

Kognitiivinen mallintaminen neuroverkoilla

Krista Lagus ja Timo Honkela

Teknillinen korkeakoulu
Adaptiivisen informatiikan tutkimusyksikkö

Versio 0.4b
23.4.2006

1. Johdanto

1.1. Kognitiivinen mallintaminen

Kognitiotieteessä usein lähdetään olettamuksesta, että mielessä on erilaisia representaatioita sekä laskennallisia prosesseja, jotka operoivat näillä representaatioilla. Tämä oletamus soveltuu yhtä lailla oletettiinpa, että representaatiot muistuttavat tietokoneiden tietorakenteita, formaalin logiikan propositioita, sääntöjä, käsitteitä, kuvia, analogioita tai piirre- ja vektoriavaruuksia¹. Tutkimuksen kohteina ovat tyypillisesti olleet ihmisen tiedonkäsittelyprosessit mukaanlukien havaintojen tekeminen, muisti, kieli, päättely ja ongelmanratkaisu ja oppiminen. Emootioiden mallintamisen tutkimus taas on jäänyt perinteisesti varsin vähälle huomiolle. Kognitiivinen mallintaminen käy vuoropuhelua käyttäytymisen tutkimuksen, aivotutkimuksen, ja tekoälyn tutkimuksen välimaastossa.

Mallinnuksella voidaan katsoa olevan kolme keskeistä kriteeriä: ennustuskyky, falsifioitavuus ja yksinkertaisuus.

Mallin ennustuskyvyllä tarkoitetaan sitä, että mallin avulla voidaan menestykkäästi ennustaa tai ennakoida sellaisia tapauksia tai osailmiöitä, joita koskevia havaintoja ei ole ollut käytettävissä mallia rakennettaessa.

Falsifioitavuudella taas tarkoitetaan sitä, että malli synnyttää niin yksityiskohtaisia hypoteeseja tai ennusteita koskien mallinnuksen kohteen luonnetta, että tutkittaessa itse kohdetta malli voi periaatteessa osoittautua myös vääräksi tai puutteelliseksi.

Yksinkertaisuusvaatimuksen taustalla taas voidaan nähdä tunnettu Occamin partaveitsi -periaate: mikäli kaksi teoriaa selittää saman ilmiön yhtä hyvin, on teorioista valittava yksinkertaisempi. Laskennallista mallia voidaan tässä pitää teorian konkreettisena, toimivana esitystapana taikka teorian toteuttavana instanssina. Käytännössä mallien yksinkertaisuutta voidaan luonnehtia esimerkiksi vapaiden parametrien määrällä (esim. neuroverkon painojen määrä) ja mallin viemien resurssien koolla (tietokoneen muistitila, laskenta-aika ja operaatioiden määrä).

Kognitiivisten järjestelmien kompleksisuutta voidaan lähestyä eri tavoin. Yksi mahdollisuus on jakaa järjestelmä erillisiin osiin, joita usein kutsutaan moduleiksi. Ajatuksena on, että kunkin modulin toimintaa ja rakennetta voidaan tarkastella erikseen.

1(ks. <http://plato.stanford.edu/entries/cognitive-science/>)

Lähdettäessä mallintamaan jotakin kompleksisen kokonaisuuden yksittäistä osaa enemmän tai vähemmän irrallaan kognitiosta kokonaisuutena (esim. kielen syntaktisten säännönmukaisuuksien oppiminen) joudutaan kuitenkin tekemään hyvin yksityiskohtaisiakin oletuksia sekä prosessin saamien syötteiden että sen tuottamien ulostulojen luonteesta. Johtuen ongelman monimutkaisuudesta kognitiota ei siten pääsääntöisesti yritetä mallintaa kerralla kokonaisuutena, vaikka tämänkaltaiset lähestymistavat ovatkin yleistymässä, ja aiheesta järjestetään jopa kilpailuja (esim. robottien jalkapallocup).

Modulaarisuuden sijasta kognitiivisen järjestelmän mallintamista on mahdollista lähestyä ”alhaalta ylöspäin” ns. bottom-up –strategialla. Siinä oletetaan, että järjestelmä saa jossakin mielessä saman syöteinformaation, jota ihminenkin saa. Tämän syöteinformaation perusteella pyritään rakentamaan kognition peruspilareita, koska tällöin mallin saaman syötteen osalta joudutaan tekemään vähiten mahdollisesti myöhemmin vääräksi osoittautuvia olettamuksia. Tässä lähestymistavassa oppimisen periaatteiden mallintaminen on keskeinen tehtävä.

Kognitiivisessa mallintamisessa on kyseessä ihmisen yleisten informaationkäsittelyn periaatteiden kartoittaminen. Mallintamisessa voidaan tällöin painottaa enemmänkin yleisten mallityyppien tai oppimisperiaatteiden löytämistä kuin jonkin tietyn osaongelman representaatioiden yksityiskohtia, jotka vieläpä voivat eri ihmisillä vaihdella. Informaationkäsittelyn tai oppimisen *perusperiaatteen* ominaisuus on, että sitä voidaan soveltaa kognition eri osa-alueilla kuten esimerkiksi näköjärjestelmän, päättelyn ja ongelmanratkaisun tai kielen tulkinnan mallinnuksessa. Perusperiaatteina voidaan pitää kutakin erilaista, yleiskäyttöistä neuroverkkomenetelmää.

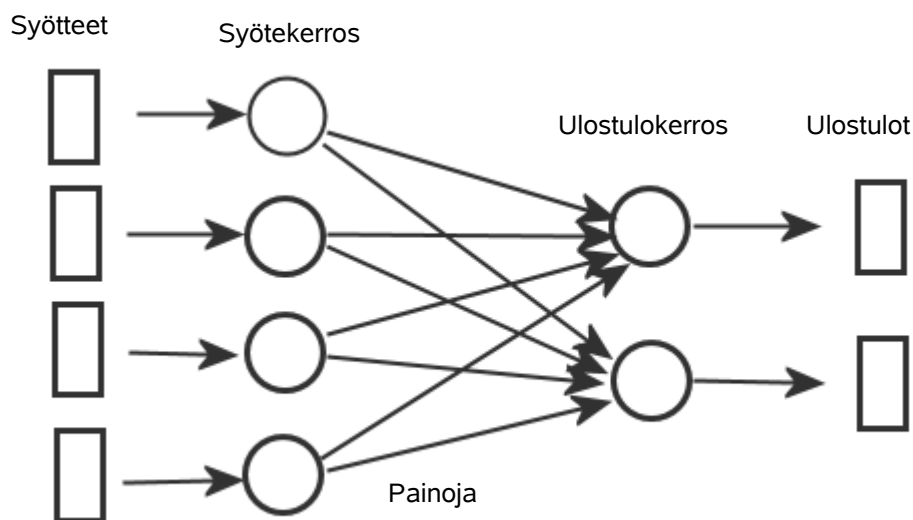
1.2. Keinotekoisien neuroverkkojen taustaa

