,40	3,70	3,50	3,00	2,60	3,20	2,10

Kuvio 10. Helppojen lauseiden taivutuskorjaamisen tulokset

6.2 Vaikeampien lauseiden korjaaminen

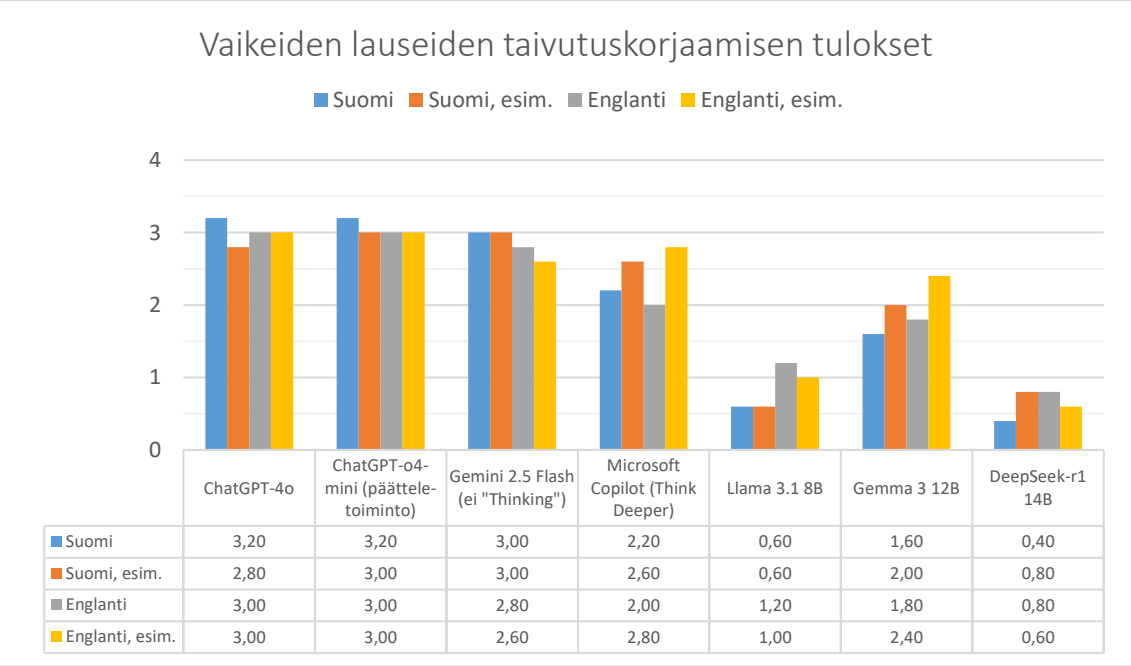
Taulukossa 11 näkyy tässä alaluvussa testatut vaikeammat lauseet, niiden alkuperäismuodot sekä niille annettu numero Suomen puupankki-korpuksessa (Helsingin yliopisto, 2023). Taulukosta huomaa, että perusmuotolauseet sisältävät taulukon 10 helpompiin lauseisiin verrattuna enemmän korjattavaa. Taivutusvirheiden tulkinta voi niiden määrän takia olla kielimalleille haastavaa.

Taulukko 11. Taivutuskorjaamisen vaikeammat lauseet

Alkuperäinen lause	Perusmuotolause	Numero
Se ketä ei tästä pidä ei ymmärrä hyvästä elokuvasta yhtään mitään	se kuka ei tämä pitää ei ymmärtää hyvä elokuva yhtään mikään	546
Koodista tulisi loputtoman pitkä ja etenkin ulkopuolisen lienisi hankala lukea sitä	koodi tulla loputon pitkä ja etenkin ulkopuolinen olla hankala lukea se	660
Turhien tutkimusten ja hoitojen karsiminen säästää työtä, aikaa ja rahaakin.	turha tutkimus ja hoito karsiminen säästää työ, aika ja raha.	883
Tänä vuonna ei rannan pohjoisosaa ole kuitenkaan siivottu, vaikka ei se mikään iso työ olisi.	tämä vuosi ei ranta pohjoisosa olla kuitenkaan siivota, vaikka ei se mikään iso työ olla.	933
Kaikesta epätodellisuudestaan ja keinotekoisuudestaan huolimatta "Kauniiden ja rohkeiden" maailma on aikuisten maailma.	kaikki epätodellisuus ja keinotekoisuus huolimatta "kaunis ja rohkea" maailma olla aikuinen maailma.	1016
Se alkaa retkipaikan ja teeman sopivan valmistusmenetelmän valinnasta, kypsennysjärjestyksen miettimisestä ja esivalmisteluista ja päättyy tähderuokien keräykseen ja pakkaamiseen, tulisijan sammutukseen, jätehuoltoon ja kierrätykseen.	se alkaa retkipaikka ja teema sopia valmistusmenetelmä valinta, kypsennysjärjestys miettiminen ja esivalmistelu ja päättyä tähderuoka keräys ja pakkaaminen, tulisija sammutus, jätehuolto ja kierrätys.	1120

Vaikeampien lauseiden tulokset näkyvät kuviossa 11. Helppojen lauseiden tulosten tavoin kuviossa näkyvät tulokset on prosenttien sijasta mitattu korjausten arvioinnissa käytetyillä pisteillä. Neljä pistettä tarvitsee täydellisyyttä, joka on vaikeampien lauseiden kanssa hyvin haastavaa. Siten vaikeiden lauseiden arvioinnissa kolme pistettä on hyvä suoritus, sillä sen saanut lause on silti hyvin lähellä alkuperäistä. Suurista kielimalleista ChatGPT-mallit olivat onnistuneet useimmiten eri kehotteilla saavuttamaan kolmen pisteen keskiarvon. Gemini 2.5 Flash onnistui tässä vain suomeksi, ja Microsoft Copilot ei onnistunut lainkaan. Pienistä kielimalleista vain Gemma 3 oli jokseenkin onnistunut englanniksi esimerkkivastauksen kanssa korjaamaan taivutuksia. Llama 3.1 ja DeepSeek-r1 eivät onnistuneet vaikeampien sanojen kanssa korjaamaan taivutuksia perusmuodoista. Koska vaikeimpia lauseita ei helposti pysty korjaamaan täysin alkuperäiseen muotoonsa,

vähintään 2,5 pisteen arvosana voidaan todeta hyväksi tulokseksi. Vaikeampien lauseiden kanssa suurista kielimalleista vain Gemini 2.5 Flash sai huonommat tulokset englanninkielisellä kehotteella. Esimerkkivastaus auttoi suurista kielimalleista vain Copilot-mallia. Muuten kehotteen kieli tai esimerkkivastaus ei niinkään vaikuttanut suurien kielimallien vastausten laatuun. Pienistä kielimalleista Llama 3.1 ja Gemma 3 hyötyivät englanninkielisestä kehotteesta, mutta vain Gemma 3 hyötyi esimerkkivastauksista.



