

Suomenkielinen abstrakti

Kysyvien lausumien vaihtelu ja muutos puhutussa suomessa

Artikkelissa tarkastellaan syntaksin muutosta keskittyen haastateltavien puheen sijasta haastattelijoihin. Kysyviä lausumia tarkasteltiin Helsingin puhekielen pitkittäiskorpuksessa, joka nykyisellään kattaa puoli vuosisataa. Kysymyksistä useimmat ovat haastattelijoiden, joten perinteistä sosiolingvististä variaatiota ei etsitty. Aineistoa työstettiin TurkuNLP-jäsentimen lisäksi litteraattien lähiluvulla. Lausumia oli lopullisessa aineistossa kultakin kolmelta tutkimusvuosikymmeneltä 2000, yhteensä 6000.

Aineiston kysymykset jaettiin viiteen kategoriaan: hakukysymykset, vaihtoehtokysymykset, yhdistelmäkysymykset, partikkelikysymykset ja väitekysymykset. Vaihtoehtokysymykset luokiteltiin edelleen kahteen alaluokkaan; yksinkertaiset ja disjunktiiviset vaihtoehtokysymykset.

Tutkituista kolmesta vuosikymmenestä 1990-luku erottui muista. Hakukysymyksiä oli runsaasti, mutta vaihtoehtokysymysten – sekä yksinkertaisten että disjunktiivisten – määrä oli pienempi kuin 1970- ja 2010-luvuilla. Yhdistelmäkysymyksiäkin oli 1990-luvulla vähiten, joskaan ero 1970-lukuun ei ollut yhtä merkittävä kuin 2010-lukuun. Kysymystyyppien osuuksien muutokset vuosikymmeneltä toiselle olivat tilastollisesti merkitseviä useimmissa tapauksissa.

Lisäksi tarkasteltiin *olla*-verbialkuisia kysymyksiä 300 lausuman otannalla, 100 kultakin tutkimusvuosikymmeneltä. *Olla*-verbialkuisuus liitettäisiin täten puhekumppanin huomioon ottamiseen, kohteliaisuuteen, mutta tässä on tilaisuus jatkotutkimukselle.

Avainsanat: *corpustutkimus, haastattelijan kieli, kysymyslauseet*

Englanninkielinen abstrakti

Variation and change of interrogative utterances in spoken Finnish

Our study covered interrogative utterances in the Longitudinal Corpus of Finnish Spoken in Helsinki, gathered in the past 50 years. The interrogatives in the database have mostly been posed by interviewers. We gathered 6000 interrogative utterances, 2000 from each decade. The novelty value of the research lies on the standpoint: examining the interviewer speech and mixing methods.

The utterances were divided into five categories: seeking questions, alternative questions, combined questions, particle questions and statement questions. We classified alternative questions into two sub-categories: simple and disjunctive alternative questions.

We found that in the 1990s the seeking questions were most common and the alternative questions – both the simple alternative questions and the disjunctive alternative questions – were less common than in the 1970s and the 2010s. The amount of combined questions was also the least in the 1990s, but the difference was not significant comparing to the 1970s; comparing to the 2010s it was. We further examined the verb *olla* ('to be') that we noticed to be extensively used in interrogative utterances, also without a clear need. Based on looking at 300 clauses in more detail, we interpreted that interviewers used frequently that verb to be polite.

Keywords: *corpus linguistics, interrogative utterance, interviewer speech*

Kysyvien lausumien vaihtelu ja muutos puhutussa suomessa

*"The art and science of asking questions
is the source of all knowledge"*

Thomas Berger

1 JOHDANTO

Sosiolingvistiksessä variaationtutkimuksessa on perinteisesti keskitytty fono- ja morfologisiin ilmiöihin, kun taas syntaktisen vaihtelun ja muutoksen havainnointi on jäänyt vähemmälle huomiolle. Tutkimuksemme vastaa tähän tarpeeseen interrogatiivien näkökulmasta. Tarkastelemme kysyviksi lausumiksi nimeämiämme puheen kokonaisuuksia, joita Helsingin puhekielen pitkittäiskorpuksessa esiintyy runsaasti paitsi haastateltavien myös ennen kaikkea haastattelijoiden puheessa. Tutkimuksen uutuusarvo piilee siten syntaksiin liittyvän aiheen lisäksi informantinäkökuulmassa ja monimenetelmällisyydessä (mixed methods). Aineistoa ja sen käsittelyä avataan tarkemmin seuraavassa luvussa. Aluksi hieman teoriataustasta.

Tutkimuksemme nojaa systeemis-funktionaaliseen (SF) kielioppiin, keskustelunanalyysiin, konstruktikielioppiin ja vuorovaikutuslingvistiikkaan. Samanlaista teoriataustaa on aiemmin sovellettu direktiivien tutkimukseen (Lauranto, 2015), ja kysymyksiä voi perustellusti pitää direktiivien alalajina (Searle, 1969 s. 69). SF-kielioppi lähtee kielen monifunktioisuudesta sekä kielenkäytön ja kielisysteemin vuorovaikutteisuudesta: kieli on samanaikaisesti funktionaalinen ja systeeminen, ja sen käyttö ja jäsenytyneisyys ovat jatkuvassa vuorovaikutuksessa (Paldanius & Sandberg, 2021 s. 452).

Konstruktikielioppi puolestaan näkee kieliopin ja leksikon saman jatkumon osina, jossa niin yksittäiset lekseemit kuin kieliopilliset rakenteetkin ovat kielellisiä merkkejä, joissa muoto ja merkitys ovat erottamattomasti kytkeytyneitä toisiinsa (Hoffmann & Trousdale, 2013 s. 1). Konstruktikielioppi on myös

tutkimusperinteenä vahvasti käyttöpohjainen. Samankaltainen ajatus aineistolähtöisyydestä on myös keskustelunanalyysissä, jossa yksityiskohtien havainnoinnista edetään laajempaan kokonaiskuvaan ja yleistykseen (Vatanen, 2021). Kokonaisuutena tällaista keskustelunanalyysin ja funktionaalisen kielitieteen synteesiä voi pitää vuorovaikutuslingvistiikkana.

Aiemmassa fennistisessä tutkimuksessa suomenkielisiä kysyviä rakenteita on tutkinut mm. Forsberg (1992, 1994, 2006), joka on ehdottanut (1992) kaksoiskysymysten olevan keino pehmentää ja lieventää kysymyksen jyrkkyyttä. Tässä työssä viittaamme samaan kysyvien rakenteiden ilmiöön termillä *yhdistelmäkysymys*. Forsberg on myös ehdottanut (2006, s. 373), että Helsingin seudun puhekielessä esiintyvä kysymysmuotojen liitteettömyys olisi itäistä vaikutusta. Siinä missä Forsbergin tutkimus on perustunut leimallisesti perinteisiin murreaineistoihin, me keskitymme Helsingin puhekielestä koottuun aineistoon. Forsberg on havainnut liitepartikkelia *-ko* voitavan yhdistää verbien lisäksi adjektiiveihin tai adverbeihin kysymysten muodostamiseksi. "*Vanhaks tiä ruuna o?*" "*Kovaako tämä kulkee?*" "*Useinko harrastat liikuntaa?*" (2006, s. 376–378), vaikkakin nykysuomessa *montako* on jo selvästi kiteytynyt merkitykseen 'kuinka monta' ja siis muuntunut vaihtoehtokysymyksestä hakukysymykseksi. Tutkimuksessamme *otat sä / otatsä* on kirjoitusasusta riippumatta tulkittu *otatko*-kysymyksen allomorfixi liitteettömän kysymyksen sijasta.

