

Distribuutio ja merkitys

ANTTI KANNER

Väitöksenalkajaisitelmä Helsingin yliopistossa 22. maaliskuuta 2022

Vuonna 1980 kielifilosofi John Searle esitteli *kiinalainen huone* -nimisen ajatuskokeen (Searle 1980). Ajatuskokeessa kuvitellaan suljettu huone, johon on kaksi luukku. Huoneessa on henkilö, jonka tehtävänä on ottaa toisesta luukusta vastaan paperille kirjoitettuja kiinankielisiä viestejä, kirjoittaa niihin vastaus ja työntää vastaus paperilapulla toisesta luukusta ulos. Henkilö huoneessa ei osaa kiinaa eikä osaa lukea sen kirjoitusmerkkejä. Sen sijaan hänellä on hyvin yksityiskohtaiset ohjeet, joiden avulla hänen tulee vastauksensa muodostaa. Ohjeet kuitenkin koskevat vain merkkien muotoa, eikä niissä ole kuvattu yhdenkään kiinan kielen sanan merkitystä. Henkilö ei siten ymmärrä sanaakaan viesteistä, joita hänelle lähetetään, eikä hän ymmärrä vastauksia, joita niihin antaa. Tästä huolimatta hän oppii tunnistamaan merkkejä ja lukemaan ohjeita niin hyvin, että henkilöt huoneen ulkopuolella erehtyvät luulemaan keskustelevansa kiinaa taitavan henkilön kanssa.

Kiinalainen huone oli puheenvuoro aikakautensa tekoälyjä koskevaan keskusteluun. Tuolloin tietojenkäsittelytieteessä oli jo vakiintunut lähtökohta, jossa filosofisesti tyydyttävästi älykkään mekanismin sijaan oli alettu tavoitella mekanismeja, jotka erehdyttävästi kykenivät jäljittelemään älyllistä toimintaa. Eikö koneellista luovuutta voisi tutkia tuottamalla mekanismi, joka toimisi aivan kuten luovana pidetty ihminen toimisi ja näin hyväksyä se luovuutena ilman keskustelua siitä, mitä luovuus tarkkaan ottaen on? Searle kohdisti huomionsa kykyyn ymmärtää kieltä tällaisena älyllisenä toimintana ja kysyi, voidaanko kiinalaisen huoneen henkilön sanoa osaavan kiinaa? Searlen vastaus oli ei, ja siihen onkin helppo yhtyä.

Nykyisin, kun erilaiset koneoppimiseen perustuvat laskennalliset menetelmät tekevät tuloaan toisaalta osaksi arkipäiväistä kielellistä maisemaa erilaisten käyttöliittymien, chat-bottien, hakukoneiden ja automaattisten käännösohjelmien muodossa, toisaalta tutkimuksen työvälineiksi niin kielitieteen kuin humanistisen tutkimuksen kentällä laajemminkin, on kiinalainen huone edelleen relevantti asettaessamme teknologisia edistysaskeleita raameihinsa. Monet modernit automaattiset luonnollisen

kielen käsittelyyn suunnitellut algoritmit nimittäin toimivat juuri kiinalaisen huoneen henkilön tavoin. Ne osaavat hyödyntää käsittelemiensä tekstien tilastollisia säännönmukaisuuksia ja kykenevät niiden avulla tarkkaan rajatuissa tilanteissa käyttäytymään usein niin kuin kieltä osaava henkilö käyttäytyisi.

Kiinalaisen huoneen kysymys–vastaus-dialogin sijaan näitä algoritmeja voidaan hyödyntää monenlaisissa tehtävissä. Kuitenkin niillä on myös sama tärkeä rajoite kuin henkilöllä kiinalaisessa huoneessa: ne eivät ymmärrä käsittelemiensä tekstien kommunikatiivisia pyrkimyksiä, sillä niiden toiminta perustuu vain formaalien elementtien ja sekvenssien lajitteluun. Kutsun tällaista automatisoitavissa olevaa lajittelua ja luokittelua sekä siihen pohjaavaa mallintamista *distributionaaliseksi operaatioksi*, sillä kielitieteellisestä näkökulmasta sellaisen toiminta nojaa olennaisesti distribuution käsitteeseen. Distributionaalilla operaatioilla on myös oma metodologinen taustansa kielitieteen historiassa.

