

Konekcionizmus v náručí výpočtovej kognitívnej vedy

Igor FARKAŠ¹

Abstrakt. Príspevok sa zaoberá princípmi konekcionistických modelov (umelých neurónových sietí) a ich porovnaním s ostatnými paradigmami kognitívnej vedy (symbolizmom a dynamickým prístupom). Venujeme sa dôležitým konceptom ako sú mentálna reprezentácia, definícia počítania, mechanizmy učenia, významy a problém ukotvenia symbolov. Uvádzame teóriu ukotvanej kognície založenú na nových empirických poznatkoch o mozgu/mysli. Využitie umelých neurónových sietí ilustrujeme na príkladoch týkajúcich sa modelovania raného kognitívneho vývinu, ako napríklad vizuálno-motorická koordinácia, modelovanie referenčných rámcov v mozgu ako súčasti priestorovej kognície, rozpoznávanie objektov, a spracovanie jazyka na lexikálnej a vetnej úrovni, a jeho integráciu s ostatnými kognitívnymi funkciami.

1 Úvod

Vznik kognitívnej vedy ako vednej disciplíny, spojený s kognitívnou revolúciou v polovici 20. storočia, bol vyvrcholením paralelných snáh o poznávanie ľudskej mysle (mind) a mentálnych procesov [57]. Otázky týkajúce sa ľudskej mysle a mentálnych procesov si filozofi kladli už počas antických čias, a psychológovia asi od polovice 19. storočia. No až v priebehu minulého storočia sa naakumulovalo dostatočné množstvo poznatkov a vznikla vôľa pre komunikáciu odborníkov z rôznych tzv. materských disciplín kognitívnej vedy, čím sa vytvorili predpoklady pre interdisciplinárny prístup v tejto oblasti. To znamená, že na myseľ sa pozeráme z rôznych perspektív a využívame pritom rôzne dostupné metódy skúmania, čo v konečnom dôsledku obohacuje každú perspektívu. To si samotný objekt skúmania priamo vyžaduje, a to kvôli svojej komplexnosti z pohľadu úrovni a metód skúmania, ako aj akémusi špeciálnemu statusu mysle/mozgu ako entity vo svete okolo nás (a v nás). Dnes už prevláda názor, že myseľ a mozog úzko súvisia, že ich nemožno úplne oddeliť (ako to kedysi tvrdil Descartes v podobe dualizmu substancií). Na mozog sa pozeráme z objektívnej perspektívy (tretej osoby), skúmajúc ho metódami prírodných vied, zatiaľ

¹ Centrum pre kognitívnu vedu, Katedra aplikovanej informatiky, FMFI, Univerzita Komenského, Mlynská dolina, 84248 Bratislava, e-mail: farkas@fmph.uniba.sk

čo pri skúmaní mysle musíme využívať iné vedecké metódy, lebo každý z nás má interný prístup len ku svojej vlastnej mysli.

V minulosti bola povaha mysle predmetom vášnivých debát filozofov mysle, no v ostatných dekádach rokov sa vzťah mysle/mozgu preniesol do portfólia aj iných disciplín, lebo aj o mozgu už toho vieme oveľa viac. Dodnes neexistuje úplný konsenzus ohľadne tohto vzťahu, aj keď isté zúženie manévrovacieho priestoru prijateľných teórií sa dosiahlo, najmä smerom k rôznym formám tzv. neredukcionistického fyzikalizmu (napr. [38]). Nad otázkami fungovania mysle (človeka, resp. zvierat'a) sa vedeckými metódami zamýšľajú kognitívni psychológovia, ktorí formulujú teórie, a na ich základe vedecké hypotézy, ktoré testujú pomocou behaviorálnych experimentov. Povahou a fungovaním mysle sa zaoberajú i lingvisti, pretože jazyk možno považovať za akúsi bránu do mysle človeka, ktorý prostredníctvom jazyka dokáže vypovedať o svojich (vnútorných) mentálnych stavoch (čo v prípade zvierat využiť nemôžeme). Psycholingvistika sa opiera o psychologické experimenty zamerané na jazyk, zatiaľ čo neurolingvistika si všíma vzťahy medzi poškodeniami mozgu a jazykovým správaním, čo umožňuje hľadať tzv. neurálne koreláty jazykových funkcií. Kognitívna neuroveda tento vzťah skúma v celom spektre kognitívnych funkcií mysle/mozgu, a to vo svojich začiatkoch hlavne pomocou behaviorálnych experimentov. K tomuto skúmaniu od 80-tych rokov minulého storočia napomáhajú zobrazovacie metódy mozgu (napr. na báze funkčnej magnetickej rezonancie, alebo merania elektrickej aktivity mozgu), ktoré nám poskytujú informáciu o činnosti rôznych častí mozgu, s rozličnými stupňami priestorového a časového rozlíšenia, v závislosti od použitej metódy. Kognitívna antropológia má tiež svoje miesto v kognitívnej vede, pretože človek je sociálna bytosť, a tak jeho poznanie a s tým súvisiaci vývin mysle závisí aj od kultúry. Napokon, vznik počítačov sprevádzajúci kognitívnu revolúciu vniesol do kognitívnej vedy nové svetlo (ale i nové otázky a paradigmy) a umelá inteligencia sa tak stala tiež jednou z materských disciplín kognitívnej vedy. Snaha výpočtovo modelovať kognitívne procesy, čo je možné vďaka počítačom, nie je samoúčelná, ale nutná na to, aby sme mohli testovať vedecké hypotézy a prostredníctvom modelov a ich predikcií ponúkať inšpiráciu pre nové experimenty.

Práve výpočtové modelovanie budeme zdôrazňovať v tomto príspevku, pretože bez neho nemôže poznanie v kognitívnej vede principiálne dosiahnuť uspokojivú úroveň, čo sa týka kognitívnych mechanizmov (na rôznych časopriestorových úrovniach fungovania). Samozrejme, týmto sa výpočtové modelovanie nestáva jediným kľúčovým prístupom v kognitívnej vede, ktorá sa nezaobíde ani bez teoretických a experimentálnych metód skúmania. Skrátka, jedine spoločný postup všetkých metód skúmania mysle môže priniesť ovocie (ak keď finálny cieľ sa v tomto kontexte stále zdá byť dosť vzdialený).

Táto kapitola je organizovaná nasledovne. Najprv sa prejdeme hlavnými míľnikmi – paradigmami, ktoré sprevádzali vývin kognitívnej vedy až dodnes (časť 2). V kontexte týchto paradigiem obrátíme pozornosť na niektoré kľúčové koncepty, ako je počítanie (časť 3), reprezentácie a významy (časť 4). Pokúsim sa vysvetliť, prečo považujem konekcionizmus za silnú paradigmu, ktorá ponúka kognitívne relevantné vysvetlenia. Použitelnosť konekcionistických modelov budem ilustrovať v spektre

príkladov týkajúcich sa raného kognitívneho vývinu zahŕňajúc nielen oblasť tzv. nižšej kognície ale aj vyššej ako napr. jazyk (časť 5). Časť 6 tvorí zhrnutie hlavných bodov kapitoly. Ak sa mi podarí nejakým spôsobom inšpirovať (alebo dokonca presvedčiť) cteného čitateľa, táto kapitola splnila svoj účel. Ak nie, nevadí, každý má právo na svoj názor aj v oblasti kognitívnej vedy. ☺

2 Paradigmy kognitívnej vedy

Od svojich počiatkov prešla kognitívna veda za 60 rokov svojej existencie vývinom, ktorý nadobudol istý trend. Ukazuje sa totiž, že v kognitívnej vede nastáva nevyváženosť, pretože kognitívna psychológia sa stáva dominantnou disciplínou. Gentner ([59]) uvádza, že v súčasnosti tvorí kognitívna psychológia už viac ako 50% vedeckej produkcie v kognitívnej vede (pričom jej nárast začal už zhruba v roku 1978, keď tvorila asi 26%), zatiaľ čo ostatné dve disciplíny, ktoré patrili k jadru kognitívnej vedy už v jej počiatkoch – umelá inteligencia a lingvistika – majú asi po 20%. Ak tento trend bude pokračovať, v priebehu ďalších 30 rokov „ovláadne“ kognitívna psychológia kognitívnu vedu. Toto by však viacerí psychológovia, vedomí si dôležitosti interdisciplinárneho poznávania, považovali za Pyrrhovo víťazstvo. Bolo by veľmi dobré, keby tento poznatok sa stal vetrom do plachiet rôznym výpočtovým modelom, pretože empirických dát je už pomerne dosť (no nikdy nie dosť ☺).