Aineiston laajentaminen pelkästä informanttien puheesta kattamaan myös haastattelijoiden puheen vaikuttaa selvästi aineiston luonteeseen. Forsbergin aineiston kysymyksistä valtaosa on retorisia ja joukkoon on osunut vain muutamia aitoja kysymyksiä, kun taas haastattelijoiden puheen mukaan ottaminen muuttaa kuvaa merkittävästi. Kokonaisia haastatteluja tarkastellessamme emme voi sanoa, että aidot kysymykset olisivat mitenkään vähemmistössä, vaan ne muodostavat nimenomaan aineistomme ytimen. Karhu (1995) on tutkinut murrehaastatteluja vuorovaikutustilanteissa. Hän esittää, että haastatteluille on tyypillistä kaksitasoisuus merkityssisällöissä. Haastattelijalla ylläpitää yleistä tasoa, kun taas haastateltava liikkuu eniten yksityistasolla. Haastattelijalla kysyy esimerkiksi "*Miten*

ennen vanhaan tehtiin X:ää?". Karhun havaintojen mukaan haastateltavat vastaavat tällaisiin kertoen niistä kokemuksista, joita heillä itsellään on X:n tekemisestä. Murrehaastattelussa täten risteillään yleisen ja yksityisen välillä edestakaisin. (Mts. 97.) Tässä työssä keskitymme haastattelijoiden kysymyksiin, joten aineistomme on enemmän yleisellä kuin yksityiskohtaisella tasolla.

Olin (1996) on tutkinut kysymysten rakenteesta heijastuvaa kuulijaan suhtautumista. Hän kuvaa kysymyksen tarkoittavan koko lausuman morfosyntaktista ilmiä tai sen kysyvää funktiota (mts. 47), ja myös tässä tutkimuksessa tarkoitamme kysymyksellä näitä molempia. Olin laati jaottelun kysymysrakenteista. Hyödynsimme tätä lähestymistapaa puhekielen korpusaineistoomme, mutta Olinin nimestysaineistoa varten laatima kysymysrakenteiden jaottelu ei kaikilta osin soveltunut käyttöömme, joten laadimme meille paremmin soveltuvan kategorisoinnin.

Käyttämämme kategorisointi muistuttaa osittain myös Raevaaran (2006) vastaavaa. Hän tutki Kelan virkailijoiden asiakkailleen esittämiä kysymyksiä. Raevaaran lajittelussa on ensinnäkin tavallisimmat kysymysrakenteet eli interrogatiivit, jotka hän jakaa hakukysymyksiin ja vaihtoehtokysymyksiin. Näistä jälkimmäiset voivat olla myös disjunktiivisia. Lisäksi Raevaara mainitsee partikkelikysymykset ja väitemuotoiset lausumat, jotka voivat toimia tarkistavina kysymyksinä (*"Ja asut Kouvolassa"*). Raevaara kuvailee, että väittämällä voidaan pyytää joko vahvistamaan tai kieltämään jonkin asiointilan paikkansapitävyys. Samoin toimii tietenkin myös interrogatiivinen vaihtoehtokysymys, mutta esittämällä kysymyksen nimenomaan väitteen muodossa puhuja ilmaisee olevansa astetta varmempi esille ottamansa vaihtoehdon todenmukaisuudesta.

Noudatamme kysymysten kategorisoinnissa pääosin Ison suomen kieliopin (2004, § 1679) linjaa, jossa kysymyslauseet ja muut interrogatiiviset rakenteet on luokiteltu vaihtoehtokysymyksiin, hakukysymyksiin ja partikkelikysymyksiin; vaihtoehtokysymyksiä voi edelleen olla yksinkertaisia ja disjunktiivisia. Olemme kuitenkin tulkinneet partikkelikysymykset hiukan laajemmin sekä ottaneet mukaan

väitekysymykset. Lisäksi aineistossamme oli yhdistelmäkysymyksiä niin runsaasti, että pidimme tarkoituksenmukaisena tarkastella niitä erillisenä kategoriana.

Lääkäreiden potilailleen esittämiä kysymyksiä on tarkastellut Korpela (2007) vuorovaikutuslingvistiikan viitekehyyksessä. Erilaisilla kysymysrakenteilla on lääkärin vastaanotolla omat tehtävänsä, ja tämä on näkynyt lausuman kokonaisrakenteen lisäksi siinä, millaisia kielenaineiksia lausumatyypeissä ollut. Esimerkiksi perinteisesti suljetuiksi kysymyksiksi tulkitut verbikysymykset (*"Liittyykö siihen mitään närästyä?"*) toimivat vastaanotolla usein avoimina kysymyksinä eli sellaisina, joihin haetaan laajaa lausevastausta. Myös perinteisesti avoimina pidetyt hakukysymykset (*"Mistä kohti se on kaikista pahin se kipu?"*) voivat toimia vastaanotolla niin suljettujen kuin avoimienkin kysymysten tapaan. Tulkintaan liittyy kysymyksen rakenteen lisäksi erityisesti se toimintakonteksti, jossa kysymykset esitetään.

Aiemmat näkemykset siitä, että kaksoiskysymys olisi tiukasti vain puhutun kielen rakenne tai peräti pelkkä itämurteisuus, on haastanut Jaskara (2009). Hän on osoittanut, että kysymystyyppi on laajempi koko puhutun suomen ja myös kirjoitetun yleispuhekielen ilmiö. Rakenne, jossa on sekä *-ko* että interrogatiivipronomini, ei ole pelkkä hakukysymyksen parafraasi, vaan sillä on vuorovaikutustilannetta kannatteleva funktio. Olemme samalla kannalla, emmekä ole kyseenalaistaneet kaksoiskysymysten eli yhdistelmäkysymysten mielekkyyttä: Pidempi pohjustus auttaa kuulemaan kysymyksen, poimimaan tärppinä varsinaisen kysymyssanan, joka ei tule lauseen alussa, ja antaa kuulijalle aikaa vastauksen miettimiseen.