Keinotekoiset neuroverkot ovat menetelmiä, jotka ovat saaneet inspiraationsa aivojen hermoverkoista. Neuroverkko koostuu tyypillisesti joukosta yksinkertaisia prosessointielementtejä, neuroneita, sekä näiden välisistä kytkennöistä ja kytkentöjen painoista. Humanistisissa tieteissä neuroverkkojen tutkimuksesta on usein käytetty nimeä konnektionismi. Näiden käsitteiden käyttöala vaihtelee jonkin verran kirjoittajasta riippuen, mutta niitä voidaan käytännössä pitää toistensa synonyymeinä.

Neuroverkkomallinnukseen perustuva näkemys kognitiosta ylipäänsä, kielikyky mukaanlukien, korostaa staattisten mallien sijaan oppimisen merkitystä sekä modulaarisen ja sekventiaalisen sijaan yhtäaikaisesti eri puolilla aivoja tapahtuvaa hajautettua tiedon käsittelyä (lisää aiheesta esim. Rumelhart ja McClelland 1986a ja b). Viime vuosikymmeninä tehdyn neuropsykologisen ja aivokuvantamiseen perustuvan tutkimuksen ansiosta tiedetään, että moneen aivoprosessiin osallistuu useita hajautetusti aivoissa sijaitsevia neuraalisia alueita, jotka voivat aktivoitua joko yhtä aikaa tai peräkkäin (Churchland ja Sejnowsky 1992). Yksittäisen homogeenisesti kytketyn neuroverkon ja siinä tapahtuvan oppimisen voitaisiin katsoa mallintavan esimerkiksi yksittäistä neuraalista aivoaluetta. Laajemman ja monimutkaisemman kokonaisuongelman ratkaisemiseen (esim. puheen ymmärtäminen) tarvittaisiin tyypillisesti useita toisiinsa kytkettyjä neuroverkkoja, joista osa voi keskittyä vaikkapa piirreanalyysiin. Verkot ovat siis yksinkertaistettuja malleja hyvin monimutkaisen, osin tuntemattoman inhimillisen järjestelmän jostakin osasta.

Tietävästi ensimmäisen neuroverkkomallin esittelivät vuonna 1943 insinöörit Warren McCulloch ja Walter Pitts. Vuonna 1949 puolestaan psykologi Donald Hebb esitti

kirjassaan “The Organization of Behavior”, että kahden neuronin kytkennän voimakkuus kehittyy sen perusteella, kuinka paljon tai vähän ne ovat aktiivisia yht’ aikaisesti. Toisin sanoin voidaan todeta, että tällaisen nk. Hebb’in lain mukaan kahden neuronin välinen kytkentä vahvistuu, kun ne aktivoituvat toistuvasti yhtä aikaa. Tätä periaatetta muunnelmiseen on hyödynnetty monissa neuroverkkomalleissa. Frank Rosenblatt esitteli vuonna 1957 klassisen yksittäisen neuronin mallin eli perseptronin. Se suorittaa sisääntulolleen yksinkertaisen laskutoimituksen. Siinä lasketaan neuronin ulostulo, kun tiedetään sen saamat syötteet, synaptisten kytkentöjen painokertoimet, ja kynnsarvo. Neuronin ulostulossa käytetään tyypillisesti epälineaarista aktivaatiofunktiota, joka rajoittaa neuronin ulostulon jollekin arvoalueelle. Kuvassa 1 on esitetty Rosenblattin perseptroniverkkomalli. Perseptroni kykenee oppimaan minkä tahansa lineaarisesti separoituvan erottelun, mutta ei esimerkiksi XOR-ongelmaa. Myöhemmin kehitetty monikerrosverkko (MLP eli Multi-Layered Perceptron) puolestaan ratkaisee oppimisen ongelman yleisessä tapauksessa.



Kuva 1: Kaksikerroksinen perseptroni. Syötekerros on täydellisesti kytketty ulostulokerrokseen, mutta kummankaan kerroksen sisällä neuronit eivät ole kytkettyjä keskenään.

Keinotekkoisten neuroverkkojen tutkimuksen pioneereja 1960- ja 1970-luvuilla olivat Sun-ichi Amari, Teuvo Kohonen ja Stephen Grossberg. Kohonen julkaisi vuonna 1978 kirjan assosiativisesta muistista ja 1981 ensimmäisen artikkelin, joka koski itseorganisoivan kartan (Self-Organizing Map, SOM) periaatetta. Käänte alalla tapahtui 1980-luvulla Rumelhartin ja McClellandin julkaistua PDP (Parallel Distributed Processing) -kirjansa. Sen aiheena oli laajasti kognition mallinnus neuroverkoilla. He mm. osoittivat, että ohjattua oppimista toteuttava neuroverkko nimeltä monikerrosverkko (multi-layer perceptron, MLP) kykenee oppimaan esimerkeistä englannin verbien taivutukset menneeseen aikamuotoon. Neuroverkoissa ”säännöt” kuvautuvat implisiittisesti: säännönkaltainen toiminta seuraa neuronien välisten kytkentöjen muodostaman verkoston yhteisvaikutuksena.

2. Oppiminen neuroverkoissa

2.1. Neuroverkon toiminta ja erilaiset oppimistilanteet

Termillä neuroverkko eli keinotekoinen hermoverkko viitataan joukkoon melko erilaisiakin malleja, joilla kullakin on omanlaisensa verkkorakenne, aktivaatioperiaate ja oppimisperiaate. Yhteistä eri malleille on että neuroverkko koostuu yleensä joukosta toisiinsa kytkettyjä neuroneita ja kytkentöihin liittyviä painoja.

Neuroverkkojen oppimisperiaatteet, tai pikemminkin oppimistilanteet, voidaan jakaa kolmeen osajoukkoon: ohjattuun ja ohjaamattomaan oppimiseen sekä vahvisteoppimiseen. Jatkossa nämä on selitetty tarkemmin. Oppimisperiaatteet ovat toisiaan täydentäviä: on mahdollista ja luultavaakin, että eri oppimisperiaatteet soveltuvat kognition eri osien mallinnukseen.