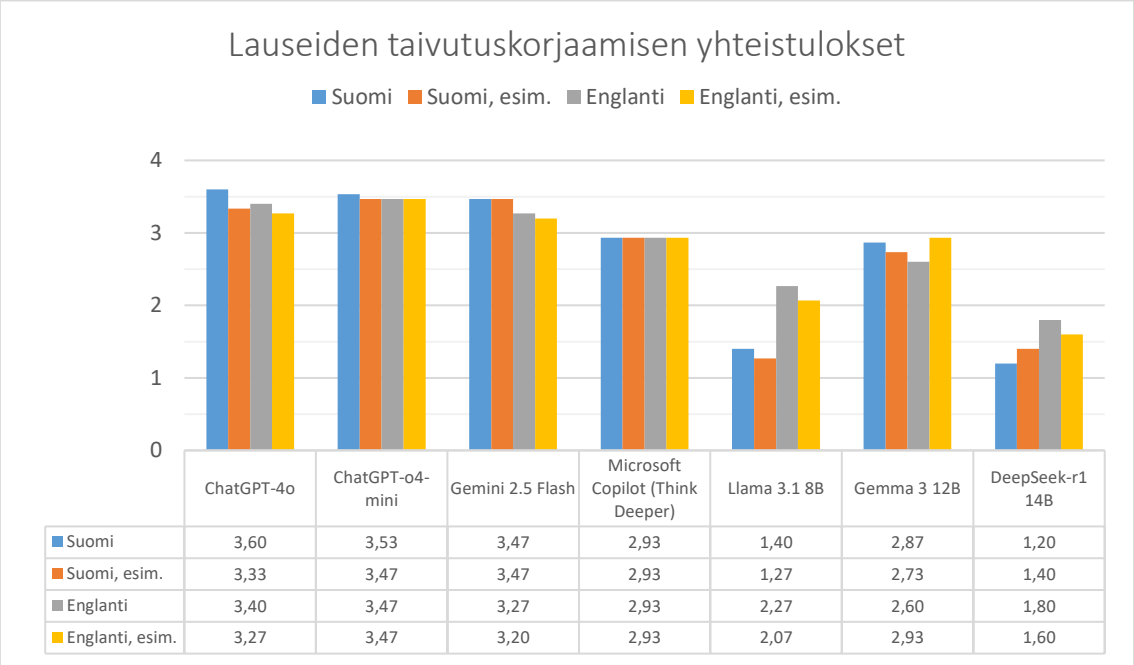
	ChatGPT-4o	ChatGPT-o4-mini (päättele-toiminto)	Gemini 2.5 Flash (ei "Thinking")	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	3,20	3,20	3,00	2,20	0,60	1,60	0,40
Suomi, esim.	2,80	3,00	3,00	2,60	0,60	2,00	0,80
Englanti	3,00	3,00	2,80	2,00	1,20	1,80	0,80
Englanti, esim.	3,00	3,00	2,60	2,80	1,00	2,40	0,60

Kuvio 11. Vaikeiden lauseiden taivutuskorjaamisen tulokset

6.3 Tulosten yhteenveto

Kuviossa 12 on kaikkien lauseiden taivutuskorjaamisten keskiarvot jokaiselle kielimallille ja kehotteelle. Koska helpompia lauseita on enemmän kuin vaikeita lauseita, niin yli 3 pisteen arvosana on hyväksyttävä tulos. Vain ChatGPT-mallit sekä Gemini 2.5 onnistuivat ylittämään tämän pistekynnyksen. ChatGPT-malleista pienempi o4-mini-malli tuotti isompaan 4o-malliin verrattuna vähän paremmat tulokset. Syväajattelu mahdollisesti vaikutti tuloksiin positiivisesti. Tulokset olivat kuitenkin lähellä toisiaan, joten se voi myös

yksinkertaisesti johtua satunnaisuudesta. Microsoft Copilot ja Gemma 3 ovat hyvin lähellä hyväksyttävää tulosta. Llama 3.1 ja DeepSeek-r1 eivät saaneet korjattua taivutuksia tyydyttävästi. Suuret kielimallit olivat tässä tehtävässä yllättäen saaneet paremmin tuloksia suomeksi, mutta tämä pieni eroavaisuus voi silti johtua satunnaisuudesta. Pienet kielimallit toimivat tässäkin parhaiten englanniksi, joka voi johtua niiden suomen kielen ymmärtämistä rajoittavasta pienestä koosta. Esimerkkivastauksella ei tässä ollut paljoa apua varsinkaan suurille kielimalleille. Gemma 3 -malli näytti kuitenkin hyötyneen englanniksi esimerkkivastauksesta.



	ChatGPT-4o	ChatGPT-o4-mini	Gemini 2.5 Flash	Microsoft Copilot (Think Deeper)	Llama 3.1 8B	Gemma 3 12B	DeepSeek-r1 14B
Suomi	3,60	3,53	3,47	2,93	1,40	2,87	1,20
Suomi, esim.	3,33	3,47	3,47	2,93	1,27	2,73	1,40
Englanti	3,40	3,47	3,27	2,93	2,27	2,60	1,80
Englanti, esim.	3,27	3,47	3,20	2,93	2,07	2,93	1,60

Kuvio 12. Lauseiden taivutuskorjaamisen yhteistulokset

6.4 Kielimallien lauseiden taivutuskorjaamisessa tekemät virheet

Tämän luvun tarkoitus on selventää, miten kielimallit saivat tässä pääluvussa esitetyt lauseiden taivutusten korjaamistulokset. Joitain suurten sekä pienten kielimallien antamia virheellisiä korjauksia käydään läpi syvemmin. Tarkoituksena on saada selville eri kielimallien tekemiä virheitä liittyen taivutusten korjaamiseen.

Suurin osa ajasta helppojen lauseiden kanssa Suurten kielimallien virheet johtuvat kontekstin puutteesta. Esimerkiksi virheellisen lauseen ”Mikä sinä nimi olla” korjaaminen alkuperäiseen muotoon ”Mikä sinun nimesi olikaan” on hyvin epätodennäköistä. Suurimpana ongelmana on viimeinen sana ”olikaan”. Infinitiivimuodosta ei pysty mitenkään selvittää edes, että sana oli alun perin imperfektimuodossa. Jokainen verkkopohjaisista malleista siten antoivat todennäköisimmän korjauksen, eli preesensmuodon. Tämän takia täysien pisteiden saaminen ei ole aina todennäköistä, ja vähän virheellinenkin vastaus on silti hyvä tulos. Jos esimerkiksi käyttäjä kuitenkin tietäisi, missä muodoissa virheellisen lauseen sanat pitäisivät olla, niin hän voi antaa muodot lisäkontekstina kielimallille. Kielimallien pitäisi pystyä korjaamaan taivutukset tarkemmin, kun se tietää lauseen alkuperäisen kontekstin sekä mihin muotoihin sanat pitäisi korjata. Tässä tutkimuksessa kielimallille ei annettu lauseiden kanssa lisäohjeistusta. Testaamisessa oli oletettu, että käyttäjä ei itse tietäisi lauseen alkuperäistä kontekstia tai sanojen alkuperäisiä muotoja.