Kysymysten muoto voi muuttua ajan myötä. Tätä on tarkastellut Suomalainen (2020). Hän on tutkinut, onko *tiedätkö*-ilmaus muuttumassa verbimuodosta partikkelistuneeksi ilmaukseksi *kato-* ja *kuule-*sanojen kaltaisesti. Kiteytyneenä ilmauksena *tiedätkö* ei enää esitä vastaanottajalle suoraa kysymystä, vaan sitä käytetään peräänkuuluttamaan osanottajien jakamaa yhteistä tietoa ja tätä kautta säätelemään heidän välistään yhteisymmärrystä. Me emme tutkimuksen rajaamisen

nimissä käsittelee *tiedätkö*-tapauksia omassa aineistossamme partikkelistuneina ilmauksina vaan aitoina kysymyksinä, kuten *muistatko*, mutta tämän kaltainen muutos on tärkeä huomioida kysymyksiä tutkittaessa.

Viittaamme sanalla *puhekieli* siihen puhuttuun suomen kieleen, josta Helsingin puhekielen pitkittäiskorpuksen materiaali on peräisin. *Kysymyksellä* tarkoitamme niin rakenteellisia kuin funktionaalisia kysymyksiä. Rakenteellisessa kysymyksessä on kysymyssana tai *-ko*-liitepartikkeli, kun taas funktionaalinen kysymys voi olla esimerkiksi "*Kirjastossa?*" riippuen kontekstista. Suurin osa aineistostamme on näitä sekä rakenteellisesti että funktionaalisesti aitoja kysymyksiä ja vain pieni osa on rajatapauksia. Niitä puhekielen kokonaisuuksia, jotka sisältävät kysymyselementin, kutsumme tässä artikkelissa *kysyviksi lausumiksi*, ja ne muodostavat tutkimuksemme ytimen.

Tarkastelemme käsillä olevassa tutkimuksessa kysyvien lausumien kategorisoinnin ohella *olla*-verbialkuisia kysymyksiä pienellä, 300 lausuman otannalla. Huomiomme on kiinnittynyt siihen, miten yleistä sen käyttö puheessa on varsinkin kysymyksen aloittajana – silloinkin, kun se ei näennäisesti kysymykseen kuulu. Katsomme, tarjoaisiko korpusaineiston lähiluku vastauksia *olla*-verbin funktioon kysymyksen aloittajan roolissa.

Olemme kiteyttäneet tutkimuskysymyksemme seuraavasti:

1. Miten puhekielen kysymyksiä voi kategorisoida korpusaineistossa?
2. Millaisia kysyviä lausumia Helpuhe-aineistosta löytyy?
3. Millaisia muutoksia (lähinnä) haastattelijan esittämät kysymykset ovat kokeneet 1970-luvulta 1990-luvulle ja edelleen 2010-luvulle?

2 AINEISTO JA MENETELMÄT

2.1 Aineiston käsittely

Tutkimuksen aineistona käytämme Helsingin puhekielen pitkittäiskorpusta, joka sisältää eri-ikäisten syntyperäisten helsinkiläisten haastatteluja 1970-, 1990- ja 2010-luvuilta ja niihin kohdistettuja litteraatteja. Taulukossa 1 on esitelty haastattelujen jakauma eri vuosikymmenille. Tässä tutkimuksessa keskitymme kysymyksiin, joita haastattelunauhoitteissa esittävät pääasiassa haastattelijat.

Taulukko 1. Helsingin puhekielen pitkittäiskorpuksessa mukana olevat haastattelut (Kielipankki 2024 ja Helsingin yliopisto 2024)

	1970-luku	1990-luku	2010-luku	Yhteensä
Haastattelut	96	45	43	184

Kysymysten löytämiseksi aineistosta käytimme apuna automaattisia työkaluja. Aineiston morfologiseen jäsennykseen käytimme TurkuNLP-ryhmän kehittämää jäsenintyökalua (Kanerva, Ginter, Miekka, Leino & Salakoski, 2018). Jäsenntystä aineistosta kysymykset voidaan etsiä kysymyssanojen ja -liitteiden avulla.

Jäsenintyökalu toimii parhaiten yleiskieliselle suomen kielelle, minkä takia yleiskielistimme aineiston konekäännösmallien avulla. Konekäännösmalleja on käytetty runsaasti murreaineistojen yleiskielistämiseen (esim. Partanen, Hämäläinen & Alnajjar, 2019; Kuparinen, Miletic & Scherrer, 2023). Koska käännettävät kielet ovat läheisiä, tulokset ovat yleensä melko hyviä. Tutkimuksemme tarpeisiin yleiskielistykseen ei myöskään tarvinnut olla täydellinen vaan riittävä kysymyslauseiden poimintaan, sillä teimme varsinaisen analyysin alkuperäisistä litteraateista.

Konekäännösmallia varten tarvitsimme pienen harjoitusaineiston, jonka yleiskielistimme käsin seuraten *Suomen kielen näytteitä* -sarjan yleiskielistyksessä käytettyjä ohjeita (Vilkuna, 2014). Harjoitusaineistossa oli yhteensä 1501 lausumaa aineiston eri vuosikymmeniltä painottuen kysymyksiin, jotta ne saataisiin

paremmalla tarkkuudella oikein. Testiaineistona käytettiin vastaavasti koottua 304 lausuman joukkoa.

Testasimme kahta eri yleiskielistysmallia: perinteistä merkkeihin perustuvaa tilastollista konekäännösmallia (Koehn et al., 2007) sekä esiopetettua, tietokonetavuihin perustuvaa ByT5-mallia (Xue et al., 2022). Yleiskielistyksen tulokset esitetään merkkivirheenä (kuinka moni merkki sadasta on väärin; pienempi arvo on parempi). Vaihtoehdoista tilastollinen malli tarjosi huomattavasti paremman tuloksen merkkivirheellä arvioituna (5.19 vs. 13.31). Niinpä yleiskielistimme koko aineiston tilastollisella konekäännösmallilla, jonka jälkeen käytimme jäsenntyyökälyä yleiskielistetyille aineistolle. Jäsenntyyökälyän morfologisesta analyysistä poimittiin lausumat, joissa esiintyi joko kysymyssana tai kysymysliite (-*ko*).

Automaattisesti poimimalla kysymyksiä löytyi huomattavasti enemmän 1970-luvulta verrattuna kahteen muuhun tutkimusvuosikymmeneen (taulukko 2). Tästä syystä keräsimme alkuperäisistä litteraateista manuaalisesti lisää kysymyksiä 1990- ja 2010-luvuilta ja vähensimme automaattisesti poimittuja 1970-luvulta. 1970-luvun poistot tehtiin aineistosta tasaisesti, jotta yksittäiset haastattelut eivät yliedustuisi. Näin tarkastelemme yhteensä 2000 kysymystä kultakin vuosikymmeneltä, jolloin tarkasteltujen kysyvien lausumien määrä tässä tutkimuksessa on kaikkiaan 6000.