Jednotlivé paradigmy, ktoré spomenieme, sa týkajú výpočtových prístupov pri formalizácii mysle a mentálnych procesov, no majú aj svoje filozofické implikácie týkajúce sa vzťahu mysle a mozog. Formálne výpočtové prístupy, ako bolo spomínané, ponúkajú **explicitné mechanizmy** fungovania mysle/mozgu, čo principiálne nemožno dosiahnuť pomocou teoretických modelov (s krabičkami a šípkami). Pre jednotlivé paradigmy kognitívnej vedy sú kľúčovými pojmami (konceptmi) **mentálne reprezentácie**, **významy** a **výpočtové operácie**, avšak ako uvidíme, tieto koncepty sa používajú nerovnakým spôsobom, čo sa pokúsím objasniť. Spomeňme stručne tri hlavné paradigmy kognitívnej vedy: symbolovú², konekcionistickú a dynamickú.

2.1 Symbolová paradigma

Vznik symbolovej paradigmy (pre ktorú sa zaužíval aj názov **kognitívizmus**³ alebo aj **klasická paradigma**) bol podnietený vynájdením moderných digitálnych počítačov, ktoré realizujú diskkrétne výpočty so symbolmi. Výpočet v počítači prebieha za pomoci dvoch odlišných kľúčových komponentov: procesora a pamäte. Procesor sériovým spôsobom spracováva symboly, uložené v pamäti a vykonáva pritom inštrukcie podľa programu uloženého v inej časti pamäte. Interakcia počítača s prostredím je na periférii záujmu a prebieha prostredníctvom vstupno-výstupných podsystemov, ktoré sprostredkovávajú vstupy pre centrálny procesor (vhodnou transformáciou) a výstupy

² Niektorí autori používajú v slovenčine, resp. češtine termín “symbolická”. Ja v kontexte kognitívnej vedy preferujem termín “symbolová” (súč inšpirovaný Martinom Takáčom), aby sme tento význam odlišili od konvenčného významu slova (napr. symbolický akt).

³ Termín kognitívizmus nebudem používať, lebo môže navodzovať dojem, že táto paradigma má najbližšie ku vysvetleniu kognitívnych procesov

z neho, konzistentne napr. s teóriou modularity ([48]). Klasická paradigma konceptualizuje **mysleľ ako výpočtový stroj**, oddeliteľný od prostredia, ktorý manipuluje s internými symbolmi podľa logických pravidiel, podobne ako to robí počítač. Mentálne operácie sú realizované pomocou vnútorného jazyka mysle, napr. mentálčiny ([49]). *Výpočtovú teóriu mysle* vystihuje hypotéza **o fyzikálnom symbolovom systéme**, ktorý podľa jej autorov „disponuje nutnými a postačujúcimi prostriedkami na všeobecné inteligentné konanie“ ([100], s. 87)⁴.

Vychádzajúc zo skorších prác symbolikov ([143],[100],[48]), Harnad ([65]) zrekonštruoval výstižnú **definíciu symbolového systému** ako (1) súboru arbitrárnych „fyzických znakov“ (tokens), zaznamenaných na papieri, páske, či inom médiu, s ktorými (2) sa dá manipulovať pomocou „explicitných pravidiel“, (3) podobných znakom, či reťazcom znakov. Pravidlami riadená symbolová manipulácia je zameraná (4) iba na tvar symbolových znakov (nie ich „význam“), tzn. je čisto syntaktická a obsahuje (5) „pravidlami riadené kombinácie“ a rekombinácie symbolových znakov. Existujú (6) primitívne atomické symbolové znaky a (7) zložené znaky (reťazce). Celý systém a všetky jeho časti – atomické znaky, zložené znaky, syntaktické manipulácie, aktuálne i možné pravidlá sú (8) „sémanticky interpretovateľné“, t.j. syntaxi možno systematicky priradiť význam, napr. pri zastupovaní objektov alebo opisovaní stavov.

Z pohľadu symbolizmu je teda podstatné, že myslenie (a tým aj konanie) človeka je vnútorne riadené algoritmami, ktoré sú realizované programami, a nejako implementované v mozgu. Ako, to symbolizmus nerieši, pretože sa to považuje len za vec implementácie. Pri počítačoch môžeme ľubovoľný algoritmus naprogramovať v nejakom programovacom jazyku a potom ho implementovať vo zvolenej hardvérovej platforme. Podobná predstava prevláda i v prípade mozgu. Nezáleží ako sú v mozgu implementované mentálne procesy.

Počítačová metafora mysle inšpirovala aj Marrova ([90]) teóriu troch nezávislých **úrovní abstrakcie** (vysvetlenia nejakého fenoménu), a to výpočtovej, algoritmickej a implementačnej, ktorú rozpracoval v kontexte modelovania vizuálneho spracovania informácie, a ktorá výrazne ovplyvnila kognitívnu vedu. *Výpočtová* úroveň opisuje to, aká informácia sa počíta a prečo, bez bližšej špecifikácie ako. *Algoritmická* úroveň už vyžaduje znalosť o používaných reprezentáciách, uchovávajúcich informácie, a o výpočtoch s nimi. *Implementačná* úroveň sa týka realizácie konkrétneho algoritmu vo fyzickom substráte. Medzi úrovňami platí vzťah, že jedno vysvetlenie na vyššej úrovni sa dá transformovať na viacero vysvetlení na nižšej úrovni. Ako ilustračný príklad uvažujme násobenie dvoch viacciferných čísel. To je teda cieľ výpočtu, ktorý sa dá dosiahnuť rôznymi algoritmami, napríklad takým, ktorý človek bežne používa, a v ktorom sa medzivýsledky násobenia (jedného čísla číslicou druhého čísla) zapíšu pod seba a potom sa sčítajú. Napokon, implementačná úroveň už predstavuje konkrétnu realizáciu tohto algoritmu v nejakom fyzickom médiu (počítač, pero a papier, a i.). Je zrejmé, že symbolová paradigma sa zaoberá výpočtovou a algoritmickou úrovňou analýzy, implementačná je považovaná za nepodstatnú.

⁴ Čo to presne je inteligentné konanie/správanie nás teraz nemusí trápiť. Postačí, že čitateľ má o tom intuitívnu predstavu. Viac na túto tému možno nájsť v [137].