Suurista kielimalleista Gemini 2.5 Flash ja Copilot Think Deeper olivat antaneet korjauksina helpommille lauseille välillä vähemmän selkeitä lauseita. Esimerkiksi Gemini 2.5 Flash oli englanninkielisellä kehotteella korjannut lauseen ”Mitä sinä haluta minä tehdä” muotoon ”Mitä sinä haluat minua tekemään”. Lauseen loppuosa ei kuulosta luontevalta, eikä se myöskään täsmää alkuperäistä lausetta, joka on ”Mitä sinä haluat minun tekevän”. Copilot oli puolestaan suomenkielisillä kehotteilla korjannut lauseen ”tuhat kiitos” muotoon ”tuhatta kiitosta”. Tämä ei myöskään kuulosta luontevalta. Alkuperäinen ”tuhannet kiitokset” tai ”tuhat kiitosta”, jonka osa kielimalleista olivat joskus antaneet, olisivat tätä selkeämpiä. Kummankin mallin vastauksista kuitenkin pystyy ymmärtämään alkuperäisten lauseiden tarkoitukset. Ne siten saivat silti niistä itselleen kaksi pistettä.

Vaikeimpien lauseiden kanssa kaikilla kielimalleilla oli ongelmia korjata lauseita täysin alkuperäisiksi. Tämä on oletettavissa, sillä mahdollisten taivutuksien permutaatioita on pidemmissä lauseissa paljon enemmän kuin lyhyemmissä. Johtuen lauseiden pituuksista

jotkin kielimallit eivät onnistu kaikkia lauseiden taivutuksia korjaamaan. Esimerkiksi Gemini 2.5 Flash -kielimallilla oli englanninkielisillä kehotteilla vähän vaikeuksia alkuperäisen virkkeen ”Tänä vuonna ei rannan pohjoisosaa ole kuitenkaan siivottu, vaikka ei se mikään iso työ olisi” korjaamisessa. Virheellinen virke, joka kielimalleille annettiin, oli: ”Tämä vuosi ei ranta pohjoisosa olla kuitenkaan siivota, vaikka ei se mikään iso työ olla.” Gemini 2.5 Flash -kielimalli antoi vastauksena englanniksi ilman esimerkkivastausta lauseen: ”Tämä vuosi ei ole rannan pohjoisosaa kuitenkaan siivottu, vaikka ei se mikään iso työ ole.” Gemini 2.5 Flash oli unohtanut korjata päälauseen alkuosassa sanat ”tämä vuosi”, jolloin siltä jäi vain kaksi sanaa korjaamatta. Koska kielimalli ei korjannut koko virkettä kieliopillisesti oikeaan muotoon, niin sille annettiin tästä kaksi pistettä.

Varsinkin pienille kielimalleille vaikeampien lauseiden korjaaminen oli haastavaa. Testatuista pienistä kielimalleista vain Gemma 3 -kielimalli oli onnistunut testauksen aikana olemaan antamatta täysin ymmärtämättömiä lauseita. Annetaan esimerkkinä taulukossa 12 pienten kielimallien vastaukset taivuttamattomalle ja virheelliselle lauseelle: ”Turha tutkimus ja hoito karsiminen säästää työ, aika ja raha.” Taulukossa vastauksille annetut arvioinnit ovat merkitty suluissa lauseiden loppuihin. Alkuperäinen lause, johonka kielimallien pitäisi tämä virheellinen lause saada korjattua, on: ”Turhien tutkimusten ja hoitojen karsiminen säästää työtä, aikaa ja rahaakin.” Gemma 3 oli onnistunut taulukon 12 mukaisesti parhaimmillaan englanninkielisen esimerkkivastauksen sisältävän kehotteen kanssa saamaan lauseen korjattua lähes alkuperäiseen versioon. Llama 3.1-malli oli saman kehotteen kanssa antanut kaksi mahdollista korjausta. Ensimmäisessä korjauksessa se oli jostain syystä yhdistänyt sanat ”karsiminen” ja ”säästää” muotoon ”vähennettäisiin”, jolloin lauseen konteksti muuttui. Llaman antama toinen versio leikkasi lauseen loppuosan pois. DeepSeek-r1-mallin englanninkielinen esimerkkikehote piti jokseenkin alkuperäisen kontekstin, mutta lauseessa on useita muita virheitä. Sana ”turha” jätettiin pois, ”säästää” kirjoitettiin ”säätää” sekä työ ja aika ovat nyt yhdyssana. Gemma oli siten ainut paikallisista kielimalleista, joka onnistui tyydyttävästi lauseen korjaamisessa vähintään yhdellä kehotepohjalla.

Taulukko 12. Pienten kielimallien taivutuskorjausesimerkki

	Llama 3.1 8B	DeepSeek-r1 14B	Gemma 3 12B
Suomenkielinen kehote	Turha tutkimus ja hoito karsimista säästää työ, aika ja raha (1)	Turha tutkimus ja hoito karsiminen säilyttää työ, aika ja raha. (1)	Turha tutkimus ja hoitojen karsiminen säästää työtä, aikaa ja rahaa. (2)
Suomenkielinen kehote esimerkkivastauksella	Turha tutkimus ja hoito karsii työtä, aikaa ja rahaa. (1)	Turha tutkimus ja hoitokarsi säästään työ, aika ja raha. (0)	Turha tutkimus ja hoito karsitaan säästää työtä, aikaa ja rahaa. / Turha tutkimuksen ja hoidon karsiminen säästää työtä, aikaa ja rahaa. (2)
Englanninkielinen kehote	Turha tutkimus ja hoito karsiminen säästää työn, ajan ja rahan. / Täysin turhankin tutkimus ja hoito vähyttämisen säästää työaikaan ja varallisuuteen. (1)	Turhan tutkimuksen ja hoidon karsiminen säilyttää työtä, aikaa ja rahaa. (2)	Turhat tutkimukset ja hoidot säästävät työ, aika ja raha. / Turha tutkimus ja hoito karsiminen säästää työ, aika ja raha. (2)
Englanninkielinen kehote esimerkkivastauksella	Turha tutkimus ja hoito vähennettäisiin työ, aika ja raha. / Turhaa olisi tehdä tutkimusta ja hoitoa (1)	Tutkinto ja hoito karsiminen säätää työaika ja rahaa. (1)	Turhan tutkimuksen ja hoidon karsiminen säästää työtä, aikaa ja rahaa. (3)

7 Johtopäätökset ja pohdintaa

Tutkimus kävi läpi kielimallien monikielisyyttä keskittyen niiden suomen kieleen morfologian osaamiseen. Suurten kielimallien on aikaisempien tutkimuksien mukaan todettu ymmärtävän englannin kieltä hyvin. Tämän tutkimuksen tuloksista voi kuitenkin todeta niiden myös osaavan muitakin kieliä. Suomen kielellä kielimallit onnistuivat osittain taivuttamaan sanoja aloittelijan tasolla. Suuret kielimallit tuottivat vähiten virheitä, mutta pienetkin kielimallit olivat jotenkin onnistuneet taivuttamisessa. Virheitä kuitenkin sattui välillä pienten ja suurten kielimallien taivutuksissa jopa helppojenkin taivutuksien kanssa, joten niiden vastauksissa löytyy vielä parannettavaa.