Matemaattisesti neuroverkkojen oppiminen tapahtuu usein soveltamalla ns. gradienttiperiaatetta. Tällöin neuroverkon toimintaperiaate kuvataan valitun kustannusfunktion minimointina, jolloin kukin oppimisaskel otetaan kustannusfunktion gradientin suuntaan. Gradientti on matemaattinen käsite, jota voidaan luonnehtia seuraavanlaisella esimerkillä. Ajatellaan, että vesipisara tipahtaa epätasaiselle pinnalle. Pisara lähtee valumaan pintaa alas siihen suuntaan, missä alamäki on jyrkin. Tätä suuntaa voidaan kutsua gradientiksi.

Neuroverkkomalleja ohjelmoidaan useimmiten perinteisillä tietokoneilla. Myös erityisiä useiden satojen tai tuhansien prosessorien rinnakkaislaskennan mahdollistavia tietokoneita on kehitetty tarkoitusta varten nopeuttamaan laskentaa, mutta niiden käyttö ei ole kovin yleistä.

Nykyisellään neuroverkkoja sovelletaan kognition mallinnusta laajemmin ongelmiin sovelluksissa, joilla ei välttämättä ole mitään tekemistä ihmisen tai aivojen toiminnan mallintamisen kanssa. Sovellusten ja menetelmätieteiden näkökulmasta ei ole keskeistä, kuinka hyvin verkko mallintaa aivojen toimintaa. Tästä syystä rajanveto neuroverkkomenetelmien ja muiden matemaattis-tilastollisten menetelmien välillä onkin ajoittain vaikeaa tai keinotekoista. Yleisemmin siksi puhutaan ns. oppivista tai adaptiivisista menetelmistä ottamatta suoraan kantaa siihen, nähdäänkö menetelmällä edes abstraktin tason yhteyttä aivojen toimintaan.

2.2. Ohjattu oppiminen, monikerrosverkko, ja vastavirta-algoritmi

Ohjatussa oppimisessa neuroverkko opetetaan joukolla syöte-vaste -esimerkkejä, jolloin oppimisen tuloksena syntyy tietynlainen yleistys siitä, miten tietyn tyyppiseen syötteeseen tulee vastata. Yleistys tarkoittaa, että verkko vastaa järkevästi myös sellaisiin syötteisiin, joita ei löydy opetusaineistosta. Yleisin ohjattua oppimista toteuttava opetusalgoritmi on nk. vastavirta-algoritmi (backpropagation), jolla lasketaan monikerrosverkon (MLP) neuronien välisten kytkentöjen painoja.

Monikerrosverkko muistuttaa Perseptronia, mutta siinä syöte- ja ulostulokerroksen välissä voi olla yksi tai useampia välikerroksia. Näitä kutsutaan myös kätkökerroksiksi (hidden layers). Neuronien määrä kaikissa verkon kerroksissa voidaan valita halutun kaltaiseksi. Kätkökerrosten merkitys neuroverkon representointikyvyssä on keskeinen: monikerrosverkon osalta on voitu todistaa että se omaa ns. yleisen funktioaprosimaattorin ominaisuuden. Tämä tarkoittaa että verkko kykenee mallintamaan minkä tahansa syötteiden ja haluttujen ulostulojen välillä olevan lineaarisen tai epälineaarisen riippuvuuden (funktion) mielivaltaisen tarkasti, ja että tämä

ratkaisu on opetusalgoritmilla löydettävissä. Mielivaltainen tarkkuus edellyttää, että välikerroksen neuroneja voidaan lisätä tarvittava määrä.

Eräs ohjatun oppimisen tyypillisesti kohtaama kritiikki koskien sen soveltuvuutta kognition mallintamiseen kokonaisuutena on, että läheskään aina ihminen ei saa suoranaista ”oikeaa vastausta”, jota ohjatun oppimisen menetelmät edellyttävät oppiakseen.

2.3 Vahvisteoppiminen

Vahvisteoppimisessa oppija, jota tyypillisesti kutsutaan agentiksi, havainnoi ja toimii jossakin ympäristössä. Tehtyään havaintoja ympäristön tämänhetkisestä tilasta, agentti suorittaa siinä jonkin teon. Tekonsa jälkeen agentti saa yksiulotteisen vahvistesignaalin eli palkkion. Tämän jälkeen agentti oppii eli tallentaa informaatiota koskien ympäristöä, omaa tekoaan ja vahvistesignaalia sekä saattaa muuttaa päätöksentekostrategiaansa (decision policy). Agentin oppiminen tähtää sen saaman kokonaispalkkion maksimointiin.

Vahvisteoppimistilanne muistuttaa kokonaisten eliöiden ja niiden ympäristön välistä vuorovaikutussuhdetta paremmin kuin ohjatun oppimisen tilanne. Agentin kokonaisarkkitehtuurissa neuroverkkoja voidaan käyttää erilaisissa osatehtävissä. Kattavan ja yleistajuisen esittelyn aiheeseen tarjoaa kirja ”Reinforcement learning” (Sutton & Barto, 1998).

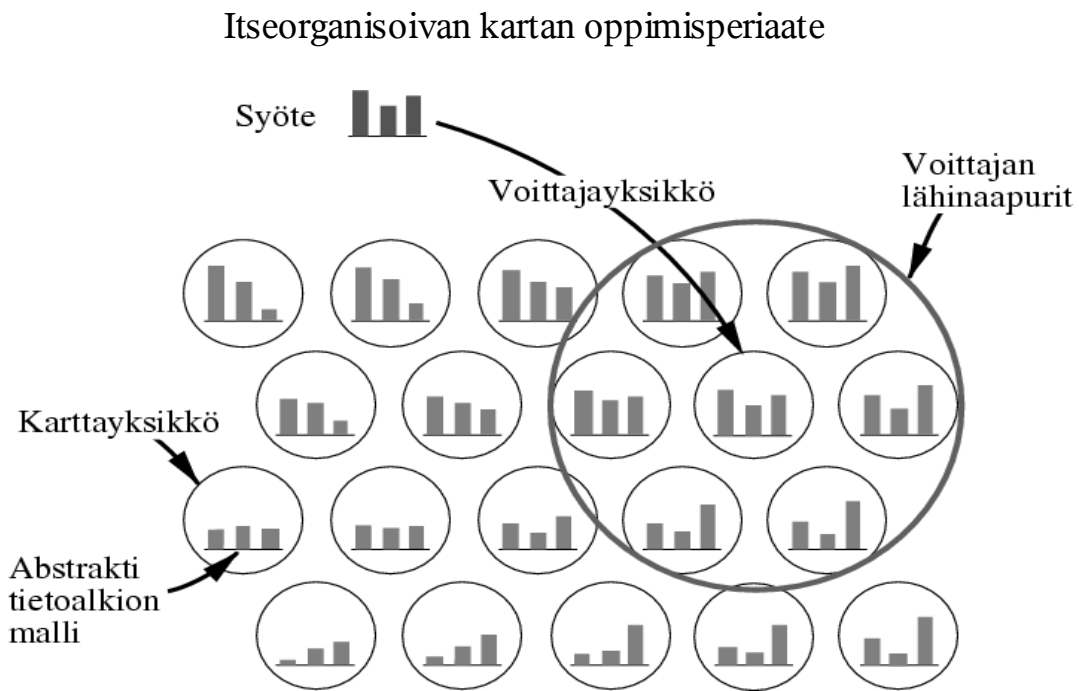
2.4. Ohjaamaton oppiminen ja itseorganisoiva kartta