Taulukko 2. Tarkasteluun valitut kysyvät lausumat

	<i>1970-luku</i>	<i>1990-luku</i>	<i>2010-luku</i>	<i>Yhteensä</i>
<i>Automaattisesti poimittu</i>	5584	1373	1987	8944
<i>Manuaalinen lisäys</i>	–	627	13	640
<i>Tarkasteltu määrä</i>	2000	2000	2000	6000

Väite- tai partikkelikysymyksiä ("*Sulla ei oo lapsia?*", "*Entäs sodan jälkeen?*") ei voi löytää jäsentimen avulla, koska niissä ei esiinny kysymyssanoja tai -liitteitä, ja

litteraateissa kysymyksiä ei ole merkitty myöskään kysymysmerkillä. Tällaisia kysymyksiä poimimme aineistosta manuaalisesti.

Kysymysten automaattisessa poiminnassa löytyi niin pää- kuin sivulauseita, jotka sisältävät kysymyselementin. Näin ollen myös epäsuorat kysymyslauseet tulkitsimme tässä tutkimuksessa kysyviksi lausumiksi. Lisäksi automaattisessa yleiskielistyksessä tai jäsennyksessä on voinut tapahtua virheitä, jotka johtivat sellaisten lausumien poimintaan, joissa ei lopulta ollut lainkaan kysymyselementtiä. Toistuvia tällaisia tapauksia ovat esimerkiksi *ts*-yhtymän sisältävät lausumat (esim. *tsiihen aikaan*), jotka sekoittuvat moniin kysymysliitteen ja persoonapronominin yhdistelmiin (esim. *otatsä*) sekä kysymyssanan sisältävät sanat, jotka eivät kuitenkaan ole kysymyksiä (*mitenkään*). Tällaisia lausumia poistimme manuaalisesti ja korvasimme ne aidoilla kysymyksillä.

2.2 Kysyvien lausumien kategorisointi

Luokittelimme kysymykset viiteen eri kategoriaan sekä edellisessä luvussa esitellyn kirjallisuuden että aineiston käsittelyn yhteydessä tehtyjen havaintojen perusteella. Käsillä olevaa tutkimusta varten luomamme kategoriat ovat: hakukysymykset, vaihtoehtokysymykset, yhdistelmäkysymykset, partikkelikysymykset ja väitekysymykset. Vaihtoehtokysymykset luokittelimme edelleen kahteen luokkaan: yksinkertaiset ja disjunktiiviset vaihtoehtokysymykset. Kategoriajako esimerkkeineen esitellään taulukossa 3.

Taulukko 3. Kysymyskategoriat

<i>Numero</i>	<i>Nimi</i>	<i>Määritelmä</i>	<i>Esimerkki</i>
1	hakukysymys	sisältää kysymyssanan	<i>Missä sä asut?</i>
2a	vaihtoehtokysymys, yksinkertainen	sisältää - <i>kO</i> tai sen allomorfin	<i>Ooksä pitkään asunu Helsingissä?</i>
2b	vaihtoehtokysymys, disjunkttiivinen	sisältää - <i>kO</i> ja antaa vaihtoehdon: <i>vai</i> (tässä myös <i>tai</i>)	<i>Kuljeksä töihin dösällä vai sporalla?</i>
3	yhdistelmäkysymys	mikä tahansa kysyvä lausuma, joka yhdistää montaa eri kategoriaa	<i>Onkse kämppä missä?</i>
4	partikkelikysymys	kysyvyyttä osoittaa pelkästään <i>vai</i> , <i>entä(s)</i> , <i>ai</i>	<i>Entä kesämökki?</i>
5	väitekysymys	Väitelause, joka tulkitaan kontekstissaan kysymykseksi ja johon vastataan kuin kysymykseen. Ei rakenteellisia vihjeitä siitä, että lausuma olisi kysymys.	<i>Ei sul lapsii oo?</i>

Kategoria 1 (hakukysymys) muodostuu kysyvistä lausumista, jotka sisältävät kysymyssanan. Tähän kategoriaan ei laskettu sellaisia kysymyksiä, joissa oli muitakin varsinaisia kysyviä elementtejä pelkän kysymyssanan lisäksi. Jos kysymyksessä oli interrogatiivisen pronominin tai adverbien lisäksi -*kO*, se lankesi kolmanteen kategoriaan (yhdistelmäkysymys). Jos kysymyksessä oli esimerkiksi *entä*- tai *vai*-sana kysymyssanan lisäksi, se oli silti hakukysymys. Esimerkeissä 1–3 on esitetty yksi hakukysymys kultakin vuosikymmeneltä.

(1) Missä te kävitte kansakoulua? (1970)

(2) Kuka sen kuusen koristeet sit ottaa pois? (1990)

(3) Eli sillon nuorena sä ainaki asuit siellä Kalliossa ni mihinkäs matka jatku sit sieltä?
(2010)

Kategoria 2 (vaihtoehtokysymys) käsittää kaikki vaihtoehtokysymykset mutta kahteen luokkaan jaettuna: a) yksinkertaiset ja b) disjunkttiiviset

vaihtoehtokysymykset. Vaihtoehtokysymyksen tunnistaa siitä, että se sisältää kysymyspartikkelin *-kO* tai sen allomorfin, kuten *ootsä, ooksmä, oottekste, luuleeksne, muistatsä, tiesikse*. Ne voivat olla kokonaan tai osittain yhteen sulautuneita persoonapronominien kanssa. Yksinkertainen vaihtoehtokysymys (2a) eli polaarinen interrogatiivilause on tunnistettavissa lauseen ensimmäiseen jäsenen kiinnittyvästä *-kO*-liitepartikkelista. Disjunkttiivinen vaihtoehtokysymys (2b) puolestaan sisältää saman liitteen mutta antaa myös yhden tai useamman vaihtoehdon, joiden yhteydessä on konjunktio *vai/tai*. Kirjallisuus on nimennyt tässä yleensä vain *vai*-konjunktin (esim. ISK 2004, § 1679), mutta aineistossa esiintyi runsaasti myös *tai*-konjunktia. Esimerkissä 4 on yksinkertainen vaihtoehtokysymys ja esimerkeissä 5 ja 6 disjunkttiivisia vaihtoehtokysymyksiä.

(4) Etteks te pitäny koulunkäynnistä? (1970)

(5) Maksaako ne sulle tästä jotain ku sä valmennat vai teek sä sitä vaan hyvää hyvyttä? (1990)

(6) Että ammuskelitiinko joillain aseilla vaikka tai räjäyteltiinkö jotai? (2010)

Kategoria 3 (yhdistelmäkysymys) sisältää kaikenlaiset kysymysten yhdistelmät (esimerkit 7–9).

(7) Mites nuo elokuvakerhot, oot sä ollut niissä mukana? (1970)

(8) Millainen ihminen sun isäs oli tulik sä sen kans hyvin toimeen? (1990)

(9) Onks sul jotain mielikuvaa et ne jotka käyttää Hesa ni mimmosii ne on? (2010)

Kysymyskategorioista neljäs ja viides, partikkelikysymykset ja väitekysymykset, olivat sellaisia, joita emme pystyneet löytämään automaattisesti. Partikkelikysymyksissä kysyvyyttä osoittaa ainoastaan partikkeli tai konjunktio: *entä(s), vai, ai, ai jaa* (ISK § 1679; ISK § 1701). Väitekysymyksistä partikkelikysymykset erosivat siten, että partikkeli on pakollinen osa kysymystä, kun taas väitekysymys toimii täysin ilman rakenteellista kysyvää osaa – kysymykseksi se on mahdollista hahmottaa ainoastaan kontekstinsa ja joskus myös puheen intonaation perusteella. Esimerkeissä 10–12 on partikkelikysymyksiä ja esimerkeissä 13–15 väitekysymyksiä.