Kielimalleilla oli testattu sanojen eri taivutusmuotojen tuottamista sekä kääntämistä englanniksi. Lisäksi oli myös testattu taivutetuille sanoille perusmuodon antamista, johon aikaisemmissa tutkimuksissa Bacherikov (2024, s. 13) oli käyttänyt erillistä Voikko-kirjastoa. Tämän työn tulokset osoittavat, että kielimallitkin pystyvät myös tähän tehtävään. Lopuksi vielä testattiin kielimalleilla lauseiden arvioimista taivutuksien näkökulmasta. Luvussa 3 esitetyistä kielimallien käyttötavoista oli siis testattu generoimista, kääntämistä sekä arvioimista. Keskusteleva käyttö jäi testaamatta, sillä se keskittyisi taivuttamisen sijasta suomen kielen laajaan osaamiseen. Jatkossa voisi kielimallien suomen kielen laajemman osaamisen kanssa testata myös keskustelevaa käyttöä.

Sanojen taivuttaminen onnistui kielimalleilta perusmuodoista yleisimpiin taivutusmuotoihin sekä taivutusmuodoista perusmuotoihin. Suuret kielimallit pystyivät myös välillä käsittelemään vähän harvinaisempiakin taivutusmuotoja, kuten verbien kanssa potentiaalimuotoa. Pienillä kielimalleilla oli kuitenkin vaikeuksia näiden harvinaisempien taivutusmuotojen kanssa. Substantiivien ja adjektiivien taivuttaminen perusmuodoista onnistui varsinkin suurilta kielimalleilta ilman mitään ongelmia. Niiden sijamuodoista testattiin kuitenkin vain yleisimpiä muotoja, joten kielimalleilla voi silti olla vaikeuksia harvinaisempien sijamuotojen kanssa. Pienet kielimallit olivat myös osanneet helpot taivutukset ilman suuria virheitä. Vähemmän monikielisyyteen keskittyvät pienet kielimallit eivät kuitenkaan onnistuneet taivuttamaan tarpeeksi laadukkaasti.

Lauseiden taivutusten korjaaminenkin onnistui kielimalleilta varsinkin helpompien lauseiden kanssa. Vaikeampien lauseiden kanssa kielimallit eivät onnistuneet täysin korjaamaan kaikkia virheitä. Suuret kielimallit olivat kuitenkin antaneet parhaimmassa tapauksessa hyväksyttäviä korjauksia lauseiden taivutuksille. Pienillä kielimalleilla oli varsinkin vaikeampien lauseiden taivutusten korjaamisessa eniten vaikeuksia. Syynä voi mahdollisesti olla niiden koosta johtuva tiedon puute. Pienet kielimallit eivät mahdollisesti osanneet suomen kieltä tarpeeksi hyvin lauseiden taivutusten korjaamista varten.

Kummassakin taivutustehtävässä kielimalleista parhaiten onnistuivat ChatGPT- ja Gemini-mallit. Syväajattelua käyttävät ChatGPT-o4-mini ja Gemini 2.5 Flash onnistuivat antamaan yhtä hyviä taivutuksia niitä suuremman ChatGPT-4o-mallin kanssa. Syväajattelu näytti siten auttaneen näitä pienempiä malleja suomen kielen taivuttamisessa. Copilot-malli oli kuitenkin suurista malleista heikoin, vaikka se on hyvin suuri ja myös syväajatteleva malli. Mahdollisesti tietojen liiallinen määrä saattoi heikentää sen ajattelukykyä, tai sitä ei ehkä koulutettu tai ohjeistettu monikielisyyteen. Suurten kielimallien vastauksiin voi suurimmilta osin luottaa helpompien sanojen ja lauseiden kanssa. Vaikeampien kanssa voi kuitenkin olla vaikeuksia.

Pienistä kielimalleista Gemma 3 oli onnistunut muihin testattuihin pieniin kielimalleihin verrattuna parhaiten taivuttamisessa. Monikielisyyteen painottuminen näytti auttaneen mallia saavuttamaan taivutustehtävissä lähes yhtä hyvät tulokset verrattuna suuriin kielimalleihin. Gemma 3 -mallin tuloksista voidaan todeta, että samankaltaiset monikieliset pienet kielimallit pystyisivät myös suurten mallien tavoin opettaa suomen kieltä ulkomaalaisille. Suomen kielen taivuttamisessa Llama 3.1 sekä DeepSeek-r1 eivät kumpikaan onnistuneet yhtä hyvin. Pienistä kielimalleista siten vain monikielisyyteen keskittyvät mallit voivat onnistua opettamaan suomen kieltä. Näiden mallien uudemmat tai suuremmat versiot saattaisivat kuitenkin ymmärtää suomen kieltä vielä paremmin. Suurempia malliversioita ei pystytty tässä tutkimuksessa testaamaan, sillä ne olisivat mahdollisesti tarvinneet kielimalleihin erikoistuvia laitteistoja.

Kehotteen kielellä sekä esimerkkivastauksen antamisella näytti myös tulosten mukaan olevan joitain vaikutuksia kielimallien vastauksiin. Esimerkkivastaukset olivat ajoittain vaikuttaneet kielimallien vastauksiin myönteisesti, mutta kielimallit eivät aina hyötyneet niistä. Kielellä ei ollut paljoa vaikutusta suurten kielimallien kanssa, mutta pienet kielimallit hyötyivät useimmiten englanninkielisestä kehotteesta. Lauseiden taivutusten korjaamisessa tuli sen sijaan parhaimmat tulokset englanniksi. Kielimallit eivät lauseiden kanssa hyötyneet lainkaan esimerkkivastauksesta. Laadukkaampi tai useampi esimerkki voisi mahdollisesti auttaa paremmin vastauksen laadun kanssa.