Symbolový (algoritmický) pohľad na myseľ je dodnes viacerými odborníkmi považovaný za tú správnu úroveň abstrakcie najmä v kontexte vysvetľovania procesov tzv. **vyššej kognície**, ako je usudzovanie, plánovanie, a tiež používanie jazyka (napr. [49][107][136]). Výhodou klasickej paradigmy je, že poskytuje silné matematické a logické formalizmy, ktoré sú zväčša transparentné a preto človeku zrozumiteľné. Je celkom možné, že ide o jednu z početných kognitívnych predispozícií človeka (cognitive biases) smerom k diskretným formalizmom, pretože operácie používané na symboloch sú blízke ľudskému mysleniu ([115]). Pre človeka je prirodzené uvažovať v termínoch diskretných podmienok, pravidiel, propozícií a logických inferencií. Napríklad, ak sa niečoho bojím (propozícia opisujúca môj emocionálny stav je pravdivá), mojím rozhodovaním sledujem cieľ, aby som vykonaním vhodnej akcie dosiahol požadovaný stav, aby daná propozícia nebola pravdivá.

2.2 Konekcionizmus

Konekcionizmus predstavuje spektrum metód, ktoré vznikli v rámci umelej inteligencie, a ktoré boli inšpirované architektúrou a fungovaním mozgu. Môžeme teda hovoriť o inšpirácii „zdola.“ Konekcionistický systém pozostáva z množiny jednoduchých prvkov – umelých neurónov, navzájom komunikujúcich cez váhované spojenia, ktoré predstavujú dlhodobú pamäť systému. Sila konekcionistického systému – umelej **neurónovej siete** – nie je v samotných neurónoch, ale v ich **vzájomnom prepojení a interakcii**. Paralelné spracovanie a distribuovanosť aktivity predstavujú základný architekturný rozdiel v porovnaní so symbolovým systémom, pretože každý neurón je súčasne procesorom aj pamäťou (aj keď elementárnou). Taktiež, povaha komunikácie medzi neurónmi má **numerický** a nie symbolový charakter, preto hovoríme v prípade sietí o **subsymbolorých reprezentáciách**⁵. Neurónová sieť môže pracovať so spojitými hodnotami, čo je veľmi vhodným nástrojom na modelovanie kognície, pretože empirické dáta nasvedčujú tomu, že mnoho kognitívnych procesov má spojitý prejav. Napríklad, nielenže perceptuálne vstupy (napr. obraz na sietnici) a motorické výstupy bývajú spojité (pohyby, zvuková podoba reči), ale podobne môžeme argumentovať aj v prípade modelovania vyšších procesov, napr. rozhodovania ([133]). Spomeňme si, aké ťažké môže byť niekedy rozhodnúť sa medzi dvoma alternatívami. Tento proces „súperenia“ alternatív sa dá vysvetliť pomocou neurónovej siete so spojitými aktiváciami. Neurónová sieť teda nespracováva symboly, ale numerické dáta v podobe vektorov. Operácie nad takýmito vektorovými dátami sú (nelineárnymi) **transformáciami**, čo má svoje implikácie. Je zrejmé, že takéto operácie nie sú vôbec transparentné, a preto sú človeku málo zrozumiteľné, a aj preto si neurónové siete vyslúžili prívlastok „čierna skrinka.“ Bez použitia techník zhľukovania a vizualizácie vysokorozmerných dát naozaj nevieme, čo sa v neurónovej sieti deje, keď sledujeme jej prebiehajúce aktivity.

⁵ Aj keď treba upresniť, že tzv. lokalistické modely neurónových sietí (teda nie distribuované) pracujú s reprezentáciami, ktoré možno nazvať symbolovými. Každý neurón reprezentuje jeden symbol (má hodnotu 1, ostatné 0), a nezáleží na tom, ktorý neurón reprezentuje ktorý symbol – forma a obsah sú oddelené.

Konekcionistický prístup, na rozdiel od symbolového, zahŕňa aj implementačnú úroveň analýzy, pretože mozog nie je počítač (časť 3), a teda pre mozog neplatí implikácia o nezávislosti implementačnej úrovne od vyšších úrovní. Niektorí filozofi navrhujú oddeliť vysvetľované fenomény od mechanizmov ich vzniku ([1]), no v konekcionistických modeloch sú tieto dva aspekty prepojené, teda existuje závislosť medzi implementovanými mechanizmami (danými interakciami medzi neurónmi) a kognitívnymi fenoménmi ([103]).

Dôležitou vlastnosťou umelých neurónových sietí je to, že na rozdiel od symbolovej umelej inteligencie, kde znalosť do systému vkladá dizajnér, ponúkajú biologicky inšpirované mechanizmy akvizície znalostí, **učenia**, aplikovateľné na rôznych úrovniach abstrakcie ([67],[35]). To znamená, že parametre siete (váhy spojení) nemusíme explicitne nastaviť (čo sa prakticky ani nedá), no treba siete poskytnúť vhodný mechanizmus učenia. Podobne ako človek, neurónová sieť sa učí na príkladoch. Učenie spôsobuje zmenu váh tak, aby sa po naučení dosiahlo požadované správanie siete, čo formálne odpovedá hľadaniu extrému (minima, resp. maxima) nejakej hodnotovej funkcie.

V prípade neurónových sietí tiež bežne hovoríme, že počítajú a teda spracovávajú informáciu. Počítajú však rovnako ako symbolové systémy? Niektorí odborníci tvrdia, že áno, pretože existujú teoretické práce o ekvivalencii medzi univerzálnym počítačom (Turingovým strojom⁶) a rekurentnou neurónovou sieťou s nekonečne veľkou pamäťou (pozri napr. prehľad v [138]). Táto ekvivalencia však predpokladá klasický pohľad na počítanie ako na vykonávanie programu (algoritmus). Neurónové siete však ponúkajú viac. Modely neurónových sietí sú spektrom metód z pohľadu typov aktivácií neurónov a ich dynamiky. Modely, ktoré pracujú v spojitom priestore, a niektoré dokonca v spojitom čase sú matematickými modelmi, ktoré sa dajú presne opísať len pomocou (nelineárnych) diferenciálnych rovníc, a nie diskretných krokov výpočtu (počítačová implementácia ponúka len aproximáciu). Zahnutie takýchto modelov si vyžaduje voľnejšiu definíciu počítania (časť 3). Od modelov neurónových sietí pracujúcich v spojitom čase sa priamo dostávame k paradigme dynamických systémov.

2.3 Dynamické systémy

Čo klasická paradigma a konekcionizmus zdieľajú, je ich izolovanosť od okolitého prostredia. V prípade symbolového modelu je to zjavné, pretože izolovanosť je jedným z predpokladov. Neurónová sieť sa síce učí, čiže interaguje s prostredím, no bežným predpokladom je, že všetky tréningové dáta máme hneď dispozícii, takže interakcia nie je „plnohodnotná,“ pretože výstupy siete nemajú dopad na prostredie. Vďaka izolovateľnosti výpočtového modelu od prostredia môžeme (no nemusíme) pracovať s vnútornými reprezentáciami entít z prostredia, ktoré systém spracováva.

Takýto prístup sa stal tŕňom v oku niektorým kognitívnym vedcom, ktorí prišli s myšlienkou kognitívneho systému ako dynamického systému, ktorý nepotrebuje reprezentácie, pretože celý proces interakcie s prostredím prebieha v spojitom čase

⁶ Bežný digitálny počítač môžeme chápať ako fyzický príklad Turingovho stroja, aj keď nemá nekonečne veľkú pamäť (čo sa teoreticky predpokladá, aby sa dal na ňom realizovať ľubovoľný algoritmus).