Kielitieteellisenä terminä, erotuksena matemaattisesta termistä, distribuutiolla tarkoitetaan jonkin kielellisen rakenteen tai ilmaisun mahdollista käyttöalaa. Niinpä kun kiinalaisessa huoneessa työskentelevä henkilö luokittelee havaintonsa tietyn ilmauksen käytöstä, kielitieteelliset näkökulmasta hän tulee kartoittaneeksi sen distribuutiota. Saman tekee automaattinen algoritmi, joka rakentaa tekstimassan pohjalta tilastollisen mallin samaisesta ilmauksesta.

Distribuutioiden luonne kielen kuvauksessa on kuitenkin hyvin teoriasidonnainen asia. Näkemys siitä, mikä kielessä on mahdollista ja mikä mahdotonta ja mikä mahdollisuuksia säätelee, on yleisesti ottaen teoreettisen kielikäsitteilyn tärkeimpiä rakennuspalikoita. Väitöskirjani keskeinen kysymys onkin se, miten kielellisten ilmausten merkitykset vaikuttavat niiden distribuutioihin. Distributionaalisten operaatioiden ja kiinalaisen huoneen relevanssi työlle on siinä, että lähestyn tätä kysymystä johtamalla sen pohjalta toisen, käytännöllisemmän kysymyksen, joka koskee sitä, mitä merkitykseen liittyviä ilmiöitä distributionaalisen operaation tuottamasta kuvauksesta voidaan löytää. Tässä kysymyksenasettelussa sellaisenaan on kuitenkin edelleen avoinna kolme olennaista seikkaa: mitä tarkoitetaan distribuutiolla, minkälaisesta distributionaalista operaatiosta on kyse ja mistä merkitykseen liittyvistä ilmiöistä puhutaan.

Olen valinnut pitää käsittelyni sellaisella tasolla, ettei sen fokus rajautuisi liiaksi koskemaan vain tiettyjä valintoja näillä kolmella akselilla, vaan tutkimukseni säilyttäisi relevanssinsa riippumatta siitä, mistä teoreettisesta kulmasta asiaa lähestytään, minkälaista operaatiota käytetään tai minkälaisista merkityksiin liittyvistä ilmiöistä ollaan kiinnostuneita. Olen pyrkinyt kuvaamaan distribuution käsitteen kahdessa hallitsevassa kielitieteellisessä paradigmassa, strukturalistisessa ja käyttöpohjaisessa, ja pyrkinyt kartoittamaan distributionaalisia operaatioita koskevaa keskustelua kummankin sisällä sekä vertaamaan näitä kahta distribuutiokäsitystä toisiinsa.

Olen tehnyt valinnan tutkia kahdessa osatutkimuksessani kahta eri distributionaalista operaatiota ja kahta erilaista merkitykseen liittyvää ilmiötä. Ensimmäisessä osatutkimuksessa lähdin liikkeelle tietystä operaatiosta ja pyrin kuvaamaan merkitykseen liittyviä ilmiöitä, joita sen tuottamat representaatiot toistavat. Tutkittavaksi operaatioksi valikoitui taaiaan käytetty menetelmätyyppi, niin kutsuttu vektoriarvumalli (ks. esim. Turney & Pantel 2010; Mikolov, Chen, Corrado & Dean 2013). Valitsin

tutkimuskohteikseni kolme vastakohtaista adjektiiviparia, mutta operaatiota itsessään en suunnitellut tavoittamaan erityisesti adjektiivien vastakohtaisuutta. Aineistona tässä osatutkimuksessa toimi Suomi24-internetkeskustelufoorumi.