(10) Entä koulun pihalla? (1970)

(11) Espoo om maaseutua vai? (1990)

(12) Ai sitä jätskiikää okei? (2010)

(13) Sä tykkäät et Suomen radiost ei tuu hyvin musiikkii? (1970)

(14) Synttärit. Ennen ja nyt? (1990)

(15) Ööö mummulla oli? (2010)

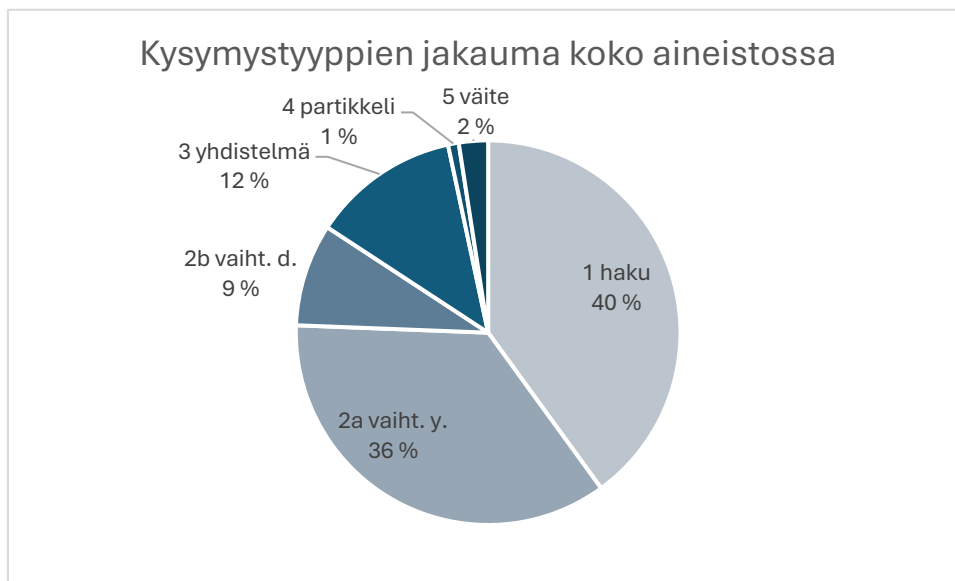
Partikkelikysymysten ja väitekysymysten kategoriat perustuivat siis kysymysten manuaaliseen poimintaan aineistosta. 1990- ja 2010-lukujen aineistoa manuaalisesti täydennettäessä näiden kysymysten poimiminen oli luontevaa. Jotta vuosikymmenkohtaiset aineistot vastaisivat toisiaan, on partikkeli- ja väitekysymyksiä poimittu manuaalisesti myös 1970-luvun aineistosta siinä yhteydessä, kun jäsentimen virheellisesti löytämiä ei-kysymyksiä on korvattu varsinaisilla kysyvillä lausumilla.

3 TULOKSET

Tarkastelemme tässä luvussa aineistomme kysyviä lausumia ensin kokonaismäärittäin tyypityksen mukaan ja sitten vuosikymmenkohtaista tyyppien edustusmäärän muutosta. Tämän jälkeen pohdimme kysymystyyppejä laadullisesti sekä esittelemme tarkemmin vielä yhden erityistapauksen, *olla*-verbialkuiset kysyvät lausumat.

3.1 Tarkastelu kokonaismäärittäin

Analysoimme tutkimuksessa siis kaiken kaikkiaan 6000 kysyvää lausumaa, joista 2000 edusti 1970-luvun aineistoa, 2000 lausumaa 1990-lukua ja 2000 oli peräisin 2010-luvun puhekielestä. Näistä 6000 lausumasta hakukysymyksiä oli yhteensä peräti 2400, yksinkertaisia vaihtoehtokysymyksiä 2137, disjunkttiivisia vaihtoehtokysymyksiä 517, yhdistelmäkysymyksiä 745, partikkelikysymyksiä 54 ja väitekysymyksiä 147. Kuvio 1 avaa kategorioiden suhteellista jakaumaa.



Kuvio 1. Kysymyskategorioiden suhteelliset osuudet koko aineistossa

Kysyvistä lausumista 85 % koostui hakukysymyksistä ja vaihtoehtokysymyksistä. Hakukysymyksiä oli 40 %, vaihtoehtokysymyksiä 45 %. Valtaosa vaihtoehtokysymyksistä oli yksinkertaisia. Loput 15 % aineistomme kysyvistä lausumista koostui yhdistelmäkysymyksistä (12 %), väitekysymyksistä (2 %) sekä partikkelikysymyksistä (1 %).

On relevanttia tarkastella vaihtoehtokysymysten määrää myös kokonaisuutena. Niiden kysymyskategoriat ovat hyvin toistensa kaltaisia: niissä kummassakin on -*kO* tai sen allomorfi, mutta disjunkttiivinen kysymys täydentyy *vai/tai*-konjunktiolla. Siksi tarkastelimme, mitä esiintymämäärille tapahtuisi, jos kumpikin vaihtoehtokysymyksen kategoria laskettaisiin yhdeksi. Kokonaismäärä tulee ilmi taulukosta 4.