([142],[113]). Kognitívne procesy teda nie sú chápané izolovane od prostredia, ale v úzkej previazanosti s ním (coupling). Kognitívne procesy sú mentálnou aktivitou v čase, ktorú môžeme opísať pomocou systému nelineárnych diferenciálnych rovníc. Táto paradigma, postavená na teórii nelineárnych dynamických systémov predpokladá existenciu dvoch kľúčových, vzájomne prepojených, typov premenných. Jedným sú tzv. kolektívne premenné, ktoré zahŕňajú vzťahy medzi interagujúcimi zložkami dynamického systému, a vysvetľujú správanie systému. Druhým typom sú riadiace parametre, ktorých kvantitatívne zmeny môžu spôsobovať kvalitatívne zmeny v správaní systému (tzv. fázové prechody). Príkladom takéhoto systému môže byť pohyb koňa, ktorého rytmus pohybu (vyjadreného periodickými priebehmi malého počtu kolektívnych premenných) závisí od rýchlosti pohybu (riadiaci parameter). Kôň sa môže pohybovať v každom čase len v jednom rytme, (fázové) prechody medzi nimi nemusia nutne nastávať pri tých istých rýchlostiach, a závisia aj od stavu prostredia (napr. hrbolatá cesta). Súčasťou takéhoto dynamického opisu kognitívneho procesu sa teda priamo stáva aj prostredie, čo rozširuje konceptuálne chápanie mysle. Bressler a Kelso ([12]) argumentujú, že dynamický pohľad je relevantný aj pre charakterizáciu samotných mentálnych procesov (vnútri kognitívneho systému), hoci identifikácia kolektívnych premenných a riadiacich parametrov v tomto prípade je zatiaľ tvrdým orieškom.

Dynamický pohľad na spoznávanie vonkajšieho sveta je konzistentný so **zjednávacím prístupom** ([141]), kde sa dôraz z vnútorných reprezentácií (vopred daného) vonkajšieho sveta presúva na **vnímanie a jednanie vo svete**, ktorý sa takto spoluvytvára. Tým vzniká prostredníctvom spätnej väzby kruhový proces, pretože zjednávaný svet je prostredím pre zjednávanie (enactment), ale zjednávanie je súčasne predpokladom zjednávaného sveta. V rámci tohto pohľadu si každý subjekt (kognitívny systém) zjedná svoj svet sám v procese dynamickej interakcie so svetom (vrátane iných jedincov), a ten bude jedinečný.

2.4 Smerom k ukotvenej kognícii

Spomínané paradigmy vznikali postupne, tak ako sme ich spomínali, a v súčasnosti stále tvoria jadro výpočtových prístupov kognitívnej vedy, pretože každá paradigma má svojich zástancov. Ešte sa im budeme venovať, teraz venujme na chvíľu pozornosť hlavným prúdom a teoretickým koncepciám vo vývoji kognitívnej vedy, ktoré prispeli k posunu v dominantnosti jednotlivých paradigmat ([59]). Už v 70. rokoch minulého storočia sa v kognitívnej psychológii objavovali skeptické pohľady na centrálnosť (vnútorných) reprezentácií a symbolového spracovania informácie v ľudskej kognícii. Hlavný predstaviteľ, Gibson ([61]), tohto prúdu známeho ako **ekologická psychológia**, zdôrazňoval, že „nie je dôležité, čo je vo vnútri hlavy, ale vnútri čoho (akého prostredia) sa hlava nachádza.“

Tento prúd plynulo prešiel v 80. rokoch do teórie **situovanej kognície**, ktorá bola postavená na predpokladoch, že ľudská inteligencia je fundamentálne interaktívna ([70]), a že ľudská kognícia je fundamentálne neoddeliteľná od kontextu ([82]). Podľa prvého predpokladu človek bezprostredne využíva prostredie (iných ľudí, zariadenia, situácie, atď.) pre svoje inteligentné konanie, a je s nimi spojený v rámci

spätnoväzbovej slučky. Druhý predpoklad viedol k rozšírenému chápaniu mysle a kognície v zmysle jej umiestnenia aj do prostredia – myseľ človeka nesídlí len v jeho mozgu, ale jej súčasťou sú aj prvky prostredia (napr. knihy, obrazy, počítač, rôzne poznámky a i.).

Dôraz na interakciu s prostredím ostal v centre pozornosti aj naďalej, no v 90-tych rokoch sa začal klásť dôraz na reprezentácie, ktoré túto interakciu umožňovali, a to v rámci prúdu **vtelenej kognície**. Vtelená (stelesnená, embodied) kognícia, podobne ako tvrdil Gibson, vychádza z predpokladu, že pochopenie mentálnych procesov nemožno oddeliť od tela človeka, konkrétne jeho zmyslov a motoriky. Avšak vtelená kognícia predpokladá, že informácia môže vstúpiť do tela organizmu v podobe modálnych reprezentácií a podieľať sa na ďalších, vyšších procesoch kognície ([6]). Zmysly určujú povahu perceptuálnych vstupov, a motorické ústrojenstvo zase určuje spektrum akcií, ktoré človek môže vykonať, a ktoré tiež predstavujú obmedzenia (constraints) pre vznik a modifikáciu vnútorných reprezentácií. Napríklad, mentálna reprezentácia objektu kladivo nezahŕňa len jeho perceptuálne vlastnosti, ale aj možnosti použitia tohto objektu (ak s ním dotýčný subjekt má skúsenosť).



Obr. 1. V klasickom systéme sú (modálne) percepčné/motorické procesy oddelené od amodálnych kognitívnych procesov, a vzájomné prepojenie je realizované cez (nešpecifikovaný) mechanizmus transdukcie (smerom ku kognitívnym procesom) a referencie (späť). V prípade ukotveného systému majú reprezentácie vzájomný prienik, čím sa kognitívne (mentálne) procesy stávajú tiež modálnymi.

V komunite kognitívnej vedy sa objavilo viacero rôznych no súvisiacich pohľadov na vtelenie a situovanosť kognície ([60],[151]). Barsalou v [7] navrhol termín **ukotvená kognícia**, ktorý zjednocuje rôzne aspekty vtelenia a situovania v prostredí (**Obr. 1**). Silnú podporu pre túto paradigmu v kognitívnej vede možno nájsť v narastajúcej empirickej evidencii (za ostatných 20 rokov), či už z behaviorálnych štúdií (napr. [10]) alebo zobrazovacích metód (napr. [116]). Teórie ukotvenej kognície sú síce opisné, no ponúkajú testovateľné hypotézy pre ďalší empirický výskum (napr. [80],[132]). Zaujímavou predikciou týchto teórií je to, že všetky kognitívne funkcie (vrátane tých vyšších) sú redukovateľné na (nízkoúrovňové) senzomotorické procesy. Inými slovami, neurálne reprezentácie vyšších a nižších kognitívnych procesov majú prienik. Tento pohľad pochopiteľne stojí v ostrom kontraste so symbolizmom, ktorý predpokladá autonómnosť mentálnych reprezentácií spojených s vyššou kogníciou, a chápe ich ako amodálne symboly. Mentálne operácie sú realizované napríklad vo vnútornom jazyku mysle, napr. mentálčine ([49]). Naopak, teória perceptuálnych symbolov ([6]) ponúka riešenie v podobe multimodálnych subsymbolových reprezentácií, ktoré sú ukotvené v modalitách a nepodliehajú tak problému transdukcie (prenosu z modálnej úrovne do amodálnej). Inými slovami, perceptuálny symbol je v podstate koncept, a úloha symbolov sa redukuje na ich priradenie ku konceptom (napr. slová k objektom, či matematické symboly

k matematickým konceptom). Operácie s perceptuálnymi symbolmi prebiehajú tiež v priestore modalít.