Ohjaamattomassa oppimisessa neuroverkko saa syötteitä mutta ei niiden vastetta. Kokonaisten eliöiden mallinnuksessa tämä vastaa tilannetta jossa eliö vain havainnoi ja ehkä valikoi, mitä syötteitä ympäristöstään poimii esimerkiksi huomiotaan suuntaamalla. Ohjaamattoman oppimisen tehtävänä on löytää annetuista syötteistä lainalaisuuksia. Tilastotieteestä tuttuja vastaavantyyppisiä menetelmiä ovat ryhmittely eli ryvästys (klusterointi) ja pääkomponenttianalyysi. Ohjaamattomassa opetuksessa syöteinformaatiosta muodostetaan malli neuroverkolla. Opetusalgoritmi on usein toteutettu siten, että siinä keinotekoiset neuronit kilpailevat keskenään siitä, mikä niistä edustaa kutakin syötettä parhaimmin. Tämän vuoksi menetelmistä käytetään usein termiä kilpailuoppiminen.

Yleisin ohjaamatonta oppimista toteuttava neuroverkko on itseorganisoiva kartta (Self-Organizing Map, SOM) (Kohonen 2001). Itseorganisoiva kartta koostuu joukosta soluja, jotka ovat tietyssä kiinteässä naapuruussuhteessa toisiinsa. Kutakin solua vastaa prototyyppi, ja alkutilassa ennen minkäänlaisten havaintojen saamista prototyyppien sisällöt voivat olla täysin satunnaisia. Verkko havainnoi syötteenä saamaansa näytevirtaa ja muodostaa vähitellen siitä järjestyneen kuvauksen tyypillisesti kaksiulotteiselle karttapinnalle. Järjestyneellä kartalla naapurisoluihin kuvautuu keskenään samankaltaisia tapauksia. Kartan eräs hyödyllinen ominaisuus on se, että tyypillisemmät havainnot vievät kartalta enemmän tilaa, jolloin niiden väliset erot kuvataan siis tarkemmin.

Itseorganisoiva kartta voi oppia jonkinlaisia prototyyppikäsitteitä ilman mitään nimitietoa, pelkästään havainnoimalla maailmassa olevia, aistijärjestelmän tai muun mittalaitteiston suodattamia tilastollisia ominaisuuksia ja niiden hahmoja.

Itseorganisoivan kartan oppimisperiaate on esitetty kuvassa 2. Oppimisperiaatteen tarkoituksena on havainnollistaa, miksi kartalle muodostuu järjestynyt representaatio. Kuhunkin karttayksikköön eli soluun liittyy abstrakti malli, joka on samaa tyyppiä kuin kartan saamat syötevektorit. Karttayksiköt on järjestetty esimerkiksi kaksiulotteiselle pinnalle jossa kullakin yksiköllä on pieni joukko välittömiä naapureita. Naapuruussuhteet ovat kiinteitä. Kun saadaan uusi syöte, voittajayksikön mallia muutetaan vastaamaan syötettä entistä paremmin. Myös voittajan lähinaapurien malleja muutetaan vastaavasti.



Kuva 2: Itseorganisoivan kartan oppimisperiaate.

Ohjaamattoman oppimisen keskeisenä tehtävänä aivojen ja kognition mallinnuksessa voidaan pitää representaatioiden muodostamista. Representaatiolla tarkoitetaan tässä jonkinlaista alkuperäisestä signaalista muunnettua esitysmuotoa, joka on hyödyllinen muita informaationkäsittelyn tehtäviä varten. Esitysmuoto voi olla vaikkapa alkuperäistä tiiviimpi tai tietyllä tavalla järjestynyt. Esimerkiksi piirteiden ja käsitteiden muodostusta voidaan pitää representaatiotehtävinä.

3. Mallintamisesimerkkejä

Neuroverkoista on tullut varsin suosittuja erilaisten hahmontunnistuksen tehtävien lisäksi myös kielikyvyn mallinnuksessa. Tälle on esitetty seuraavassa joitakin syitä.

- o Kielen luokat ja merkitykset voidaan nähdä sumeina ja ajan myötä muuttuvina, mitä perinteisen logiikan puitteissa on vaikea mallintaa.
- o Kielen tulkinta vaikuttaa olevan hyvin yksilöllistä: tulkinta perustuu yhteisössä muotoutuviin ja opittaviin konventioihin, sekä omiin kokemuksiin maailmasta, joten kunkin oma elinpiiri vaikuttaa siihen minkälaisia omat kielelliset merkitykset ovat.
- o Kielikyky riippuu ihmisen muista ajattelullisista kyvyistä kuten havaitsemisesta ja muistamisesta, joihin liittyviä prosesseja myös mallinnetaan neuroverkoilla

- o Neuroverkoilla on mahdollista mallintaa ihmiselle kielen tulkinassa, tuottamisessa ja oppimisessa tapahtuvia virheitä, sekä yksilöllisiä ominaisuuksia jotka riippuvat esimerkiksi siitä minkälaisia kielellisiä havaintoja kielen oppijalla on ollut.
- o Neuroverkot mahdollistavat hienovaraisten ja monisäikeisten riippuvuussuhteiden oppimisen esimerkkien avulla.