Kielimallit osaavat tämän työn tulosten mukaisesti käsitellä suomen kieltä. Tuloksista voi todeta, että kielimalleilla voisivat mahdollisesti pystyä opettamaan suomen kielen kaltaisia vähemmän yleisempiäkin kieliä. Kielimallien taivutukset eivät kuitenkaan ole täydellisiä, joten ne eivät korvaa ihmisopettajien antamaa kunnon koulutusta. Tällöin kielimallien käyttöä voisi suositella esimerkiksi suomen kielen aloittelijoille, jotka eivät ole vielä täysin varmoja suomen kielen oppimisesta. Kokeneempien suomen kielen oppijoiden kannattaa etsiä muita laadukkaampia lähteitä, kuten esimerkiksi suomen kielen kursseja tai oppikirjoja.

Jatkotutkimusaiheina voisi laajentaa tätä tutkimusta eri sanojen ja erilaisempien kehotteiden kanssa. Esimerkiksi kielimalleilla voisi testata taivuttamista laajemmalla sanavalikoimalla. Kehotteiden laadun vaikutusta kielimallien vastauksiin kannattaisi myös tutkia syvemmin. Myös voisi testata kielimallien käyttöä käytännössä suomen kielen aloittelijoiden opettamiseen esimerkiksi keskustelevalla käytöllä, jota ei tässä tutkimuksessa testattu. Myös olisi hyvä tutkia kielimallien suomen kielen osaamista muiden kielitieteen osa-alueiden, kuten syntaksin eli lauseopin tai oikeinkirjoituksen kannalta. ChatGPT-5 oli julkaistu tämän työn kirjoittamisen aikana tulosten keräämisen jälkeen, joten sitä ei ehditty tässä työssä testata. Jatkotutkimuksissa olisi mahdollisesti hyvä testata, olisiko ChatGPT-5 sen aikaisempiin versioihin verrattuna tehokkaampi.

Lähteet

Ali, M., Fromm, M., Thellmann, K., Rutmann, R., Lübbering, M., Leveling, J., Klug, K., Ebert, J., Doll, N., Buschhoff, J. S., Jain, C., Weber, A. A., Jurkschat, L., Abdelwahab, H., John, C., Suarez, P. O., Ostendorff, M., Weinbach, S., Sifa, R., … Flores-Herr, N. (2024). *Tokenizer Choice For LLM Training: Negligible or Crucial?* (No. arXiv:2310.08754). arXiv. https://doi.org/10.48550/arXiv.2310.08754

Alomar, K., Aysel, H. I., & Cai, X. (2024). *RNNs, CNNs and Transformers in Human Action Recognition: A Survey and a Hybrid Model* (No. arXiv:2407.06162). arXiv. https://doi.org/10.48550/arXiv.2407.06162

Alsafari, B., Atwell, E., Walker, A., & Callaghan, M. (2024). Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants. *Natural Language Processing Journal*, *8*, 100101. https://doi.org/10.1016/j.nlp.2024.100101

An, S., Ma, Z., Lin, Z., Zheng, N., Lou, J. G., & Chen, W. (2023). *Learning from mistakes makes llm better reasoner.* (No. arXiv: 2310.20689). arXiv. https://doi.org/10.48550/arXiv.2310.20689

Bacherikov, D. (2024). *Development of web application for practicing Finnish language writing skills with the help of LLMs* [AMK-opinnäytetyö]. https://urn.fi/URN:NBN:fi:amk-2024051411629

Barman, D., Guo, Z., & Conlan, O. (2024). The dark side of language models: Exploring the potential of llms in multimedia disinformation generation and dissemination. *Machine Learning with Applications*. https://doi.org/10.1016/j.mlwa.2024.100545

Baytak, A. (2024). The Content Analysis of the Lesson Plans Created by ChatGPT and Google Gemini. *Research in Social Sciences and Technology*, *9*(1), 329–350. Noudettu 24.1.2025 osoitteesta https://eric.ed.gov/?id=EJ1426976

Briganti, G. (2024). How ChatGPT works: A mini review. *European Archives of Oto-Rhino-Laryngology*, *281*(3), 1565–1569. https://doi.org/10.1007/s00405-023-08337-7

Chang, L.-H., & Ginter, F. (2024). Automatic Short Answer Grading for Finnish with ChatGPT. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), Article 21. https://doi.org/10.1609/aaai.v38i21.30363

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2023). *Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review*. (No. arXiv: 2310.14735). arXiv. https://doi.org/10.48550/arXiv.2310.14735

Cong, Y. (2024). Demystifying large language models in second language development research. *Computer Speech & Language*, *89*, 101700. https://doi.org/10.1016/j.csl.2024.101700

Creutz, M. (2024). Correcting Challenging Finnish Learner Texts With Claude, GPT-3.5 and GPT-4 Large Language Models. Teoksessa R. van der Goot, J. Bak, M. Müller-Eberstein, W. Xu, A. Ritter, & T. Baldwin (Eds.), Proceedings of the Ninth Workshop on Noisy and User-generated Text (W-NUT 2024): Collocated with EACL 2024 (pp. 1-10). Association for Computational Linguistics (ACL). Noudettu 26.2.2025 osoitteesta https://aclanthology.org/2024.wnut-1.1/

Crystal, D. (2008). Two thousand million? *English Today*, *24*(1), 3–6. https://doi.org/10.1017/S0266078408000023

CSC – IT Center for Science. (2004). *Frequency Lexicon of the Finnish Newspaper Language* [Data set]. Kielipankki. http://urn.fi/urn:nbn:fi:lb-201405271

Dahl, Ö. (2008). Kuinka eksoottinen kieli suomi on? *Virittäjä*, *112*(4), 545–559. Noudettu 18.9.2025 osoitteesta https://journal.fi/virittaja/article/view/40707

Ding, Y., & Shi, T. (2024). Sustainable LLM Serving: Environmental Implications, Challenges, and Opportunities : Invited Paper. *2024 IEEE 15th International Green and Sustainable Computing Conference (IGSC)*, 37–38. https://doi.org/10.1109/IGSC64514.2024.00016

DeepSeek-AI (2025). *DeepSeek-R1: Incentivizing reasoning capability in llms via reinforcement learning* (No. arXiv: 2501.12948). arXiv. https://doi.org/10.48550/arXiv.2501.12948

Gemini Team, Google (2024). *Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context* (No. arXiv:2403.05530). arXiv. https://doi.org/10.48550/arXiv.2403.05530

Gemini Team, Google (2025). *Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities* (No. arXiv:2507.06261). arXiv. https://doi.org/10.48550/arXiv.2507.06261

Gemma Team, Google DeepMind (2025). *Gemma 3 Technical Report* (No. arXiv:2503.19786). arXiv. https://doi.org/10.48550/arXiv.2503.19786

George, A. S., George, A. S. H., & Martin, A. S. G. (2023). The Environmental Impact of AI: A Case Study of Water Consumption by Chat GPT. *Partners Universal International Innovation Journal*, *1*(2), Article 2. https://doi.org/10.5281/zenodo.7855594

Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. *Annals of biomedical engineering*, *51*(12), 2629-2633. https://doi.org/10.1007/s10439-023-03272-4

Greco, C. M., & Tagarelli, A. (2023). Bringing order into the realm of Transformer-based language models for artificial intelligence and law. *Artificial Intelligence and Law*, 1-148. https://doi.org/10.1007/s10506-023-09374-7

Hauhio, I., & Friberg, T. (2024). Mitra: Improving Terminologically Constrained Translation Quality with Backtranslations and Flag Diacritics. Teoksessa *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)* (ss. 100–115). European Association for Machine Translation (EAMT). https://aclanthology.org/2024.eamt-1.12/

Haukka, K. (2024). *Kuinka suuret kielimallit oppivat ymmärtämään ja tuottamaan kieltä?* http://urn.fi/URN:NBN:fi:jyu-202402061757

Heikkilä, J. (2024, 10. kesäkuuta) *Tutkimus ChatGPT: llä luotujen käännösten laadusta*. Turun yliopisto. https://urn.fi/URN:NBN:fi-fe2024061452800

Heikkola, L. M., Toivola, M., Kim, J., & Saloranta, A. (2024). Kielivalinnat ja kielenvaihtaminen ensikielisten ja ei-ensikielisten suomenpuhujien keskusteluissa – Kyselytutkimus: Language choices and language switching in conversations between L1

and L2 Finnish speakers – A survey. *AFinLA-teema*, *17*, Article 17. https://doi.org/10.30660/afinla.132056

Helsingin yliopisto (2023). *Suomen puupankki FinnTreeBank 1:n ladattava versio* [aineisto]. Kielipankki. http://urn.fi/urn:nbn:fi:lb-2022090201

Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., & Gu, X. (2024). Teaching Plan Generation and Evaluation With GPT-4: Unleashing the Potential of LLM in Instructional Design. *IEEE Transactions on Learning Technologies*, *17*, 1445–1459. IEEE Transactions on Learning Technologies. https://doi.org/10.1109/TLT.2024.3384765

Huang, J. (2023). Engineering ChatGPT prompts for EFL writing classes. *International Journal of TESOL Studies*, *5*(4), 73-79. https://doi.org/10.58304/ijts.20230405

Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: A review of emerging educational technology. *Smart Learning Environments*, *11*(1), 22. https://doi.org/10.1186/s40561-024-00310-z

Jadhav, A. (2025, 1. tammikuuta). *LLM Transformer Overview*. Noudettu 19.9.2025 osoitteesta https://www.anup.io/p/llm-transformer-overview

Jiao, J., Afroogh, S., Xu, Y., & Phillips, C. (2024). *Navigating llm ethics: Advancements, challenges, and future directions.* (No. arXiv: 2406.18841). arXiv. https://doi.org/10.48550/arXiv.2406.18841

Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). *Why Language Models Hallucinate* (No. arXiv:2509.04664). arXiv. https://doi.org/10.48550/arXiv.2509.04664

Katz, J. M. (2024, 16. tammikuuta). *Understanding Large Language Models—Parameters*. Noudettu 12.9.2025 osoitteesta https://kelvin.legal/understanding-large-language-models-what-are-paramters

Kettunen, K. (2005). *Sijamuodot haussa – tarvitseeko kaikkea hakutermien morfologista vaihtelua kattaa?* Noudettu 7.3.2025 osoitteesta https://www.researchgate.net/publication/240240603_Sijamuodot_haussa-tarvitseeko_kaikkea_hakutermien_morfologista_vaihtelua_kattaa_Ms

Kimppa, L., Shtyrov, Y., Hut, S. C. A., Hedlund, L., Leminen, M., & Leminen, A. (2019). Acquisition of L2 morphology by adult language learners. *Cortex*, *116*, 74–90. https://doi.org/10.1016/j.cortex.2019.01.012

Kotimaisten kielten keskus (2024, 15. joulukuuta). *Suomen kieli.* Noudettu 21.8.2025 osoitteesta https://kotus.fi/kotus/kielet-ja-kielipolitiikka/kansalliskielet/suomen-kieli/

Kielitoimiston sanakirja. (2024, 19. maaliskuuta). Kotimaisten kielten keskus. http://urn.fi/urn:nbn:fi:kotus-201433.

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology*, *3*(1). Noudettu 20.11.2024 osoitteesta https://langedutech.com/letjournal/index.php/let/article/view/48

Kytö, M. (2024). *Copilot for Microsoft 365: A Comprehensive End-user Training Plan for Organizations*. https://urn.fi/URN:NBN:fi:amk-202404257869

Laato, S., Morschheuser, B., Hamari, J., & Björne, J. (2023, July). AI-assisted learning with ChatGPT and large language models: Implications for higher education. *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)* (pp. 226-230). IEEE. https://doi.org/10.1109/ICALT58122.2023.00072

Lesonen, S., Steinkrauss, R., Suni, M., & Verspoor, M. (2020). Lexically specific vs. Productive constructions in L2 Finnish. *Language and Cognition*, *12*(3), 526–563. https://doi.org/10.1017/langcog.2020.12

Llama Team, AI @ Meta. (2024). *The Llama 3 Herd of Models* (No. arXiv:2407.21783). arXiv. https://doi.org/10.48550/arXiv.2407.21783

Luukkonen, R., Komulainen, V., Luoma, J., Eskelinen, A., Kanerva, J., Kupari, H. M., Ginter, F., Laippala, V., Muennighoff, N., Piktus, A., Wang, T., Tazi, N., Scao, T. L., Wolf, T., Suominen, O., Sairanen, S., Merioksa, M., Heinonen, J., Vahtola, A., … & Pyysalo, S. (2023). *Fingpt: Large generative models for a small language.* arXiv preprint. https://doi.org/10.48550/arXiv.2311.05640

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom:

AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*. https://doi.org/10.1016/j.caeai.2023.100199

Microsoft Learn. (2025, 5. tammikuuta). *Decide which Copilot is right for you*. Noudettu 12.9.2025 osoitteesta https://learn.microsoft.com/en-us/copilot/microsoft-365/which-copilot-for-your-organization

Moisio, A., Creutz, M., & Kurimo, M. (2024). *LLMs' morphological analyses of complex FST-generated Finnish words* (No. arXiv:2407.08269). arXiv. https://doi.org/10.48550/arXiv.2407.08269

Nalpas, M. (2024, toukokuuta 30). *Understand LLM sizes*. Web.Dev. Noudettu 12.9.2025 osoitteesta https://web.dev/articles/llm-sizes

Nguyen, C. V., Shen, X., Aponte, R., Xia, Y., Basu, S., Hu, Z., Chen, J., Parmar, M., Kunapuli, S., Barrow, J., Wu, J., Singh, A., Wang, Y., Gu, J., Dernoncourt, F., Ahmed, N. K., Lipka, N., Zhang, R., Chen, X., … Nguyen, T. H. (2024). *A Survey of Small Language Models* (No. arXiv:2410.20011). arXiv. https://doi.org/10.48550/arXiv.2410.20011