3 Počítanie z pohľadu paradigiem

Spomínané tri paradigmy ponúkajú rôzne odpovede na otázky spojené s modelovaním mentálnych procesov. Symbolizmus a konekcionizmus ponúkajú kvalitatívne rozdielne vysvetlenia, no niektorí odborníci hovoria o výpočtovej ekvivalencii medzi nimi. Základom ich argumentácie je to, že symbolový systém dokáže realizovať akýkoľvek algoritmus (lebo Turingov stroj je univerzálny výpočtový stroj) a že umelým neurónovým sieťam boli tiež pripísané univerzálne vlastnosti. Dvojvrstvová dopredná neurónová sieť dokáže aproximovať ľubovoľnú spojitú funkciu ([69]) a rekurentná neurónová sieť (t.j. so spätnou väzbou) je výpočtovo ekvivalentná univerzálnemu Turingovmu stroju ([131]). To znamená, že akýkoľvek algoritmus sa dá simulovať rekurentnou neurónovou sieťou, a naopak, akákoľvek neurónová sieť sa dá simulovať na bežnom počítači (ktorý môžeme chápať ako Turingov stroj). Toto je pravda, a bežne simulujeme konekcionistické modely na digitálnych počítačoch. Z toho však nevyplýva, že je jedno, či sa nejaké kognitívne procesy snažíme vysvetliť konekcionisticky alebo pomocou klasickej paradigmy. Jedno to nie je z viacerých dôvodov (uvádzaných nižšie), okrem iného aj preto, že tieto existenčné dôkazy neriešia otázku ako skonštruovať sieť (t.j. nastaviť jej architektúru a váh), tak aby realizovala daný algoritmus. Symbolisti môžu nastaviť parametre (pravidlá) svojich modelov ručne, v konekcionizme to prakticky možné nie je. Vysvetlenie fenoménov v kognitívnej vede často zahŕňa akvizíciu poznatkov a to nie je možné riešiť bez modelovania zmien v systéme (jeho pamäti).

Musíme si vysvetliť koncept počítania, a že počítanie v neurónovej sieti nemusí byť to isté ako počítanie v symbolovom systéme. Otázka počítania začala byť vedecky aktuálnou krátko potom, keď bol v 40-tych rokoch minulého storočia vynájdený digitálny počítač, ktorý inšpiroval vznik teórie vypočítateľnosti a počítačovej vedy. Tie poskytujú teoretické poznatky o tom, čo to znamená počítať na Turingovom stroji. Počítanie v symbolovom systéme v podstate znamená vykonávanie programu, realizujúceho nejaký algoritmus sériovým spôsobom⁷. Pracujeme v diskretnom čase, s diskretnými entitami, symbolmi. Aj človek bežne pri počítaní má na mysli diskkrétne výpočtové úkony s číslami, ktoré vedú k výsledku.

V prípade neurónových sietí však tiež predpokladáme, že počítajú. Je povaha ich počítania odlišná od symbolových systémov? V kognitívnej vede prevláda názor, že ľudská myseľ realizuje výpočty, a teda spracováva informácie, pričom my sa tieto procesy snažíme vysvetliť pomocou formálnych modelov. Viacerí zástancovia konekcionizmu zastávajú názor, že mozog vykonáva výpočty a že neurálne počítanie vysvetľuje ľudskú kogníciu (napr. [45],[21],[8],[102]). V čom sa líšia názory konekcionistov a symbolikov je otázka, či konekcionistické modely sú lepšími výpočtovými modelmi než klasické symbolové modely. Väčšina konekcionistov tvrdí,

⁷ Samozrejme, paralelizácia výpočtu sa v bežných počítačoch dá dosiahnuť, ale tento fakt nie je z konceptuálneho hľadiska podstatný. Ako teoretický koncept, Turingov stroj pracuje sériovo.

že áno, symbolisti zase, že nie, lebo obe vysvetlenia sú vzájomne transformovateľné (opierajúc sa o spomenuté dôkazy o ekvivalencii). Niektorí symbolici tvrdia, že neurónové siete nepočítajú vôbec, alebo ak, tak len vtedy, keď implementujú symbolový systém ([50]). Niektorí prívrženci konekcionizmu zase zastávajú názor, že mozog nie je výpočtovým systémom (napr. [33],[54]), a že naše výpočtové modelovanie je len jeho diskretnou aproximáciou.

3.1 Je mozog výpočtový stroj?

Zastavme sa najprv pri tejto otázke. Otázka, či počítač je výpočtovým strojom je triviálna, pretože počítač bol navrhnutý ako výpočtový stroj (a stanovil tak klasickú definíciu počítania). V počítači prebieha spracovanie informácie pomocou elektrických aktivít (prostredie je homogénne) a implementačná úroveň, na ktorej výpočty prebiehajú, je jednoznačne určená. Sú to logické hradlá, ktoré sa preklápajú medzi dvoma možnými diskretnými stavmi. V prípade mozgu však implementačná úroveň vôbec nie je jednoznačná, pretože mozog nikto nezostrojil za účelom počítania. Ten sa evolučne vyvinul tak, aby umožnil organizmu efektívne fungovať v dynamickom prostredí ([21]). Na akej úrovni potenciálne prebiehajú výpočty v mozgu? Najpriateľnejšou hypotézou sa javí úroveň buniek mozgu (neurónov), hoci existujú neurofyziologické poznatky o tom, že nižšia (molekulárna) úroveň fyzikálnej reality ovplyvňuje činnosť samotných neurónov. Na druhej strane, na každom jednom neuróne nezáleží⁸, tak možno vyššia úroveň (neurálnych zhlukov) by bola vhodnejšia a efektívnejšia. Okrem toho, mozog pracuje „v zajatí“ všadeprítomných späťoväzobných slučiek, ktoré fungujú na rôznych časopriestorových škálach ([5]), a ktoré vnášajú do systému paralelizmus procesov operujúcich v spojitom čase. Napokon, prostredie v mozgu nie je homogénne, pretože spracovanie informácie má elektricko-chemický charakter: elektrický signál šíriaci sa pozdĺž axónu neurónu sa premieňa na chemickú difúziu na synapsách medzi neurónmi. Okrem toho, rôzne (pomalšie) hormonálne a neuromodulátorové systémy tiež ovplyvňujú správanie neurónu ([73]). Berúc do úvahy všetky tieto skutočnosti dospievame k presvedčeniu, že mozog nie je výpočtovým strojom v klasickom ponímaní, podobne ako mnohé iné komplexné prírodné systémy, ktoré majú nelineárny dynamický charakter. Napriek tomu, na ceste k jeho poznávaniu formálnymi metódami my nemáme inú možnosť, len aproximovať činnosť mozgu diskretizáciou v čase a v priestore (analogové počítače sa prakticky nepoužívajú).

3.2 Klasické a neklasické počítanie

Rozdiel medzi paradigmami reflektovaný aj v odlišnosti medzi klasickým a neklasickým počítaním. Neurónové siete nie sú homogénnou skupinou metód, pretože niektoré modely neurónových sietí počítajú klasicky, niektoré neklasicky a niektoré nepočítajú vôbec ([106]). Tieto kategórie modelov prirodzene vznikali aj chronologicky. Prvý model neurónových sietí pozostával z logických neurónov, ktoré

⁸ Je známe, že fyzicky blízke neuróny majú často identickú funkciu, čím vzniká redundancia i robustnosť systému.

pripomínali logické hradlá, a ktoré boli pospájané do vrstiev ([95]). Tento model sa nevedel učiť, pracoval s diskretnými hodnotami aktivít (0 a 1) a z výpočtového hľadiska je preto ekvivalentný so symbolovým systémom. Fungovanie takejto siete by sa dalo opísať programom, ktorý sieť realizuje.