Seuraavaksi esitettävien esimerkkisovellusten tavoitteena ei niinkään ole ollut esittää, että ihmisen kielikyky toteutuisi täsmälleen tietyllä, mallin esittämällä tavalla. Sen sijaan kysymys johon usea malli pyrkii vastaamaan on, voidaanko edes periaatteessa tämä osa kielikykyä mallintaa neuroverkkojen avulla, Tai vaihtoehtoisesti, mikä tyyppisellä neuroverkkoarkkitehtuurilla kielikyvyn tietty osa-alue voidaan periaatteessa mallintaa.

3.1. Foneemikartat

Kielen mallinnuksen eräs perusongelma on puheentunnistus, ja siihen liittyen foneemien tunnistus.

Foneemien järjestymistä kuvaava itseorganisoiva kartta (Kohonen 1988) sai syötteen pelkästään lyhyiden puhesegmenttien spektraalisia (akustisia) ominaisuuksia. Ohjaamattomasti, siis ilman foneemien nimitietoa tapahtuvan opetuksen jälkeen kukin kartan neuronit representoi tiettyä foneemia, usein niin että foneemin eri variantit löytyvät lähekkäisistä neuroneista. Järjestyneelle kartalle siis muodostuu implisiittisiä foneemikategorioita, jotka heijastavat datassa olleita rykelmiä tai tihentymiä. Mikäli halutaan saada kartan järjestys näkyväksi, voidaan kartta myöhemmin *nimikoida* niiden näytteiden avulla, joiden nimike tiedetään. Foneemikartan tapauksessa tämä tapahtuu käyttämällä ääninäytteitä, joita vastaava foneemin nimi tai muu kategoriatieto on tunnettu. Tutkimuksessa aikaansaadussa kartassa esimerkiksi lähdetäessä vasemmasta yläkulmasta oikealle karttayksiköitä parhaiten vastaavat foneemit ovat järjestyksessä /a/ /a/ /ä/ /e/ /e/ /e/ /h/ /h/ /h/ /s/ /s/ /s/ /s/ ja lähdetäessä alaspäin ne ovat /a/ /a/ /o/ /a/ /a/ /o/ /o/ /u/ /u/. Kartan aikaansaaman järjestyksen hyvyys riippuu luonnollisesti siitä minkälaisia piirteitä kartalle ääninäytteistä syötetään. Kohosen esimerkissä piirteinä oli 15 frekvenssikomponenttia 9,83 millisekunnin kestoista puhesegmenteistä. Vaikka malli tuskin yksityiskohdiltaan toimii kuten ihmisen puheentunnistusjärjestelmä, se tarjoaa kognition mallinnuksen kannalta kiinnostavan ja varsin yksinkertaisen esimerkin siitä, miten puheen tunnistaminen periaatteessa voisi tapahtua.

3.2. Implisiittisten sääntöjen oppiminen: Englannin verbien taivutus

Eräs klassinen vastaväite neuroverkkoja vastaan kognitiivisen mallinnuksen piirissä on ollut että kognition eri osa-alueiden toteuttaminen edellyttää kykyä representoida erilaisia sääntöjä, ja että neuroverkot eivät kykenisi oppimaan tai esittämään sääntöjä. Rumelhart ja McClelland (1996b) sovelsivat monikerrosverkkoa (MLP) pyrkiessään osoittamaan että kieliopillinen tieto, jota voidaan kuvata esimerkiksi säännöillä, on mahdollista kuvata ja oppia neuroverkoilla *implisiittisesti*: vaikka sääntöä itsessään ei ole koodattu sellaisenaan mihinkään osaan neuroverkkoa, verkko käyttäytyy aivan kuin se olisi oppinut kyseisen säännön. Mallinnuksen kohteeksi he valitsivat Englannin verbien taivutuksen. Erityisesti heidän pyrkimyksensä oli tarjota selitysmalli sille, miten lasten kielenoppimisessa havaitut kolme oppimisvaihetta syntyvät. 1-vaiheessa lapsi käyttää vain pientä joukkoa verbejä, eikä tee niiden taivutuksessa virheitä sen enempää epäsäännöllisissä kuin säännöllisissä tapauksissa. 2-vaiheessa lapsi käyttää

suurempaa joukkoa verbejä, ja osaa tuottaa säännöllisiä taivutuksia myös täysin keksityille verbeille. Lisäksi lapsi tekee ns. yliyleistysvirheitä, eli muodostaa epäsäännöllisillekin verbeille toisinaan säännöllisiä taivutuksia. 3-vaiheessa näyttää kuin lapsi olisi oppinut taivutuksen ”kokonaan”, verbejä on vielä laajempi joukko ja niiden taivutus on jälleen virheetöntä myös epäsäännöllisten verbien osalta.

Neuroverkon opetus tapahtuu tässä ohjatun oppimisen tilanteen mukaisesti: syötekerrokselle tarjotaan verbin perusmuotoa, ja tavoiteulostulona on taivutettu muoto, kumpikin koodattuna 3-ulotteisella vektorilla. Tämä toistetaan useita kertoja joukolle verbejä. Verkko oppii tällöin assosioimaan verbin juurimuodon ja sen menneen taivutusmuodon. Oppiminen näkyy siten että saadessaan itselleen tutun juurimuodon, se tuottaa sille oikean päätteen. Lisäksi saadessaan juurimuodon jota ei ole nähnyt, se yleistää oppimastaan tuottaen sille jonkin aineistossaan yleisen taivutusmuodon. Nämä assosiaatiot representoituvat neuronien välisissä painoissa. Kun verkolle annetaan aluksi kaikkein tyypillisimpiä verbejä, sitten yhä suurempi ja suurempi joukko verbejä, myös verkko näyttäisi käyvän läpi oppimisvaiheet joita lasten on havaittu käyvän läpi. Itse oppimisvaiheet johtunevat osaksi siitä että rajallisen kokoinen neuroverkko oppiessaan yleistää, ja osaksi englannin epäsäännöllisten ja säännöllisten verbien yleisyysjärjestyksestä, joka vaikuttaa siihen mitä asioita verkko kulloinkin parhaiten osaa.