Nikolova-Stoupak, I., Bibauw, S., Dumont, A., Stas, F., Watrin, P., & François, T. (2024). Generating Contexts for ESP Vocabulary Exercises with LLMs. *Swedish Language Technology Conference and NLP4CALL*, 153–175. https://doi.org/10.3384/ecp211012

OpenAI (2024a). Gpt-4o system card. arXiv preprint. https://doi.org/10.48550/arXiv.2410.21276

OpenAI (2024b) *GPT-4o mini: advancing cost-efficient intelligence.* Noudettu 14.1.2025 osoitteesta https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/

OpenAI (2025, 16. huhtikuuta). *Esittelyssä OpenAI o3 ja o4-mini.* Noudettu 11. syyskuuta 2025, osoitteesta https://openai.com/fi-FI/index/introducing-o3-and-o4-mini/

orgtre. (2025). *Top OpenSubtitles sentences*. Noudettu 15.9.2025 osoitteesta https://github.com/orgtre/top-open-subtitles-sentences

Pan, M., Kitson, A., Wan, H., & Prpa, M. (2024). *ELLMA-T: An Embodied LLM-agent for Supporting English Language Learning in Social VR* (No. arXiv:2410.02406). arXiv. https://doi.org/10.48550/arXiv.2410.02406

Park, J., Bae, J., Lee, U., Ahn, T., Lee, S., Kim, D., Choi, A., Jeong, Y., & Kim, H. (2024). *How to Align Large Language Models for Teaching English? Designing and Developing LLM based-Chatbot for Teaching English Conversation in EFL, Findings and Limitations.* arXiv. https://doi.org/10.48550/arXiv.2409.04987

Qin, L., Chen, Q., Zhou, Y., Chen, Z., Li, Y., Liao, L., Li, M., Che, W., & Yu, P. S. (2025). A survey of multilingual large language models. *Patterns*, *6*(1). https://doi.org/10.1016/j.patter.2024.101118

Rajaraman, N., Jiao, J., & Ramchandran, K. (2024). *Toward a Theory of Tokenization in LLMs* (No. arXiv:2404.08335). arXiv. https://doi.org/10.48550/arXiv.2404.08335

Salmela, R. (2024, 13. joulukuuta). *Finnish as a Second Language: Perspectives on Vocabulary Assessment, Morphological Processing, and Morphological Training*. Åbo Akademi - Åbo Akademi University. https://urn.fi/URN:ISBN:978-952-12-4445-2

Sanakirja.org (2011). *Sanakirja.* Noudettu 20.5.2025 osoitteesta https://www.sanakirja.org/

Shahriar, S., Lund, B. D., Mannuru, N. R., Arshad, M. A., Hayawi, K., Bevara, R. V. K., Mannuru, A. & Batool, L. (2024). Putting gpt-4o to the sword: A comprehensive evaluation of language, vision, speech, and multimodal proficiency. Applied Sciences, 14(17), 7782. https://doi.org/10.3390/app14177782

Srivastava, G., Cao, S., & Wang, X. (2025). Towards Reasoning Ability of Small Language Models (No. arXiv:2502.11569). arXiv. https://doi.org/10.48550/arXiv.2502.11569

Subramanian, S., Elango, V., & Gungor, M. (2025). *Small Language Models (SLMs) Can Still Pack a Punch: A survey* (No. arXiv:2501.05465). arXiv. https://doi.org/10.48550/arXiv.2501.05465

The Copilot Team (2025, 25. helmikuuta). *Announcing Free, Unlimited Access to Think Deeper and Voice.* Microsoft Copilot Blog. Noudettu 12.9.2025 osoitteesta

https://www.microsoft.com/en-us/microsoft-copilot/blog/2025/02/25/announcing-free-unlimited-access-to-think-deeper-and-voice/

Tieteen termipankki. (2022, 29. elokuuta). *Kielitiede:morfologia*. Noudettu 18.9.2025 osoitteesta https://tieteentermipankki.fi/wiki/Kielitiede:morfologia

Tilastokeskus. (2025, 18. kesäkuuta). *Väestö ja yhteiskunta*. Noudettu 21. elokuuta 2025 osoitteesta https://stat.fi/tup/suoluk/suoluk_vaesto.html#vakiluvun-kehitys

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models (No. arXiv:2302.13971). arXiv. https://doi.org/10.48550/arXiv.2302.13971

Tyen, G., Caines, A., & Buttery, P. (2024). LLM chatbots as a language practice tool: A user study. *Swedish Language Technology Conference and NLP4CALL*, 235–247. https://doi.org/10.3384/ecp211018

Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive Cross-Linguistic Differences: More Morphological Computation in English Than in Finnish. *Journal of Psycholinguistic Research*, *31*(2), 83–106. https://doi.org/10.1023/A:1014934915952

Ye, J., Wang, S., Zou, D., Yan, Y., Wang, K., Zheng, H.-T., Xu, Z., King, I., Yu, P. S., & Wen, Q. (2025). *Position: LLMs Can be Good Tutors in Foreign Language Education* (No. arXiv:2502.05467). arXiv. https://doi.org/10.48550/arXiv.2502.05467

Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2024). GPT (Generative Pre-Trained Transformer)—A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access*, *12*, 54608–54649. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3389497

Yuan, F., Yuan, S., Wu, Z., & Li, L. (2024). How Vocabulary Sharing Facilitates Multilingualism in LLaMA?. *Findings of the Association for Computational Linguistics ACL 2024* (pp. 12111-12130). https://doi.org/10.48550/arXiv.2311.09071

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica, 10(5), 1122-1136. https://doi.org/10.1109/JAS.2023.123618

Zhang, Q., Liu, Z., & Pan, S. (2025). The Rise of Small Language Models. *IEEE Intelligent Systems*, *40*(1), 30–37. https://doi.org/10.1109/MIS.2024.3517792

Liitteet

Liite 1. Perusmuodosta taivuttamisen kehotepohjat

Suomeksi:

-Verbi-

Haluaisin selvittää verbin "[sana]" joitain taivutusmuotoja. Halutut taivutusmuodot ovat: indikatiivin 1. persoonan preesens, indikatiivin 3. persoonan imperfekti, konditionaalin 3. persoonan preesens, potentiaalin 3. persoonan preesens, imperatiivin 3. persoonan preesens, 2. partisiippi ja passiivin imperfekti. Anna lisäksi infinitiivi- eli perusmuotoon enintään kolme englanninkielistä käännöstä, jos useita käännöksiä on olemassa.