Neskôr vznikli modely dopredných sietí na báze neurónov typu perceptrón ([125]) so spojitou aktivačnou funkciou, čím sa stavový priestor neurónovej siete stal spojitým. Tieto modely sa dajú trénovať pomocou malých zmien ich parametrov, čím sa stali použiteľnými pre riešenie rôznych úloh (v rámci vtedajších obmedzení). Učiacie pravidlo v adaptívnom systéme je jeho kľúčovým prvkom, pretože umožňuje implementovať *elementárnu formu zmeny znalosti* (internej reprezentácie časti vonkajšieho sveta), a to sa dá dobre realizovať v priestore numericky založených reprezentácií. V prípade týchto modelov už hovoríme o **neklasickom počítaní**, čo znamená, že model realizuje požadovanú funkciu – výpočet (t. j. nejakému vstupu priradí požadovaný výstup), bez toho, že by nutne vykonával program. Voľnejšie chápanie počítania teda nepredpokladá realizáciu programu. Napríklad, dopredná sieť sa vie naučiť násobiť dve viacciferné čísla bez toho, že by vypočítavala medzivýsledky, ako to bolo v prípade vyššie spomínaného algoritmu v symbolovom systéme. Výpočet prebehne v podstate na jeden krok: na vstup sa dajú čísla, aktivácia pretransformuje na výstup siete, čím sa generuje výsledok. Prečo učiacia sa neurónová sieť v princípe nevykonáva program vyplýva aj z toho, že pracuje dynamicky v spojitom priestore, a elementárne zmeny parametrov vedú k zmenám vo vstupno-výstupnom zobrazení, pričom takéto meniaci sa vzťah sa nedá redukovať na sériu diskretných operácií – potrebovali by sme nekonečnú presnosť. Avšak dá sa argumentovať, že tým že simulujeme neurónovú sieť pomocou symbolového systému (počítača), jej nedávame nekonečnú presnosť, a tak redukcia sa zabezpečiť dá. To je pravda, preto na ideálnu simuláciu neurónovej siete pracujúcej v spojitom priestore aktivít a váh by sme potrebovali analógový počítač, aby sme dosiahli nekonečnú presnosť pri simulácii dynamického procesu učenia.

Treťou kategóriou konekcionistických modelov sú tie, ktoré nielenže pracujú v spojitom priestore, ale aj v spojitom čase. Tieto modely nájdeme medzi umelými neurónovými sieťami, ale častejšie medzi sieťami s impulznými neurónmi ([86]). V tomto prípade už **nemôžeme hovoriť ani o neklasickom počítaní**, pretože počítanie si vyžaduje diskretný čas. Z rovnakého dôvodu tu nehovoríme ani o reprezentáciách, pretože tie tiež predpokladajú diskretný čas. Presná simulácia takýchto modelov sa tiež dá dosiahnuť iba na analógových počítačoch, kde sa všetky zmeny hodnôt dejú plynulo.

Takže konceptuálny rozdiel medzi symbolovým systémom a neurónovou sieťou závisí od toho, aký model siete máme na mysli, a tento rozdiel sa týka matematického opisu, nie implementácie. Implementácia akejkoľvek neurónovej siete v digitálnom počítači znamená jej diskretizáciu v čase a stáva sa tak aproximáciou modelu, ktorá môže byť viac alebo menej presná. Simulácia neurónovej siete už pochopiteľne je príkladom klasického alebo neklasického počítania, v závislosti od modelu. Súčasne 32, resp. 64-bitové počítače dosahujú vysokú presnosť, a problém by mohol nastať len v prípadoch, ak by sme potrebovali modelovať zložité dynamické procesy s chaotickým správaním, čo sa týka rekurentných sietí. V prípade väčšiny úloh tento

problém nemáme, lebo pracujeme s výstupmi siete, kde využívame istú robustnosť. To znamená, že na presnej hodnote výstupu nezáleží, a v prípade malých odchýlok sa interpretácia výstupu nezmení (napr. pri spomínanom násobení).

4 Podobnosti a rozdiely medzi paradigmami

V kontexte modelov spojených neurónových sietí je dynamický prístup úplne konzistentný s rekurentnými neurónovými sieťami pracujúcimi v spojitom čase a priestore. Okrem toho, obe paradigmy predstavujú príklady *kolektívnych systémov* pozostávajúcich z množstva interagujúcich prvkov, čoho výsledkom sú *emergentné javy* sprevádzajúce dynamiku systému. V čom sa obe paradigmy líšia, je (1) dimenzia systému (t.j. stavového priestoru), ktorá býva v prípade neurónových sietí (daná počtom neurónov) omnoho vyššia, a (2) fakt, že len pri neurónových sieťach boli navrhnuté rôzne mechanizmy učenia. V neurónovej sieti teda existujú dve, vzájomne sa ovplyvňujúce dynamiky, operujúce na rôznych časových škálach – pomalá adaptačná a rýchla aktivačná dynamika. V dynamických systémoch sú síce významné fázové prechody (externe spôsobené zmeny parametrov ktoré vedú k zmene dynamiky), ale ako k nim prirodzene dochádza, resp. ako zabezpečiť adaptáciu je už zložitejšie. Dôraz ostáva na aktivačnej dynamike. Neexistencia učiacich mechanizmov v dynamických systémoch je ich slabou stránkou. Keďže dimenzia oboch prístupov je výrazne rozdielna, zaujímavé prepojenie oboch paradigiem by mohlo napríklad spočívať v tom, že učiacia sa rekurentná neurónová sieť by mohla slúžiť ako „substrát“ pre hľadanie dynamického systému, ktorý operuje na hypotetickom stavovom priestore s nižšou dimenziou⁹. To by sa dalo teoreticky dosiahnuť identifikovaním kolektívnych premenných na základe monitorovania aktivácií existujúcich neurónov (napr. ich zoskupovaním). Podobne, napríklad meranie aktivity mozgu pomocou EEG prístroja ponúka niekoľkokanálové signály odpovedajúce aktivite mozgu na vyššej úrovni než je úroveň neurónov.

Čo má klasická paradigma spoločné s konekcionistickými modelmi pracujúcimi v diskretnom čase, a čím sa líšia od dynamickej paradigmy, je existencia vnútorných reprezentácií (no s rozdielnymi vlastnosťami). Tie sú v prípade klasickej paradigmy bežne dizajnované ručne ako symboly, resp. symbolové štruktúry, a v prípade neurónovej siete sú dizajnované len čiastočne (napr. požadované výstupy siete), pretože sieť si vytvára ešte svoje vnútorné reprezentácie. Neurónová sieť sa bežne učí na základe vopred pripraveného prostredia, avšak, dá sa principiálne používať spôsobom, aby interagovala s prostredím (či už simulovaným alebo fyzickým, ak jej implementovaná v agentovi), čím by si mohla svoju reprezentáciu prostredia zjednávať. V tom prípade by už išlo o tréning, kde by sa tréningové vstupy siete vytvárala sama vďaka svojej aktivite. Napríklad, robot riadený neurónovou sieťou, ktorý sa učí interagovať s objektmi vo svojom okolí upriamovaním pozornosti rôznymi

⁹ Dynamika vysokorozmerných komplexných systémov (s dimenziou danou počtom neurónov v sieti) nezriedka prebieha v nízkorozmernom nelineárnom podpriestore, čo by mohlo zodpovedať opisu pomocou dynamického systému v priestore tejto nižšej dimenzie.

smermi počas interakcie pripravuje nové perceptuálne vstupy (tie teda nie sú vopred dané).