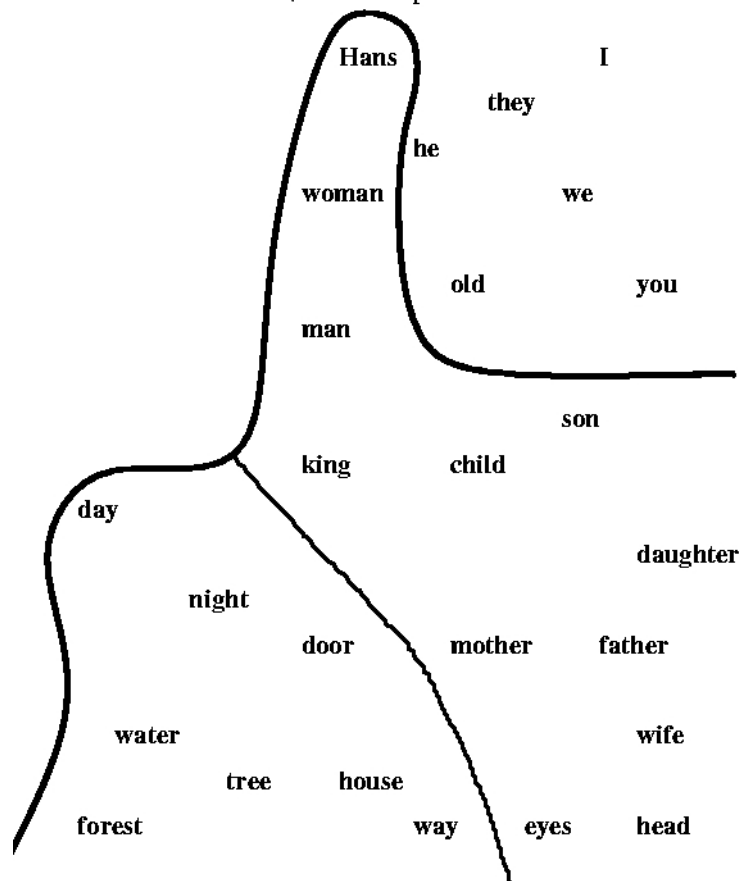
3.3. Käsitteiden representoituminen: Itseorganisoivat käsitekartat

Ritter ja Kohonen (1989) ovat tutkineet itseorganisoituvia käsitekarttoja. Yksittäiset symbolit (esim. luonnollisen kielen sanat tai logiikan predikaatit) eivät itsessään sisällä mitään tietoa siitä, mikä on niiden keskinäinen suhde. Tarvitaan jotain lisäinformaatiota symbolien välisten suhteiden määrittämiseen. Verkon oppimisvaiheessa voidaan esittää symbolit asianmukaisessa kontekstissa, esimerkiksi joidenkin todellisuutta kuvaavien piirrearvojen yhteydessä. Ritter ja Kohonen antoivat itseorganisoituvalle verkolle esimerkkeinä joukon vektoreita, joissa määriteltiin kutakin symbolia vastaavat piirrearvot. Tutkimuksessa käytettiin 16 eläimen nimeä ja niille määriteltiin koko (karkeasti kolmeen luokkaan jaotellen), rakenteen pääpiirteet (jalkojen määrä yms.) ja joitakin toiminnallisia ominaisuuksia (lentokyky yms.). Oppimisvaiheen tuloksena saatu kartta organisoitui tavalla, joka kuvaa topologiassaan esimerkeissä käytettyjen eläinten keskinäisiä sukulaisuussuhteita. Lisäksi voitiin todeta, että yleisemmät käsitteet (kuten lintu) rajautuivat omalle alueelleen, josta löytyivät myös käsitteen alakäsitteet (kuten haukka).

Ritter ja Kohonen (1989) käyttivät myös hyväkseen sanojen yhdessäesiintymisen ja merkityssuhteiden välistä yhteyttä antaen itseorganisoituvalle kartalle esimerkkeinä joukon sana-konteksti -pareja. Järjestelmä tuotti oppimisvaiheessaan kuvauksen -kartan, joka mielekkäällä tavalla kuvaa sanojen välisiä merkityssuhteita. Topologisesti kartalla toisiaan lähempänä olevien sanat ovat tyypillisesti merkitykseltään läheisempiä kuin kauempana toisistaan olevat sanat. Esimerkiksi sanat 'Mary' ja 'Jim' ovat syntyneellä kartalla lähempänä toisiaan kuin sanaa 'horse' tai 'eats'. Vastaavasti parien 'runs' - 'walks', 'likes' - 'hates' ja 'water' - 'beer' jäsenet ovat kartalla lähekkäin. Tällainen järjestys syntyy kartalle pelkästään sillä perusteella, missä yhteydessä sanat ovat kartan saamassa syötetekstissä esiintyneet.

Kuvassa 3 on esitetty osa sanakartasta, joka on muodostettu Ritterin ja Kohosen (1989) kuvaamalla periaatteilla. Sanakartta tuotettiin Grimmin satujen englanninkielisistä käännöksistä (Honkela ym. 1995). Kartalla kaksi sanaa on tyypillisesti lähellä toisiaan,

jos ne esiintyvät tekstissä samankaltaisissa lauseyhteyksissä. Tällä periaatteella itseorganisoiva kartta muodosti automaattisesti kuvauksen sanoista. Sanakartalla esimerkiksi verbit muodostivat oman alueensa kartan yläosassa ja substantiivit sijoittuivat kartan alaosaan (ks. kuva 3). Substantiivien alue jakaantui lisäksi elottomia ja elollisia olentoja kuvaaviin alueisiin. Luontevaa on myös se, että jälkimmäinen alue sijoittuu persoonapronomien alueen viereen.



Kuva 3: Sanakarttaesimerkki: joukko Grimmin saduissa esiintyviä sanoja itseorganisoivan kartan järjestämänä sen perusteella, minkälaisessa lauseyhteydessä ne esiintyvät (Honkela ym. 1995).

3.4. Neuroverkon leesiointi

Eräs tapa saada epäsuoraa informaatiota mallin pätevyydestä on häiriöiden tuottaminen malliin, esim. leesioimalla, ja vertaamalla tuloksia ihmisen toiminnan erilaisiin häiriötilanteisiin.

Miikkulainen (1997) on rakentanut yksityiskohtaisen neuroverkkomallin mentaalista leksikosta ja tutkinut, seuraako paikallisista malliin aiheutetuista häiriöistä afasialle tyypillisiä virheitä. DISLEX-nimisessä mallissa on komponentteina useita yhteen kytkettyjä itseorganisoivia karttoja. Ortografiselle, fonologiselle ja semanttiselle representaatiotasolle on kullekin oma karttansa.

Ortografisella kartalla *visuaalisesti* samankaltaiset sanahahmot kuvautuvat lähekkäin, ja fonologisella kartalla taas *ääniasultaan* samankaltaiset sanat löytyvät vierekkäin. Assosiativisten kytkentöjen ansiosta sanan ”dog” syöttäminen ortografiselle kartalle sumeutettuna kuvana (bittikarttana) aktivoi hetken kuluttua semanttisella kartalla käsitteen DOG.