Toisessa osatutkimuksessa lähdin liikkeelle tietystä merkitykseen liittyvästä rakenteesta, tiettyyn rekisteriin liittyvästä assosiaatiosta, ja pyrin kehittämään distributionaalisen operaation, joka tuottaisi tähän soveltuvan mallin. Yleiseltä arkkitehtoniselta malliltaan tämä operaatio kuitenkin suuresti muistuttaa niin kutsuttua aihemallinnusta (ks. esim. Blei, Ng & Jordan 2003). Tutkittavana kielen ilmiönä tässä osatutkimuksessa oli usein substantiivisesti käytetty adjektiivi *vaivainen*, jonka asemaa juridisessa ja hallinnollisessa rekisterissä pyrin kartoittamaan 1800-luvun sanomalehtiaineistossa.

Havaintojeni mukaan distributionaaliset operaatiot tuottavat representaatioita, joissa merkitykseen liittyvät rakenteet ovat aina tavalla tai toisella läsnä. Yleisellä tasolla merkityksen ja distribuutioiden hyvin laaja-alainen yhteenliittyvyys ei kielen tutkijaa yllätä, mutta yksityiskohtainen kuva, jonka distributionaaliset operaatiot kykenevät tarjoamaan, on yhtä kaikki mielenkiintoinen. Silloinkin kun distributionaaliset representaatiot näyttävät nostavan esiin tilastollista rakenteisuutta, joka ei sinällään vaikuta liittyvän millään ilmeisellä tavalla merkityksiin, lähempi tarkastelu usein osoittaa, että kyse on vain merkityksen lajista, jota kontekstista irrallaan on vaikea tulla ajatelleeksi. Näin usein toistettu hypoteesi distributionaalisen ja semanttisen samankaltaisuuden korrelaatiosta on samaan aikaan sekä tosi että yleistyksenä lähes hyödyttömän karkea: erilaisilla distributionaalisilla operaatioilla saadaan esiin erilaista merkityksen samankaltaisuutta ja kahden eri kielellisen ilmauksen merkityksen samankaltaisuutta voidaan arvioida eri näkökulmista.

Lisäksi yksinkertaisilla operationaalisilla valinnoilla oli mahdollista muuttaa representaatiot vastaamaan joko strukturalistista tai käyttöpohjaista distribuutiokäsitystä. Tällöin merkitykseen liittyvät ilmiöt, joita näistä representaatioista oli tulkittavissa, olivat sopusoinnussa niiden merkitykseen liittyvien ilmiöiden kanssa, joita kohti näissä paradigmoissa on totuttu suuntautumaan.

Strukturalistisessa kielikäsitelyssä kieli perustuu kullekin luonnolliselle kielelle ominaiseen sääntöjen joukkoon, jotka säätelevät, mitkä ilmaukset ovat kielenmukaisia ja mitkä kielenvastaisia. Yleisesti ottaen distribuutiot ovat näiden kielenmukaisten ilmausten joukkoja. Täydestä kielellisten ilmausten kombinatorisesta avaruudesta hahmottuvat siis osajoukkona kieliopilliset ja merkitykselliset ilmaukset, ja kunkin kielellisen ilmauksen distribuutio on sen omien mahdollisten kombinaatioiden ja kaikkien kielenmukaisten ilmausten leikkaus.

Strukturalistisesta näkökulmasta distribuutioiden suhde kielen normeihin onkin samankaltainen kuin tieverkoston suhde liikennesääntöihin; kunkin kielen rakenteen distribuutio ovat ne reitit, joita sen on mahdollista liikennesääntöjen mukaan tieverkon sisällä kulkea. Ajoneuvon tyyppi, se, onko kyseessä kuorma-auto, henkilöauto vai polkupyörä, vaikuttaa mahdollisiin reitteihin, ajoneuvon väri yleensä ei. Samalla tavoin merkitys heijastuu distribuutioihin niiltä osin kuin se sääntelee, minkälaisissa konteksteissa jokin tietty kielellinen ilmaus voi esiintyä. Kun pohditaan kysymystä merkityksen ja distribuution suhteesta, voidaan kysyä, mitkä osat kielellisen ilmauksen merkityksestä vertautuvat ajoneuvon tyyppiin, mitkä sen väriin.