Niektorí zástancovia tzv. radikálnej tézy dynamickej paradigmy ([142]) zdôrazňujú jej odlišnosť od symbolizmu a konekcionizmu práve kvôli jej nevýpočtovému charakteru, s čím súvisí predpoklad o neexistencii (mentálnych) reprezentácií. Zdôvodňujú to tým, že dynamický systém vyvíjajúci sa v čase sa nedá implementovať pomocou diskretných výpočtov, a preto nemôže podporovať ani žiadne vnútorné reprezentácie vonkajšieho sveta. Tu však možno namietat, že tento pohľad je odôvodniteľný len v kontexte modelovania senzomotorickej interakcie človeka s prostredím, zúženej na koordináciu pohybov, kde možno naozaj nepotrebujeme vnútorné reprezentácie (hoci aj o tom sa dá debatovať). Ťažko si ho však predstaviť v širšom kontexte kognitívnych procesov človeka, ktoré si vyžadujú existenciu mentálnych reprezentácií, či už vyvolateľných (cued) vonkajším podnetom, alebo oddelených (detached), t.j. vyvolateľných vnútorne ([58]). Aproximácia spojitých dynamických systémov pomocou modelov s diskretným časom je schodnou cestou, ktorá napr. celkom dobre funguje aj v biologicky prijateľnejších modeloch neurónových sietí s impulznými neurónmi ([71]). Preto je reprezentačno-výpočtový prístup kompatibilný s mäkkšou verziou dynamickej paradigmy, resp. S pohľadom na mozog ako na samoorganizujúci sa systém ([76],[37]). Podobnú optiku vidieť aj u menej radikálnych zástancov dynamickej paradigmy ([113]).

Môžeme konštatovať, že konekcionistická a dynamická paradigma zdieľajú (spolu s pravdepodobnostnými, bayesovskými prístupmi) viacero spoločných charakteristík a možno dôjde k zblížovaniu týchto paradigiem vo výpočtovej kognitívnej vede ([87]). Symbolová paradigma stojí vo výraznom kontraste s týmito paradigmami, no pravdepodobne si zachová svoju rolu „hrubozrnného,“ no zrozumiteľného opisu kognitívnych fenoménov. Ten možno obohatiť o detailnejšie, štatisticky orientované konekcionistické vysvetlenie mechanizmov napríklad pomocou neurónovej siete ([1]). Napríklad, vysvetlenie tvorby minulého času anglických sloviac pomocou jedného pravidla a výnimiek je transparentným symbolovým opisom, ktorý možno obohatiť o vysvetlenie procesu akvizície ako ho pozorujeme u detí (časť 5.5).

4.1 Rerezentácie

Podobne ako v prípade rozdielneho chápania pojmu *počítanie* by bolo dobré si vysvetliť dôležitý koncept *reprezentácie*. Tento pojem (reprezentácia) sa bežne používa nielen v umelej inteligencii, ale aj v kognitívnej psychológii, či lingvistiky. Jazykový pôvod tohto pojmu napovedá, že ide o re-representovanie niečoho vonkajšieho (vo svete) niekde inde, vo vnútri nejakého systému (živého alebo umelého) alebo i na papieri. V kognitívnej psychológii sa hovorí o *mentálnych reprezentáciách*, čo je v podstate prvý teoretický konštrukt v kognitívnej vede. Mentálna reprezentácia je základným konceptom výpočtovej teórie mysle, podľa ktorej sú kognitívne stavy a procesy konštituované výskytom, transformáciou a uchovávaním (v myšli/mozgu) reprezentácií (t.j. štruktúr nesúcich informáciu) nejakého typu ([111]). Korene pojmu mentálna reprezentácia však siahajú až do antických čias, keď širšie chápanie tohto

pojmu medzi filozofmi neznamenal výpočtový charakter¹⁰. Až neskorší predpoklad, že reprezentácia je (abstraktný) objekt so sémantickými vlastnosťami, dala tomuto konceptu výpočtový charakter, pretože s takto definovanými objektmi by sa už podľa predpokladu dalo manipulovať.

S nástupom umelej inteligencie sa hlavným cieľom stala **reprezentácia znalostí**, t.j. snaha o formalizáciu znalostí takým spôsobom, aby sa dali uložiť a používať v počítači (symbolovom systéme). Tu zohrali dôležitú úlohu lingvistika a logika, ktoré ponúkli existujúce formalizmy na rôzne implementácie reprezentácie znalostí. Môžeme povedať, že všetky znalosti (tzv. báza znalostí) sú z pohľadu klasickej paradigmy **explicitné a symbolové**, a keď systém disponuje pravidlami odvodzovania (inferenčné mechanizmy), dokáže si odvodiť ďalšie znalosti. Dôležitou vlastnosťou reprezentácií v symbolovom systéme je, že sú stabilné. Môžu sa síce v čase meniť (ak je to súčasťou programu), ale možno povedať, že sú väčšinou kontextovo nezávislé. Napríklad, koncept „strom“ je zväčša reprezentovaný rovnakým amodálnym symbolom (alebo štruktúrou) vo všetkých prípadoch, a ak nie, tak ku symbolu reprezentujúceho strom treba pridať iný symbol (alebo štruktúru), ktorý daný kontext špecifikuje (pozri Harnadovu definíciu symbolového systému v časti 2.1).

Otázka reprezentácií je relevantná aj v prípade konekcionizmu, kde má však inú povahu. Keďže vieme, že dlhodobou pamäťou natrénovanej siete sú spojenia medzi neurónmi i so svojimi hodnotami, znalosť je v sieti distribuovaná, a reprezentovaná **implicitne a numericky**. Analýzou natrénovanej siete však vieme zväčša identifikovať, čo konkrétne neuróny a spojenia medzi nimi zabezpečujú, aj keď komplikáciou je práve distribuovanosť informácie. Okrem toho, interné reprezentácie v neurónovej sieti bývajú prirodzene kontextovo závislé (napríklad, pozri príklad Elmanovej siete v časti 5.6 , ktorá si vytvorí vnútorné kontextovo-závislé sémantické reprezentácie slov). Kontextová závislosť v kontexte distribuovaných systémoch má zaujímavú implikáciu, pretože sa môžeme pýtať na mieru distribuovanosti. Existuje v mozgu neurón, ktorý páli **vtedy a len vtedy**, keď na senzorickom vstupe (nie nevyhnutne vizuálnom) je nejaký strom? Táto úvaha vedie k tzv. bunkám starej mamy (grandmother cells), ktoré túto podmienku spĺňajú (daný neurón reprezentuje moju starú mamu, t.j. je jedinečný). Máme viacero dôvodov predpokladať, že neurónov s touto vlastnosťou je v mozgu pravdepodobne veľmi málo (medzi také môžu patriť napr. bunky, ktoré detegujú konkrétne ľudské tváre). Takže reprezentácie v kontexte neurónových sietí je lepšie chápať ako vnútorné aktivácie systému pri danom vstupnom podnete v danom kontexte. Doslovné chápanie pojmu reprezentácia je mätúce.

Radikálna téza v doméne dynamických systémov nepredpokladá existenciu reprezentácií. Dynamické procesy pracujú v spojitom čase, a preto v žiadnom časovom okamihu nemôžeme predpokladať, že niečo z vonkajšieho sveta by mohlo byť aktuálne reprezentované vnútri systému, pretože všetko sa stále mení. Antireprezentačný postoj sa objavil aj v umelej inteligencii, keď Brooks ([13]) prišiel s alternatívou k tradičnej

¹⁰ No už bolo možné hovoriť o reprezentačnej teórii mysle, ktorá pracovala s konceptami ako sú myšlienky, presvedčenia, túžby, či obrazy, a ktorá sa v kognitívnej vede voľne zamieňa s výpočtovou teóriou mysle.

symbolovej architektúre. Tzv. *subsumpčná architektúra*, aplikovaná v oblasti robotiky, zahŕňa hierarchické usporiadanie vrstiev, ktoré zabezpečujú organizáciu správania robota podľa priorít (napr. cieľ robota dostať sa na nejakú pozíciu môže byť prechodne podradený cieľu dobiť si akumulátor). Vrstvy medzi sebou komunikujú, pričom využívané výpočtové prostriedky sú symbolové. Napriek tomu si takáto architektúra nevyžaduje uchovávanie vnútorných reprezentácií, pretože správanie robota je vždy len reakciou na vonkajší podnet (ide o reaktívny systém). Tento trend v robotike mal snahu priblížiť umelú inteligenciu ku kognitívnej vede, aj keď nezachytáva aspekty vyššej kognitívnej činnosti, ktorá sa nezaobíde bez vnútorných mentálnych reprezentácií.

