

## 2.1 Tutkimusprosessi ja hypoteesin muodostaminen

Seuraavassa tekstissä käydään läpi tutkimuskysymyksen muodostaminen ja ns. nollahypoteesin muodostaminen. Videossa ”Tutkimusprosessi ja hypoteesin muodostaminen” käsitellään näitä koskevat keskeiset asiat.

Tutkimuskysymyksen muodostaminen lähtee liikkeelle tutkimusalueen kartoittamisesta. Aiheesta kerätään tietoa ja selvitetään, mitä siitä on sanottu aiemmin, millaisia tuloksia on saatu, miten luotettavia tulokset ovat ja mitä on jäänyt tutkimatta. Selvittämällä vastauksia mm. näihin kysymyksiin on mahdollista alkaa saada selville, millaisia selvittämättömiä asioita tutkimusalueessa on ja millainen tutkimuskysymys voisi alkaa vastata niihin.

Tutkimuskysymys muodostetaan aina suhteessa tutkimusalaan ja sen tulee olla omalla alalla mielekäs ja teoreettisesti kiinnostava. Voisimme esimerkiksi testata, esiintyykö maailman kielten passiivin merkinnässä yleisemmin foneemi /i/ kuin jokin muu foneemi, mutta näiden ilmiöiden välinen suhde on niin tuulesta temmattu, että vaikka niiden välillä löytyisikin jokin tilastollinen yhteys, se ei voi olla muuta kuin sattumaa. Tämän vuoksi on erittäin tärkeä muistaa, että korrelaatio ei ole kausaatio – eli syyn ja seurauksen suhde – eikä tilastollisesti osoitettu korrelaatio kerro välttämättä mitään kausaatiosta. Hauskoja esimerkkejä tällaisista näennäiskorrelaatiosta tarjoavat mm. Roberts ja Winters (2013; <https://doi.org/10.1371/journal.pone.0070902>).

Kun tutkimustausta on kunnolla hallussa, on aika muodostaa tutkimuskysymys. Se on syytä esittää aidosti kysymysmuodossa eli aidon kysymyslauseen muodossa. Kysymys voi olla sisältökysymys tai kyllä/ei-kysymys. Esimerkki kyllä/ei-kysymyksestä on Antti Arppen (2008 [<https://helda.helsinki.fi/handle/10138/19274>]) väitöskirjaan liittyvä kysymys: vaikuttaako subjektin persoona (1., 2., 3. persoona) siihen, mitä ajatella-verbiä suomen kielessä käytetään (*pohtia, mieltä, ajatella, ...*). Esimerkki sisältökysymyksestä voisi olla: miten pohjoissaamassa käytetään metaforaa tarinoissa? Tai: missä määrin sanajärjestyksen ja sijanmerkinnän välillä esiintyy vuorovaikutusta maailman kielissä?

Tutkimuskysymys muutetaan testattavaksi hypoteesiksi muuttamalla kysymys väitemuotoon. Esimerkiksi Tuomas Huumo

(2010; <https://journal.fi/virittaja/article/view/4371>) tutkii suomen kielen ns. bipositioita, jotka voivat esiintyä joko ennen komplementtia A: *Hän käveli läpi metsän*) tai sen jälkeen B: *Hän käveli metsän läpi*. Tutkimuskysymyksemme voisi olla:

(1) Esiintyykö suomen bipositio *läpi* useammin prepositiona (A) vai postpositiona (B).

Kysymys ilmaisee epävarmuuden siitä, miten usein bipositio *läpi* esiintyy prepositiona tai postpositiona; se ei kuitenkaan ota kantaa siihen, kumpi niistä esiintyy useammin. Tällöin testattavaksi lähtökohdaksi voidaan ottaa, että bipositio *läpi* esiintyy yhtä usein prepositiona ja postpositiona. Tutkimuskysymys voidaan näin olla muuttanut väitemuotoon seuraavasti. Nollahypoteesin yhteydessä on esitetty myös tapa, jolla nollahypoteesi muunnetaan tieteelliseen muotoon.

(2) Suomen bipositio *läpi* esiintyy prepositiona (A) yhtä usein kuin se esiintyy postpositiona (B).