Käyttöpohjainen näkökulma distribuutioon sen sijaan lähtee oletuksesta, että binääristen normien sijaan kieltä järjestelmänä pitävät paikoillaan kielenkäyttäjien yhtenevät kokemukset siitä, miten kieltä käytetään. Käyttöpohjaisessa paradigmassa kielellisten ilmausten yleisyys toimiikin ankkurina, johon nähden harvinaisempien ilmausten kielenmukaisuutta tilanteisesti arvioidaan. Näin jyrkkä raja kielenmukaisen ja kielenvastaisen välillä liudentuu stokastiseksi todennäköisyydeksi.

Kun aiemmin käytettyä liikennesääntövertausta sovelletaan käyttöpohjaiseen kielikäsitteeseen, havainnollistuu, miten epätodennäköiseltä tuntuu, että mikään yksittäinen ajoneuvo koskaan ajaisi läpi kaikkia sille mahdollisia reittejä. Sen sijaan ajoneuvot käyttöesineinä toteuttavat funktionaalisia tarkoituksiperiään ja niiden kulkemat reitit heijastelevat niitä tarpeita, joihin niitä yleensä käytetään. Kielellisten ilmausten distribuutiot keskittyvät tavanomaisimpien käyttötarkoitusten ympärille samalla tavoin kuin esimerkiksi kuorma-autojen reitit keskittyvät tavarakuljetusten logistiikan ja henkilöautojen reitit puolestaan työ- ja tarhamatkojen perusteella tiukemmin kuin mitä pelkästään liikennesääntöjen perusteella olisi odotettavissa.

Kun vektoriavaruusmalli säädettiin strukturalistisen distribuutiokäsityksen mukaiseksi käyttämällä niin kutsuttua Jaccard-etäisyyttä etäisyysmittana, analyysi osoitti, että esimerkiksi adjektiiviparit *sairas* ja *terve* ja *vanha* ja *nuori* saavat tavanomaisesti pääsanoikseen henkilöviitteisten substantiivien lisäksi myös substantiiveja, jotka ovat osa-kokonaisuussuhteessa henkilöihin. Tällaisia pääsanoja ovat erilaiset sekä konkreettiset että metaforiset ruumiinosia ja psykologisia taipumuksia ilmaisevat substantiivit, kuten *nuori sydän* tai *sairas huomiontarve*. *Rikas* ja *köyhä* eivät tällaisia pääsanoja tavallisesti saa. Tämä liittyy siihen miten *sairas*, *terve*, *vanha* ja *nuori* tuottavat henkilöviitteisen pääsanansa kanssa konkreettisemmän tarkoitteen kuin *rikas* tai *köyhä*. Näin mitattuna *sairaana*, *terveen*, *vanhan* ja *nuoren* kanssa distributionaalisesti samankaltaisimpien sanojen joukossa oli runsaasti sanoja, joilla oli tämä sama ominaisuus. Se, että adjektiivin kuvaama ominaisuus etenee pääsanana tarkoiteen osa-kokonaisuushierarkiassa, on samankaltainen merkityksen piirre kuin esimerkiksi suhteellisuuden ja asteittaisuuteen liittyvät adjektiiviluokitukset, jotka strukturalistisessa perinteessä ovat olleet adjektiivien kuvauksessa keskeisiä.

Toisaalta kun etäisyysmittana käytettiin dimensioiden frekvenssit huomioivaa kosinietäisyyttä, esimerkiksi *sairaana* läheisimmät naapurit olivat adjektiiveja ja substantiiveja, joissa toistuivat käsitteellisen samankaltaisuuden sijaan sellaiset asiat kuten taaja affektiivinen käyttö loukkauksissa. Näin *hullu* ja *idiootti* ovat *sairaalle* läheisempiä kuin sen lähisynonymi *kipeä*. Käyttöpohjaisesta perspektiivistä on selvää, miksi näin tapahtuu: affektiivisuuteen liittyvät konnotatiiviset merkitykset, jotka perustuvat tietynlaisiin, aineistossa yleisiin kielenkäytön tapoihin, ovat käyttöpohjaisessa paradigmassa keskeisempi osa sanojen merkitystä kuin strukturalistisessa.