Esimerkkivastaus verbin "sanoa" kanssa: indikatiivin 1. persoonan preesens = sanon; indikatiivin 3. persoonan imperfekti = sanoi; konditionaalin 3. persoonan preesens = sanoisi; potentiaalin 3. persoonan preesens = sanonee; imperatiivin 3. persoonan preesens = sanokoon; 2. partisiippi = sanonut; passiivin imperfekti = sanottiin; englanniksi = say, tell, speak.

-Substantiivi-

Haluaisin selvittää substantiiville "[sana]" joitain taivutuksia tietyille sijamuodoille yksikössä sekä monikossa. Halutut sijamuodot ovat: nominatiivi, genetiivi, partitiivi, illatiivi. Anna lisäksi nominatiivin yksikölle enintään kolme englanninkielistä käännöstä, jos useita käännöksiä on olemassa.

Esimerkkivastaus substantiivin "asia" kanssa: Yksikkö: Nominatiivi = asia; Genetiivi = asian; Partitiivi = asiaa; Illatiivi = asiaan; Monikko: Nominatiivi = asiat; Genetiivi = asioiden; Partitiivi = asioita; Illatiivi = asioihin; Englanniksi = thing, case, matter.

-Adjektiivi-

Haluaisin selvittää adjektiiville "[sana]" joitain taivutusmuotoja. Halutut taivutusmuodot ovat: nominatiivi, genetiivi, partitiivi ja illatiivi yksikössä ja monikossa, sekä komparatiivi

ja superlatiivi. Lisäksi anna adjektiivin perusmuodolle enintään kolme englanninkielistä käännöstä, jos useita käännöksiä on olemassa.

Esimerkkivastaus adjektiivin "pieni" kanssa: Yksikkö: Nominatiivi = pieni; Genetiivi = pienen; Partitiivi = pientä; Illatiivi = pieneen; Monikko: Nominatiivi = pienet; Genetiivi = pienten; Partitiivi = pieniä; Illatiivi = pieniin; Komparatiivi = pienempi; Superlatiivi = pienin; englanniksi = tiny, small, little.

Englanniksi:

-Verbi-

I would like to figure out some of the Finnish conjugation forms for the Finnish verb "[sana]". The conjugation forms are: indicative 1st person present tense, indicative 3rd person past tense, conditional 3rd person present tense, potential 3rd person present tense, imperative 3rd person present tense, 2nd participle and passive past tense. Also, give at most three English translations for the verb, if more than one exists.

An example response with the Finnish word "sanoa": indicative 1st person present tense = sanon; indicative 3rd person past tense = sanoi; conditional 3rd person present tense = sanoisi; potential 3rd person present tense = sanonee; imperative 3rd person present tense = sanokoon; 2nd participle = sanonut; passive past tense = sanottiin; in English = say, tell, speak.

-Substantiivi-

I would like to figure out some of the Finnish conjugation forms for the Finnish noun "[sana]" for both singular and plural. The conjugation forms are: nominative, genitive, partitive, illative. Also, give at most three English translations for the singular nominative of the noun, if more than one exists.

An example response with the Finnish word "asia": Singular: Nominative = asia, Genitive = asian; Partitive = asiaa; Illative = asiaan; Plural: Nominative = asiat; Genitive = asioiden; Partitive = asioita; Illative = asioihin; In English = thing, case, matter.

-Adjektiivi-
I would like to figure out some of the Finnish conjugation forms for the Finnish adjective "[sana]". The forms include: nominative, genitive, partitive, and illative in both singular and plural, alongside comparative and superlative. Also, give at most three English translations for the adjective, if more than one exists.

An example response with the Finnish word "pieni": Singular: Nominative = pieni; Genitive = pienen; Partitive = pientä; Illative = pieneen; Plural: Nominative = pienet; Genitive = pienten; Partitive = pieniä; Illative = pieniin; Comparative = pienempi; Superlative = pienin; in English = tiny, small, little.

Liite 2. Perusmuotoon taivuttamisen kehotepohjat

Suomeksi:

-Verbi-

Haluan selvittää suomenkielisen taivutetun verbin "[sana]" infinitiivi- eli perusmuodon. Kerro lisäksi annetun verbin taivutusmuoto, sekä anna taivutetulle verbille yksi englanninkielinen käännös.

Esimerkkivastaus verbin "sanon" kanssa: Infinitiivimuoto = sanoa, Taivutusmuoto = indikatiivin 1. persoonan preesens, Englanniksi = i say.

-Substantiivi-

Haluan selvittää suomenkielisen taivutetun substantiivin "[sana]" nominatiivin yksikön, eli perusmuodon. kerro lisäksi taivutetun substantiivin taivutusmuoto, sekä anna taivutetulle substantiiville yksi englanninkielinen käännös.

Esimerkkivastaus substantiivin "asiat" kanssa: Nominatiivin yksikkö = asia, Taivutusmuoto = nominatiivin monikko, Englanniksi = things.

-Adjektiivi-

Haluan selvittää suomenkielisen taivutetun adjektiivin "[sana]" nominatiivin yksikön, eli perusmuodon. kerro lisäksi taivutetun adjektiivin taivutusmuoto, sekä anna taivutetulle adjektiiville yksi englanninkielinen käännös.

Esimerkkivastaus adjektiivin "pienet" kanssa: Nominatiivin yksikkö = pieni, Taivutusmuoto = nominatiivin monikko, Englanniksi = small (plural).

Englanniksi:

-Verbi-

I would like to figure out the infinitive- or base form of the conjugated Finnish verb "[sana]". Also, give the form of the conjugated verb and one English translation for the conjugated verb.

An example response with the verb "sanon": Infinitive form = sanoa, Current form = indicative 1st person present tense, In English = i say.

-Substantiivi-
I would like to figure out the nominative singular, or the base form, of the Finnish noun "[sana]". Also, give the form of the conjugated noun and one English translation for the conjugated noun.

An example response with the noun "asiat": Nominative singular = asia, Current form = nominative plural, In English = things.

-Adjektiivi-
I would like to figure out the nominative singular, or the base form, of the Finnish adjective "[sana]". Also, give the form of the conjugated adjective and one English translation for the conjugated adjective.

An example response with the adjective "pienet": Nominative singular = pieni, Current form = nominative plural, In English = small (plural).

Liite 3. Lauseiden korjaamisen kehotepohjat

Suomeksi:
Haluan arvioida tämän suomen kielellä mahdollisesti virheellisesti kirjoitetun lauseen "[lause]" taivuttamisen näkökulmasta. Selvitä mahdollisia taivutusvirheitä lauseessa ja korjaa taivutukset.

Esimerkkikorjaus virheellisen lauseen "Minä pitää mennä." kanssa: Korjattu lause = "Minun pitää mennä."

Englanniksi:
I would like to assess this possibly incorrectly written Finnish sentence "[lause]" from a conjugation standpoint. Figure out the conjugation mistakes in the Finnish sentence and fix them.

An example fix with the incorrectly written sentence "Minä pitää mennä.": Fixed sentence = "Minun pitää mennä."