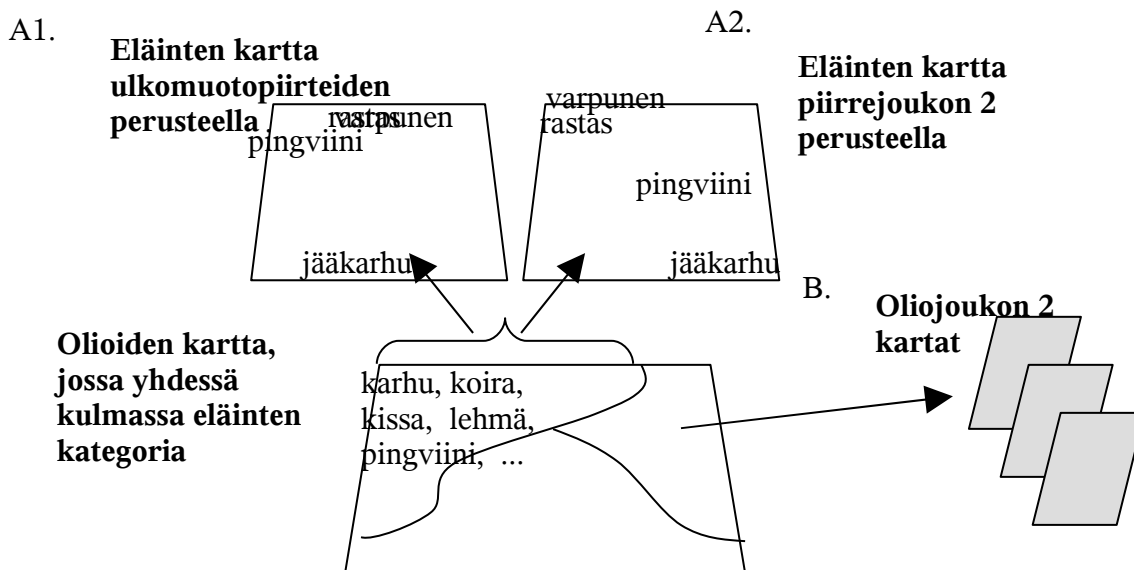
Kun karttojen välisiin kytkentöihin lisättiin kohinaa, mallin toiminta häiriintyi muistuttaen ihmisen tekemiä virheitä: ortografiselta kartalta semanttiselle kartalle kulkevaan kytkentään lisätty kohina aiheutti syvädysleksialle (deep dyslexia) ominaisen väärän vaikkakin lähekkäisen semanttisen kategorian aktivoitumisen (esim. DOG-käsitteen sijaan PREDATOR tai PREY). Jos taas kohinaa lisättiin semanttisen kartan ulostulokytkentään, joka johti ortografiselle kartalle, saattoi BALL-käsitettä vastaavan sanan tilalla aktivoitua melko samannäköinen sana, esimerkiksi DOLL. Miikkulainen (emt.) toteaa, että DISLEX-mallissa paikallisen häiriön aiheuttamat virheet ovat hyvin samankaltaisia kuin afasiassa ilmenevät kategoriaspesifit virheet, joita syntyy mm. paikallisen leesio seurauksena.

4. Neuroverkkomallinnus ja kognition modulaarisuus

Eräs kysymys kognition mallinnuksen piirissä on, kuinka modulaarisia mielen eri prosessit ovat. Esimerkiksi, kuinka riippumattomia tai modulaarisia kielikyvyyn mallinnuksessa ovat syntaktinen tieto, ja toisaalta tieto sanoista ja niiden merkityksistä. Varhaisessa konnektionismissa korostettiin hyvin vahvasti representatioiden hajautuneisuutta koko verkon alueelle. Ajateltiin, että neuroverkkomallinnus tarkoittaisi yksinomaan täysin homogeenista tai täysin hajautettua kognition kokonaisuutta. Kuitenkin koko verkko silloisissa esimerkkitapauksissa tarkoitti usein mallia, joka vastasi yhden pienen osatehtävän suorittamisesta.

Keinotekkoisten neuroverkkojen ominaisuuksien näkökulmasta ei ole uskottavaa, että aivot kokonaisuutena tai edes aivojen kuorikerros (korteksi) olisi yksi valtava, homogeenisesti kytketty neuroverkko. Eräs syy on että täysin kytketty neuroverkko sisältää valtavan määrän vapaita parametreja. Neuroverkon oppimiseen vaadittujen näytteiden määrä taas kasvaa yleisesti samaa tahtia kuin vapaiden parametrien määrä. Näin ollen suuren, täysin kytketyn neuroverkon saattaminen toimintakuntoon veisi kohtuuttoman kauan. Kuitenkin täysin kytketyn, täysin rinnakkaistetun neuroverkon ja erittäin modulaarisesti toteutetun tietokonearkkitehtuurin kaltaisen mallin välissä on suuri kirjo vaihtoehtoja, jotka ovat eri määrissä modulaarisia. Mallin korkean tason rakenne voi siis sisältää jonkinasteista neuraalista modulaarisuutta, vaikka mallin kaikki osa-alueet olisi toteutettu neuroverkein.

Kuvassa 3 on esitetty skemaattisesti, miten käsiterepresentaatiot voisivat koostua useista eri itseorganisoivista kartoista. Voidaan ajatella, että kuvan 3 rakenne on yhtäaikaan esimerkki sekä hajautetuista representaatioista että modulaarisuudesta. Hajautettua representaatiota on se, että tapaus ”jääkarhu” aktivoi jonkin alueen sekä olioiden kartalla että kahdella kartalla, A1 ja A2, jotka kuvaavat eläimiä eri näkökulmista. Neuraalisen tason modulaarisuutta taas heijastaa kunkin erillisen kartan olemassaolo: yksi kartta saa kaikista mahdollisista piirteistä ja kaikista mahdollisista tapauksista syötteenään vain jonkin osajoukon. Samantyyppistä yhtäaikaista modulaarisuutta ja hajautettua tiedon representointia edustaa Miikkulaisen (1997) DISLEX-mallissa jako ortografiseen, fonologiseen ja semanttiseen representaatioon.



Kuva 3: Eri tyyppisillä piirrejoukoilla saadaan samoille tapauksille eri tavoin järjestyneitä representaatioita. Kuvassa näkyy kahdentyyppistä neuraalisen tason modulaarisuutta: 1. tapausjoukon jakaantuminen aliryhmiin (A1 ja A2 eläinkartat vs. B. oliojoukon 2 kartat). 2. Eri tyyppisen piirretiedon kulkeutuminen eri representaatioille (eläimille on karttaa, A1 ja A2, jotka ovat organisoituneet eri piirteillä).