Sen lisäksi, että operationaalisilla valinnoilla distributionaaliset representaatiot voidaan saada vastaamaan tiettyä teoreettista distribuutiokäsitystä, tilanne on mahdollista nähdä myös toisinpäin: distributionaalisille representaatiolle voidaan löytää tulkinta, kun tunnustetaan kielitieteellinen teoria, jonka muodon kanssa ne ovat yhteneviä. Näin distributionaalinen analyysi ei siis mekanistisesta ja näennäisen objektiivisesta luonteestaan huolimatta voi myöskään toimia empiirisenä pohjana arvioitaessa struktura-

listista ja käyttöpohjaista paradigmaa vastakkain, sillä analyysiin on välttämättä jo leivottu sisään jommankumman teoreettiset lähtöoletukset.

Olenaisena pidän myös havaintoa siitä, että välittämättä siitä, kumman distribuutiokäsityksen mukaiseksi operaatio suunnataan, distributionaaliset representaatiot näyttävät sisältävän vain niukasti tietoa sanojen denotatiivisten merkitysten prototyypisimmistä piirteistä. Kuten tietokone-lingvisti Katrin Erk (2016) on huomauttanut, *alligator*-sanan distributionaaliset representaatiot onnistuvat identifioimaan alligaattorin muista eläimistä akvaattisen elämäntavan ja käsilaukkuihin liittyvien vihjeiden perusteella, mutta siihen niiden tarjoama tieto alligaattoreista tyypillisesti jää. Myös omien havaintojeni perusteella yleisen tason distributionaalisen samankaltaisuuden ja merkityksen samankaltaisuuden välisen yhteyden ylläpitoon näyttävätkin riittävän epäsuorat ja summittaiset vihjeet. Implikaatio tästä on se, että pelkästään se, että distributionaalinen operaatio kykenee hyvin arvioimaan kahden sanan merkitysten välisen samankaltaisuuden, ei osoita, että mikä tahansa tietty sanoja yhdistävä tai erottava merkityksen piirre olisi mukana tässä samankaltaisuudessa.

Tutkimuksen työvälineenä erilaiset distributionaaliset operaatiot lienevät tulevaisuudessa korvaamattomia, mutta niiden tuloksista tulee ymmärrettäviä ainoastaan suhteessa tunnistettuihin teoreettisiin lähtöoletuksiin siitä, minkälaiset tekijät hallitsevat distribuutioiden muodostumista. Yhtä lailla on hylättävä ajatus siitä, että distributionaalisten operaatioiden tuloksena syntyvät representaatiot ja mallit voisivat sellaisenaan olla analogisia erilaisille merkityksen ilmiöille. Sen sijaan ne tulee nähdä kuvauksina distribuutioista, ja distribuutiot puolestaan nähdä kielitieteellisinä tutkimuskohteina keskeisimpien joukossa.

Lähteet

- BLEI, DAVID M. – NG, ANDREW Y. – JORDAN, MICHAEL I. 2003: Latent dirichlet allocation. – *Journal of Machine Learning Research* 3 s. 993–1022.
- ERK, KATRIN 2016: What do you know about an alligator when you know the company it keeps. – *Semantics and Pragmatics* 9 s. 17:1–63. <https://doi.org/10.3765/sp.9.17>.
- MIKOLOV, TOMAS – CHEN, KAI – CORRADO, GREG – DEAN, JEFFREY 2013: Efficient estimation of word representations in vector space. – *CoRR*. <https://doi.org/10.48550/arXiv.1301.3781>.
- SEARLE, JOHN 1980: Minds, Brains and Programs. – *Behavioral and Brain Sciences* 3 (3) s. 417–424. <https://doi.org/10.1017/S0140525X00005756>.
- TURNEY, PETER – PANTEL, PATRICK 2010: From frequency to meaning. Vector space models of semantics. – *Journal of Artificial Intelligence Research* 37 s. 141–188. <https://doi.org/10.1613/jair.2934>.

Antti Kanner: *Meaning in distributions. A study on computational methods in lexical semantics*. Helsinki: Helsingin yliopisto 2022.

Kirjoittajan yhteystiedot:
etunimi.sukunimi@helsinki.fi