H<sub>0</sub> (nollahypoteesi): frekvenssi (*läpi* prepositiona) = frekvenssi (*läpi* postpositiona)

Kysymyksen muuttamisessa väitemuotoon on syytä olla tarkkana. Hypoteesi on laadittava sellaiseen muotoon, jotta se on mahdollista kumota. Tämä on erittäin tärkeää huomioida, koska tilastollisten testien avulla testataan, kumotaanko nollahypoteesi vai pysyykö se voimassa.

Pohdi esimerkiksi, mikä seuraavassa hypoteesissa on vikana, jos vertaat sitä yllä olevaan väitteeseen?

(3) Suomen bipositio *läpi* voi esiintyä prepositiona (A) yhtä usein kuin se esiintyy postpositiona (B).

Ensimmäisessä tapauksessa vastauksena on kaksi toisensa pois sulkevaa vaihtoehtoa:

(1) suomen bipositio *läpi* esiintyy prepositiona yhtä usein kuin postpositiona ja (2) suomen bipositio *läpi* ei esiinny prepositiona yhtä usein kuin postpositiona. Jälkimmäisessä kohdassa puolestaan on vain yksi vaihtoehto: jos todellisuudessa bipositio *läpi* esiintyy yhtä usein prepositiona ja postpositiona, tämä ei kumoa väitettä; jos todellisuudessa bipositio *läpi* ei esiinny yhtä usein prepositiona kuin postpositiona, tämäkään ei kumoa väitettä. Tämä on vikana sellaisissa hypoteeseissa, joita aineiston avulla ei voi kumota eikä hyväksyä.

Nollahypoteesin lisäksi täytyy muodostaa myös ns. vastahypoteesi. Tilastollisten testien avulla nimittäin pyritään löytämään evidenssiä, jonka perusteella nollahypoteesi kumotaan tai hyväksytään. Jos nollahypoteesi kumotaan, tällöin hyväksytään vastahypoteesi. Vastahypoteeseja voi olla useita. Yksi mahdollinen vastahypoteesi voisi olla:

(4) Suomen bipositio *läpi* ei esiinny prepositiona (A) yhtä usein kuin postpositiona (B).

H1 (vastahypoteesi): frekvenssi (*läpi* prepositiona)  $\neq$  frekvenssi (*läpi* postpositiona)

Toisin sanoen, bipositio *läpi* frekvenssi prepositiona poikkeaa sen frekvenssistä prepositiona. Tämäkään väite ei vielä ota kantaa siihen kysymykseen, esiintyykö bipositio *läpi* useammin prepositiona vai postpositiona vaan ainoastaan siihen, että niiden frekvenssit poikkeavat toisistaan. Tätä kysymystä voitaisiin selvittää muodostamalla uudet vastahypoteesit:

(5) Suomen bipositio *läpi* esiintyy useammin prepositiona (A) kuin postpositiona (B).

H2 (vastahypoteesi): frekvenssi (*läpi* prepositiona)  $>$  frekvenssi (*läpi* postpositiona)

(6) Suomen bipositio *läpi* esiintyy harvemmin prepositiona (A) kuin postpositiona (B).

H3 (vastahypoteesi): frekvenssi (*läpi* prepositiona) < frekvenssi (*läpi* postpositiona)

Kun tutkimuskysymys ja sitä vastaava nollahypoteesi on muodostettu, voimme siirtyä tarkastelemaan aineistoon liittyviä asioita.

Viitteet:

Arppe, Antti 2008. *Univariate, Bivariate, and Multivariate Methods in Corpus-Based Lexicography – a Study of Synonymy*. Väitöskirja, yleinen kielitiede, Helsingin yliopisto. <https://helda.helsinki.fi/handle/10138/19274>

Huumo, Tuomas 2010. Suomen väyläadpositioiden prepositio- ja postpositiokäyttöjen merkityseroista. *Virittäjä* 114(4): 531–561. <https://journal.fi/virittaja/article/view/4371>

Laippala, Veronika & Minna Palander-Collin 2020. Korpusaineistot. Teoksessa Milla Luodonpää-Manni, Markus Hamunen, Reetta Konstenius, Matti Miestamo, Urpo Nikanne & Kaius Sinnemäki (toim.), *Kielentutkimuksen menetelmiä II: Aineistonkeruusta analyysiin*, 460–486. Helsinki: Suomalaisen Kirjallisuuden Seura. <https://doi.org/10.21435/skst.1457>

Roberts Seán & James Winters 2013. Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PLoS ONE* 8(8): e70902. <https://doi.org/10.1371/journal.pone.0070902>