Neuraaliset, fysiologisen tai rakenteellisen tason moduulit voivat kukin osallistua usean eri osaongelman tai psykologisen tason tehtävän ratkaisemiseen. Churchland ja Sejnowski (1992) erottavat aivoissa ns. systeemien tarkastelutason. Systeemejä ovat esimerkiksi yksittäisen aistin kautta tulevaan tiedonkäsittelyyn liittyvät alueet ja mm. autonominen systeemi. Yksittäiseen systeemiin voi osallistua kartoja tai muita neuronikokoelmia useilta kaukana toisistaan olevia aivoalueilta pitkien, näitä alueita yhdistävien kytkentöjen välityksellä.

Itseorganisoivaa karttaa muistuttavia järjestyneitä representaatioita on toistaiseksi löydetty lähinnä aistikohtaista tietoa keskeisesti käsitteleviltä aivokuoren alueilta (Kohonen, 2001). Korkeamman kognition osalta prosessointia tunnetaan vähemmän, mutta joidenkin abstraktimpien piirteiden järjestyneitä, karttamaisia representaatioita on tiettävästi löydetty (emt). Kuitenkin koska aivokuoren fysiologinen rakenne on tiettävästi samankaltainen alueesta riippumatta, on mielekäästä olettaa, että samankaltaiset oppimisperiaatteet pätevät eri alueilla. Tähän viittaa myös se, että useita erillisiä kortikaalisia alueita voidaan mallintaa itseorganisoivan kartan avulla: se mikä muuttuu, on kunkin alueen saama piirretieto.

5. Lisää aiheesta

Kirjassaan ”Self-Organizing Maps” (2001) Teuvo Kohonen kuvaa itseorganisoivan kartan matemaattisena menetelmänä, sen variantteja, sekä sovelluksia moninaisiin insinöritieteen ongelmiin. Lisäksi Kohonen käsittelee SOMin fysiologista tulkintaa aivojen mallinnuksessa. Varsinaista kognition mallinnusta kirjassa ei juurikaan käsitellä. Tähän puoleen panostaa vahvemmin Gärdenfors (2000), joka esittää nk.

käsitteellisten avaruuksien (conceptual spaces) teorian kautta, miten neuraalinen representaation taso, käsitteiden taso ja symbolinen taso voisivat kytkeytyä yhteen. Hän pitää itseorganisoivaa karttaa eräänä mahdollisena käsitetason mallina.

Kirjassa "Introduction to Neural and Cognitive Modeling" (2000) Daniel Levine tarjoaa laaja-alaiselle lukijakunnalle melko lähestyttävän tavan perehtyä neuroverkkomallinnuksen historiaan, joukkoon neuroverkkomenetelmiä, sekä niiden käyttöön nimenomaan kognition mallinnuksessa. Menetelmien osalta kirjan keskiössä ovat erityisesti Grossbergin ja kumppanien neuroverkkomallit, kuten ART-verkko (Adaptive Resonance Theory).

Kirja "Rethinking Innateness" (Elman ym., 1996) on vahva kannanotto neuroverkkomallinnuksen puolesta. Soveltamalla tiettyjä neuroverkkomalleja psykolingvistiikan koeasetelmissa havaitaan muun muassa, miten pinnalta tarkastellen diskreeteilä ja synnyntäisiltä vaikuttavat ilmiöt kuten "kehitysvaiheet" (stages) voivat emergoitua pohjimmiltaan jatkuvien, epälineaaristen dynaamisten prosessien toiminnasta. Kirjassa sovelletaan erityisesti sekventiaalisen tiedon käsittelyyn soveltuvia takaisinkytkettyjä verkkoja kuten SRN (Simple Recurrent Network).

Hyvärinen, Karhunen ja Oja (2001) esittelevät riippumattomien komponenttien analyysi (Independent Component Analysis, ICA) -nimisen menetelmän, jota on sovellettu myös neuroverkkomallina. ICA-mallilla voidaan menestyksekkäästi selittää ilmiöitä, jotka liittyvät erilaisten lähteiden tai piirteiden erotteluun toisistaan. Perinteinen esimerkki on nk. cocktail party -ongelma: miten kuulija erottaa toisistaan eri puhujat.

Kirjallisuus

- Churchland, P.S. & Sejnowski, T.J. (1992). Neuroscience Overview. Luku 2 kirjassa *The computational brain*, 17-29. Cambridge MA: MIT Press.
- Elman, J.L. (1990) Finding Structure in Time. *Cognitive Science* 14:179-211.
- Elman, J.L. (1996). Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, Km Plunkett. *Rethinking Innateness*. MIT Press.
- Elman, J.L. (2001). Connectionism and Language Acquisition. Teoksessa M. Tomasello & E. Bates (toim.), *Language Development: The Essential Readings*, 295-306. Oxford: Blackwell Publishing.
- Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge, MA: MIT Press.
- Honkela, T., Pulkki, V. & Kohonen, T. (1995). Contextual Relations of Words in Grimm Tales Analyzed by Self-Organizing Map. Proceedings of ICANN-95, Paris: EC2 et Cie., ss. 3-7.
- Kohonen, T. (1988). "Neutral" phonetic typewriter. *IEEE Computer Magazine*, 27: 3, 11-22.
- Kohonen, T. (2001). *Self-Organizing Maps*. 3rd extended edition. Berlin: Springer.
- Levine, D.S. (2000). *Introduction to Neural and Cognitive Modeling*. 2nd Edition. Mahwah, NJ: Lawrence Erlbaum.
- Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon and memory*. Cambridge, MA: MIT Press.
- Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language*, 59:34-366.

- Rumelhart D.E. & McClelland, J.L. (1986a). Feature discovery by competitive learning. Teoksessa Rumelhart D. E. & McClelland, J. L. (toim.), *Parallel Distributed Processing: Explorations in the microstructure of cognition*, 151-193. Volume 1: Foundations. Cambridge, MA: MIT Press.
- Rumelhart D.E. & McClelland, J.L. (1986b). On learning the past tenses of English verbs. Teoksessa Rumelhart D. E. & McClelland, J. L. (toim.), *Parallel Distributed Processing: Explorations in the microstructure of cognition*, 216-271). Volume 2: Psychological and biological models. Cambridge, MA: MIT Press.