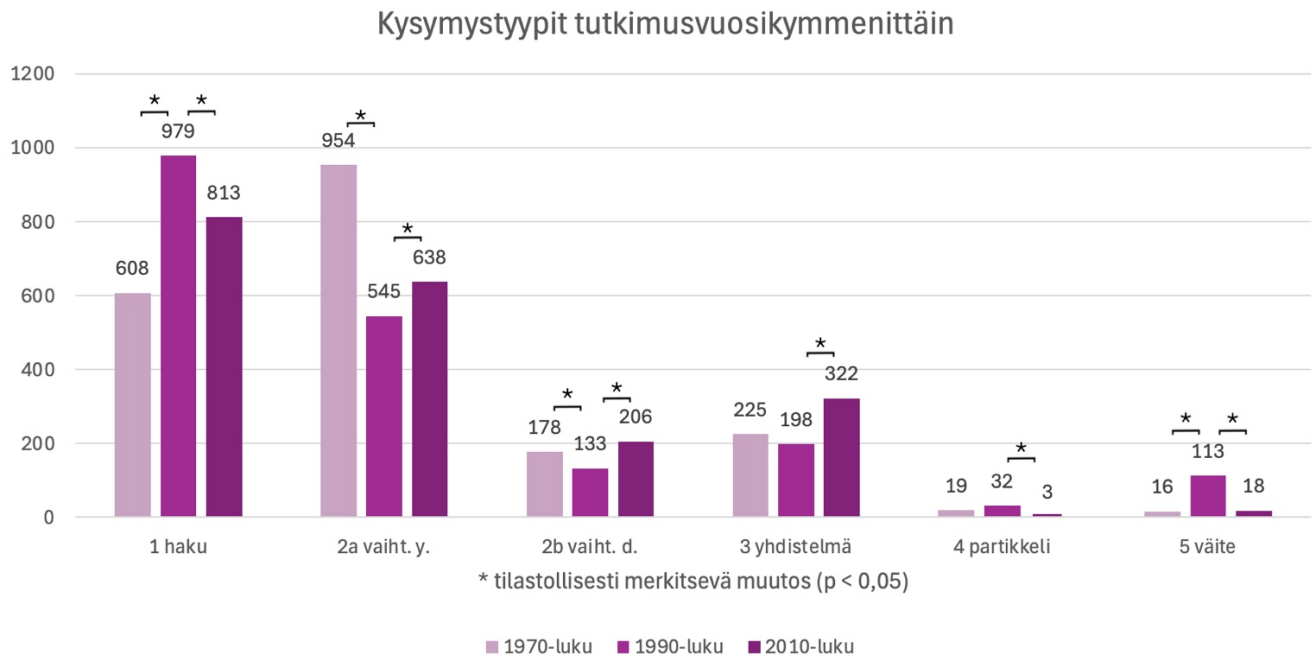
Taulukko 4. Vaihtoehtokysymykset

	<i>2a yksinkertainen</i>	<i>2b disjunkttiivinen</i>	<i>Yhteensä</i>
<i>1970</i>	954	178	1132
<i>1990</i>	545	133	678
<i>2010</i>	638	206	844

Hakukysymyksiä oli 1970-luvulla 608, 1990-luvulla 979 ja 2010-luvulla 813 eli yhteensä 2400. Verrattaessa niiden määrää yllä oleviin vaihtoehtokysymysten (joita yht. 2654) määriin, huomataan, että 1970-luvulla vaihtoehtokysymysten määrä oli liki kaksinkertainen. 1990-luvulla hakukysymyksiä oli kolmisensataa enemmän kuin vaihtoehtokysymyksiä, ja 2010-luvulla nämä kaksi kategoriaa olivat lähes tasasuuruiset.

Näiden tulosten perusteella hakukysymykset ja vaihtoehtokysymykset vaikuttavat olevan ne syntaktiset systeemit, joilla suomenkielisessä puheessa, tässä tutkimuksessa haastattelupuheessa, kysymys muodostetaan. Koska aineiston haastattelijoiden on ollut käytössään haastattelurunko, he kysyivät toistuvasti samoista teemoista. Laadullisessa tarkastelussa vaikuttaa siltä, että kaikista aiheista on keskusteltu kaikenlaisilla kysymyksillä, joten teemalla ei vaikuttaisi olleen ainakaan suoraa vaikutusta kysymyskategoriaan.

3.2 Tarkastelu vuosikymmenittäin



Kuvio 2. Kysymyskategorioiden ajallinen vaihtelu tutkimusvuosikymmenittäin

Kysymyskategorioiden esiintymämäärät vaihtelivat eri vuosikymmenillä. Kuten kuvio 2 käy ilmi, yleisimmät kysymyskategoriat aineistossamme olivat haku- ja vaihtoehtokysymykset eli kategoriat 1 ja 2. Vuosikymmenet erosivat kategorioissa. Yleisimmät kategoriat 1970-luvulla olivat vaihtoehtokysymykset, 1990- sekä 2010-luvulla puolestaan hakukysymykset. Muita kategorioita, joissa oli havaittavia eroja vuosikymmenten välillä, olivat yhdistelmäkysymykset, joita löytyi eniten 2010-luvulta, sekä väitekysymykset, joita löytyi eniten 1990-luvulta. Kaikilla vuosikymmenillä partikkelikysymysten osuus oli pieni, samoin kuin väitekysymysten lukuun ottamatta 1990-lukua.

1990-luku vaikuttaa kuvion 2 perusteella poikkeavan selkeimmin muista vuosikymmenistä: sen väitekysymysten määrä oli huomattavan suuri ja myös hakukysymyksiä oli kaikista vuosikymmenistä suurin joukko. 1970-luvulla korostuivat yksinkertaiset vaihtoehtokysymykset, kun taas 2010-luvulla oli vuosikymmenistä eniten yhdistelmäkysymyksiä.

Teimme eniten manuaalista poimintaa nimenomaan 1990-luvun aineistosta (taulukko 2). Vaikka manuaalista poimintaa ja väärin poimittujen kysymysten korvaamista tehtiin myös muiden vuosikymmenten aineistoille, väitekyymysten piikki selittynee ainakin osin manuaalisesti poimittujen kysymysten määrällä. Sen sijaan hakukysymysten runsautta manuaalinen poiminta ei selitä, sillä ne on poimittu ensisijaisesti automaattisesti jäsentimen avulla.

Kuviossa 2 asteriski tarkoittaa, että kysymystyyppin suhteellisessa osuudessa on tapahtunut tilastollisesti merkitsevä muutos vuosikymmenten välillä. Merkitsevyys on todettu khiin neliö -testeillä, joiden tuloksiin tehtiin lopuksi Holm-Bonferroni-korjaus. Merkitsevyyden rajana pidetään yleisesti $p < 0,05$. Testissämme nollahypoteesina oli, että ryhmän ja vuosikymmenen vaikutukset esiintymämäärään ovat toisistaan riippumattomat. Testien tuloksena merkitseviä ovat kaikkien muiden kysymystyyppien määrän muutokset paitsi yhdistelmäkyymykset 1970-luvulta 1990-luvulle ja partikkelikyymykset 1970-luvulta 1990-luvulle, joissa p-arvo ylittää 0,05:n. Kaksi kolmasosaa muutoksista oli lisäksi testien mukaan tasoltaan tilastollisesti erittäin merkitseviä ($p < 0,001$). Näin oli esimerkiksi hakukysymyksissä ja väitekyymyksissä sekä 1970-luvulta 1990-luvulle että 1990-luvulta 2010-luvulle tultaessa.

3.3 Laadullinen pohdinta

Kuten todettua, hakukysymykset ja vaihtoehtokysymykset yhdessä muodostivat 85 % kaikista tutkituista kysyvistä lausumista. Näiden keskinäinen jakauma oli melko tasainen, 40 ja 45 prosenttia. Mistä kertoo, että juuri nämä kysymystyyppit nauttivat kansan suosiota? Osaltaan syytä voi hakea niiden yksiselitteisyydestä. Niiden merkityksestä ei juuri voi erehtyä. Ne ovat perinteisimpiä tapoja tiedustella suomeksi jotakin asiaa, johon halutaan vastaus.

Tulkintaa kysymyksestä tukee vahvasti lausuman rakenne. Suomeksi kysymyslauseen intonaatio on usein samankaltainen kuin väitelauseen lukuun ottamatta aivan lyhyitä tarkistuskysymyksiä, joten tällainen rakenteellinen tuki

merkityksen muodostamiselle voi olla tarpeen. Esimerkiksi ukrainan kielessä puolestaan lauseet *"Hän on lääkäri"* ja *"Onko hän lääkäri?"* (ukr. *вона лікар*) erottaa ainoastaan kysymyslauseen nouseva intonaatio, koska lauseessa ei ole minkäänlaista kysymyssanaa. Intonaatiolla ei suomea puhuttaessa ole niin merkittävää roolia kuin indoeurooppalaisissa kielissä, koska intonaatio harvoin, jos koskaan, johtaa suomessa väärinymmärryksiin.

Puhekielessä kysymysten esittäminen on sikäli moniulotteinen operaatio, että puhtaiden hakukysymysten tai vaihtoehtokysymysten lisäksi kysyjä voi hyödyntää näiden yhdistelmää, yhdistelmäkysymystä, mikäli katsoo sen palvelevan kontekstia parhaiten. Niin ikään aiemmin tässä tutkimuksessa esitetyt partikkeli- ja väitekysymykset ovat mahdollisia, vaikkeivät puristisessa mielessä kieliopillista kysymystä edustakaan. Ne antavat puhujan repertuaariin kuitenkin lisää valinnanvaraa ja voivat toimia joustavina, lyhyinä ja tarkentavina välikysymyksinä kesken puhekumppanin kertoman tarinan. *"Ai Jonna vai?"*, *"Jaa silloin kun sä asuit Saksassa?"* Haku- tai vaihtoehtokysymys veisi huomattavasti enemmän niin sanotusti tilaa ja aikaa keskustelusta – siihen eivät partikkeli- ja väitekysymykset tähtää, vaan ne tarvitsevat vain nopean *joo* tai *ei* -tarkennuksen tultuaan kenties epäkohteliaasti väliin keskeyttämään puhujan. Mielestämme eri kysymystyyppien rooli vuorovaikutuksessa olisi antoisa jatkotutkimuksen paikka.

Lisäksi näitä kaikkia puhekielessä esiintyviä kysymystyyppisiä täydentämään voidaan valita ns. ylimääräinen *olla*-verbi, joka näennäisesti kysymyksen rakenteen kannalta tarkasteltuna mutkistaa mutta toisaalta vuorovaikutuksen tavoitteen näkökulmasta voi auttaa selkeyttämään kysymystä. Seuraavassa alaluvussa luomme tähän erityistyyppiin vielä tarkemman katsauksen.

3.4 *Olla*-verbilliset kysyvät lausumat

Käsitlemme tässä alaluvussa tarkemmin aineistomme *olla*-verbillisiä kysymyslauseita ja varsinkin *olla*-verbin roolia kysyvän lausuman aloittajana. Sen käyttö suomenkielisissä kysymyksissä tuntuu olevan silmiinpistävän runsasta, ja

siltä se vaikutti myös meidän aineistossamme. *Olla*-verbillä halutaan aloittaa kysymyslause, vaikkei kysymys lainkaan tarvitsisi sitä kieliopillisessa mielessä.

Etsimme aineistostamme yhteensä 300 *olla*-verbillistä kysyvää lausumaa, 100 kultakin tutkimusvuosikymmeneltä. Manuaalinen poiminta suoritettiin kronologisessa järjestyksessä dokumenteista, jotka jäsenin oli jo suodattanut. Havaitimme, että *olla*-verbin suosituin paikka oli lauseen alussa tai vähintään sen alkupuolella. *Olla*-verbin vieras käyttö ja vieras sijainti lauseessa yleiskieleen nähden toivat aineistossa kysymyksiin leimallista puhekielisyttä. Esimerkit 16–18 havainnollistavat *olla*-verbillisiä kysymyksiä.

(16) Onks sul muistat sä mimmosii käsitöitä sä oot tehny kässätunneilla? (2010)

(17) Mites tota, onks se sun puku nii vuokrattiinks se jostai vai ostettiin vai teitkö ite vai? (2010)

(18) No onks jotai meetsä mielellään Espooseen tai Vantaalle tai? (2010)

Tulkintamme mukaan *olla*-rakenteen käyttö kysyvän lausuman alussa liittyy kuulijan huomion herättämiseen ja puhutteluun sekä tarjoaa kuulijalle aikaa valmistautua varsinaiseen kysymykseen. Aloittaminen jollain muulla kuin oleellisella kysymyssanalla, joka on tarpeen varmasti kuulla, liittyy täten ennen kaikkea kohteliaisuuteen. Tällä tavoin vältetään rasittamasta kysymyksen kuulijaa sillä, että hän joutuisi mahdollisesti kysymään "*Mitä?*", jolloin hän joutuisi puolestaan rasittamaan tällä toiston toiveella kysyjää. Kysyjä haluaa siis jo alun perin välttää toistamasta kysymystään uudelleen, joten hän aloittaa *olla*-verbillä eikä vaikkapa kysymyssanalla, jonka kuuleminen olisi lausuman ymmärtämisen ja vastauksen oikean muotoilun kannalta ehdotonta.

Olla-verbialkuisilla kysymyksillä on siis yhtenäinen funktio aiemmin tässä artikkelissa esitettyjen yhdistelmäkysymysten kanssa, eli kysymyksen pehmittäminen: "*Muistaksä mitä te leikitte – –*". Oikeastaan halutaan kysyä vain "*Mitä te leikitte?*", mutta halutaan antaa keskustelukumppanille aito mahdollisuus

sanoa "Emmä muista" ilman, että sillä lailla vastaaminen tuntuisi hänestä epäkohteliaalta tai vaivaannuttavalta.

Olla-verbillä alkavat kysymykset vaikuttavat olevan leimallisesti puhekielen ilmiö. Olisikin varmasti hedelmällistä vertailla, miten vastaava kysymystyyppi edustuu puhutun ja kirjoitetun kielen välimaastossa liikkuvissa kieliympäristöissä, kuten pikaviestimissä ja sosiaalisessa mediassa. Nähdäksemme niissä konteksteissa, joissa kieli välittyy nimenomaan tekstinä kielenkäyttäjältä toiselle, *olla*-verbin ylenpalttinen käyttö ei ole perusteltua, vaan keskustelijat voivat esittää kysymykset suoremmin vailla pelkoa siitä, että toiselta menisi jotakin ohi. Mutta tämä on ehdottomasti kiintoisa jatkotutkimuskohde.

4 LOPUKSI

Olemme tutkineet enimmäkseen haastattelijoiden esittämien kysyvien lausumien eroja ja niiden muutosta Helsingin puhekielen pitkittäiskorpuksessa 1970-, 1990- ja 2010-luvuilla. 6000 kysyvän lausuman aineistomme poimittiin automaattisia työkaluja ja manuaalista työtä yhdistelemällä.

Kun litteraatteja käy läpi manuaalisesti ja ottaa myös vastaajan repliikit huomioon kysymyksen tulkinnassa, partikkeli- ja väitekysymysten osuus nousee väistämättä. Näiden automaattinen poiminta on huomattavasti hankalampaa. Näin ollen pelkästään käsin poimimalla kysymyskategorioiden keskinäinen jakauma voisi olla hyvin erilainen. Tässä tutkimuksessa menetelmien yhdistäminen on mahdollistanut tutkimuksen tekemisen ylipäätään: pelkkä manuaalinen poiminta korpusaineistosta olisi ollut epärealistista toteuttaa. HavaitSIMME, että automaattisten työkalujen toimintaa ja niiden tuottamia tuloksia on kuitenkin hyvä seurata ja tarvittaessa tehdä muutoksia virheiden välttämiseksi.

Huomasimme, että erityisesti 1990-luku erottuu määrällisessä tarkastelussa muista vuosikymmenistä. Hakukysymysten määrä 1990-luvulla oli runsas ja vastaavasti vaihtoehtokysymyksiä – sekä yksinkertaisia että disjunkttiivisia – oli vähemmän

kuin 1970- ja 2010-luvuilla. Myös väitekysymyksiä on selvästi eniten 1990-luvun aineistossa, mutta tämä johtui luultavasti osittain kysymysten runsaasta manuaalisesta poiminnasta. 1970-luvulla korostuivat vaihtoehtokysymykset muihin vuosikymmeniin verrattuna, ja 2010-luvulla taas yhdistelmäkysymyksiä esiintyy muita vuosikymmeniä runsaammin. Kaiken kaikkiaan hakukysymykset ja vaihtoehtokysymykset ovat aineiston runsaslukuisimmat kysymystyypit: yhdessä ne kattoivat noin 85 % kaikista aineiston kysyvistä lausumista.

Havaitsimme tutkimuksessa myös sellaisia kiinnostavia seikkoja aineiston kysymyksistä, jotka eivät suoraan liittyneet kysymysten kategorisointiin, mutta jotka kiinnittivät huomionsi, kuten *olla*-verbi-alkuiset kysyvät lausumat. Niiden osalta näkisimme mielellämme jatkotutkimusta. Lisäksi sanajärjestykseen ja sen syihin voisi pureutua vielä lisää verraten puhuttua puhekieltä, kirjoitettua puhekieltä ja normitettua yleiskieltä.

LÄHTEET

Forsberg, H. (1992). "Onko se minkäläistä?" – kysyvyyden redundanssia puhesuomessa. *Virittäjä*, 96(1), 17–30.

Forsberg, H. (1994). Havaintoja kysymystyypistä otat sie ~ sä. *Virittäjä*, 98(1), 60–68.

Forsberg, H. (2006). *Montako-kysymystyyppin synty ja kehitys*. *Virittäjä*, 110(3), 365–396.

Helsingin yliopisto (2024). *Helsingin puhekielen pitkittäiskorpus -aineisto- ja tutkimushanke*. <https://researchportal.helsinki.fi/fi/projects/the-longitudinal-corpus-of-finnish-spoken-in-helsinki-1970s-1990s>

Hoffmann, T. & Trousdale, G. (2013). Construction Grammar: Introduction. Thomas Hoffmann & Graeme Trousdale (toim.), *The Oxford Handbook of Construction Grammar*. New York: Oxford University Press.

ISK = Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T. R. & Alho, I. (2004). *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura. Verkko-versio. Saatavissa: <http://scripta.kotus.fi/visk>

Jaskara, V. (2009). *Onkos se Jussi nyt missä. Kaksoiskysymyksen rakenne ja funktio keskustelussa*. Pro gradu -tutkielma. Helsingin yliopisto.

Kanerva, J., Ginter, F., Miekka, N., Leino, A. & Salakoski, T. (2018). Turku Neural Parser Pipeline: An End-to-End System for the CoNLL 2018 Shared Task. *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Association for Computational Linguistics, Brussels, Belgium, s. 133–142.

Karhu, A. (1995). Murrehaastattelu vuorovaikutustilanteena. – Marjatta Palander (toim.), *Murteiden matkassa. Juhlakirja Alpo Räisäsen 60-vuotispäiväksi* s. 92–109. Joensuu: Joensuun yliopiston humanistinen tiedekunta.

Kielipankki. (2024). <https://www.kielipankki.fi/aineistot/helpuhe>

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A. & Herbst, E. (2007). *Moses: Open source toolkit for statistical machine translation*. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, s. 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Korpela, E. (2007). *Oireista puhuminen lääkärin vastaanotolla. Keskustelunanalyttinen tutkimus lääkärin kysymyksistä*. Suomalaisen

Kirjallisuuden Seura.

Kuparinen, O., Miletić, A. & Scherrer, Y. (2023). *Dialect-to-Standard Normalization: A Large-Scale Multilingual Evaluation*. Helsingin yliopisto.

Lauranto, Y. (2015). *Direktiivisyyden rajoja. Suomen kielen vaihtokauppasyntaksia*. Suomen kielen, suomalais-ugrialaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos. Helsingin yliopisto.

Olin, K. (1996). Kysymysten rakenteesta heijastuva kuulijaan suhtautuminen. *Kieli 10. Suomalaisen keskustelun keinoja II* s. 47–79. Helsingin yliopiston suomen kielen laitos.

Paldanius, H. & Sandberg, K. (2021). Systemis-funktionaalinen lauseoppi. Kirjainarvio teoksesta Susanna Shore: Lauseita ja vesinokkaeläimiä. Perinteisestä funktionaaliseen lauseoppiin. *Virittäjä*, 125(3), 452–457.

Partanen, N., Alnajjar, K. & Hämäläinen, M. (2019). *Dialect Text Normalization to Normative Standard Finnish*. Helsingin yliopisto.

Raevaara, L. (2006). Kysymykset virkailijan työkaluna. Arjen asiointia. *Keskusteluja Kelan tiskiltä* s. 86–116. Tietolipas 210. Suomalaisen Kirjallisuuden Seura. Toimittaneet Marja-Leena Sorjonen ja Liisa Raevaara.

Searle, J. R. (1969). *Speech acts. An Essay in the Philosophy of Language*. Cambridge University Press.

Suomalainen, K. (2020). Tiedätkö – kysymyksestä kiteymäksi. *Sanajalka*, 62, 33–53.

Vatanen, A. (2021). *Keskustelunanalyysi. Laadullisen tutkimuksen verkkokäsikirja*. Tietoarkisto. Tampereen yliopisto.

<https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvali/teoreettis-metodologiset-viitekehykset/keskustelunalyysi/>

Vilkuna, M. (2014). *SKN-aineiston yleiskielistys*.

https://www.kielipankki.fi/download/SKN/yleiskielistys_skn.pdf

Xue, L., Barua, A., Constant, N., Al-Rfou, R., Narang, S., Kale, M., Roberts, A. & Raffel, C. (2022). ByT5: Towards a token-free future with pre-trained byte-to-byte models. *Transactions of the Association for Computational Linguistics*, 10: 291–306.