Dynamická paradigma je teda výstižná pre opis tej časti prostredia, s ktorou človek môže interagovať, a na ktorú má tak povediac „dosah.“ Medzi percepciou a motorikou je istá asymetria v tom zmysle, že človek dokáže spracovávať signály z celého pre neho vnímateľného sveta (vd'aka senzorum), no motorický repertoár a jeho dosah je obmedzený. Napriek tomu si vytvárame vnútorný model prostredia obsahujúci aj tie nedostupné časti, v prípade ktorých už nemožno hovoriť o nejakej slučke. Takže mentálne reprezentácie zrejme majú buď senzomotorické charakteristiky alebo len perceptuálne charakteristiky, čo je v súlade s teóriou ukotvenej kognície, a čo sa premieta aj do reprezentácie významov ([44]).

4.2 Významy a sémantika

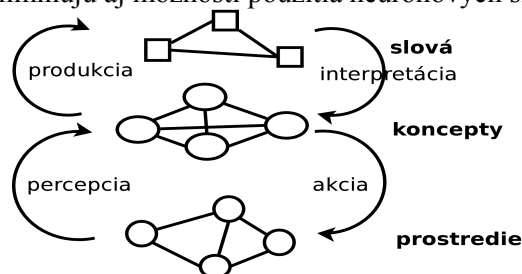
Otázka vlastností mentálnych reprezentácií a s ňou súvisiaci problém významov (sémantiky) patrí ku kľúčovým otázkam, ktoré sa kognitívna veda snaží riešiť. V tejto súvislosti boli identifikované dve kľúčové otázky: (1) ako sú významy vnútorne reprezentované a (2) či sú ukotvené vo svete. Podľa niektorých autorov sú tieto otázky nezávislé ([126],[134]). Otázka povahy vnútorných (mentálnych) reprezentácií závisí od paradigmy (hlavne symbolová verzus subsymbolová) a dá sa na ňu pozeráť ako na pretrvávajúci spor o tom, či myseľ reprezentuje významy pomocou (symbolových) propozícií ([117]) alebo pomocou (subsymbolorých) obrazov ([79]). Táto dichotómia je konzistentná s tvrdením, že (1) tradičné symbolové teórie uzatvárajú významy do separátnej sémantickej pamäti (uloženej niekde v mysli), ktorá je založená na amodálnych (t.j. modálne nezávislých) symboloch a operáciách s nimi, a že (2) teórie ukotvenej kognície a jazyka predpokladajú, že významy musia byť reprezentované v rôznych modalitách, v rámci senzomotorickej skúsenosti človeka so svetom ([7]).

Problém ako ukotviť významy identifikoval Harnad ([65]) ako **problém ukotvenia symbolov**. Formulácii tohto problému predchádzali dva iné významné míľniky v histórii kognitívnej vedy. Prvým bol Turingov ([143]) návrh, ako operacionalizovať otázku zistenia inteligencie stroja pomocou tzv. imitačnej hry, ktorá sa v literatúre stala známou ako **Turingov test**¹¹. Tento test vnukol predstavu, že myslenie (a teda aj ľudskú kogníciu) možno chápať ako manipuláciu so symbolmi.

¹¹ Jednoduchšia forma Turingovho testu predpokladá, že človek komunikuje v prirodzenom jazyku s neznámym systémom cez terminál, a jeho cieľom je zistiť (na základe odpovedí), či za systémom je skrytý človek alebo počítač.

Druhým míľnikom bola kritika tohto myšlienkového experimentu, podľa ktorej iba na základe pozorovania vonkajšieho správania (napr. odpovedí systému), nemožno určiť či systém naozaj rozumie tomu, o čom hovorí ([130]). A ak nerozumenie, potom nemôže byť inteligentný. Searlov **argument čínskej izby** ilustruje situáciu, pri ktorej sa navonok môže zdať, že osoba zavretá v izbe, komunikujúca s okolím cez terminál, rozumie významom svojich odpovedí, pričom je zrejme (nielen nám, ale aj samotnej osobe v izbe), že tomu tak nie je¹². Problém čínskej izby naráža na otázku vlastnej (intrinsic) sémantiky (významu). Argument spočíva v tom, že ak je význam symbolov v symbolovom systéme systémom nevlastný (v porovnaní s vlastnými významami v našich hlavách), tak to nemôže byť prijateľný model pre sémantiku podobnú tej ľudskej. Kognícia preto nemôže byť manipuláciou so symbolmi ([65]), pretože tie nie sú ukotvené. Základnou otázkou ukotvenia je, ako vytvoriť interný mechanizmus prepojenia mentálnych reprezentácií s vonkajším svetom takým spôsobom, aby mentálne reprezentácie boli systému vlastné, teda nezávislé na interpretácii externým pozorovateľom ([152]).

Problém ukotvenia symbolov je vo svojej podstate zložitejší, než sa na prvý pohľad zdá, a k jeho riešeniu pristupovali viacerí autori so svojimi modelmi (pozri prehľad v [149]). Tadeo a Floridi ([139]) prišli s návrhom kritéria tzv. nulového sémantického záväzku, ktoré podľa nich nespĺňa žiadne z existujúcich modelov, a preto dospeli k záveru, že problém ukotvenia symbolov sa doteraz nepodarilo vyriešiť. Avšak spomínané kritérium sa javí ako príliš reštriktívne, pretože maximálne zužuje úlohu dizajnéra ([148]). Napríklad, kritérium nepripúšťa ani existenciu učiacich mechanizmov v systéme (čo v prípade umelých systémov predstavuje účasť dizajnéra), čím sa principiálne eliminujú aj možnosti použitia neuronových sietí.¹³



Obr. 2. Vzťah medzi objektmi prostredia (referentmi), konceptmi a symbolmi jazyka. V symbolovom systéme sú koncepty amodálnymi symbolmi, ktoré vznikajú transdukciou z percepcie objektov vo svete, a ktoré sa mapujú na slová jazyka (symboly). Úloha ukotvenia sa týka samotných konceptov, pričom ich prepojenie s lingvistickým systémom je už symbolovo-symbolové. V prípade ukotveného systému sú koncepty modálne (napr. perceptuálne symboly) a preto sú priamo ukotvené v senzomotorickom systéme. Slová sú súčasťou lingvistického systému, ktorý je prepojený na konceptuálny systém a úloha ukotvenia symbolov (jazyka) spočíva v konštrukcii tohto prepojenia.

¹² Osoba A ktorá nehovorí po čínsky je zamknutá v miestnosti a komunikuje cez terminál s osobou B vonku, ktorá hovorí po čínsky. Predpokladáme však, že osoba A má k dispozícii kompletnú sadu pravidiel na manipuláciu s čínskymi znakmi, ktorá umožňuje osobe A odpovedať zmysluplne na ľubovoľnú otázku od osoby B.

¹³ Okrem toho, ako správne poznamenáva Vavrečka, také kritérium nespĺňa ani učiace sa dieťa.

Obr. 2 ilustruje vzťah medzi vonkajším svetom a internou štruktúrou agenta, ktorý disponuje konceptuálnym systémom a lingvistickým systémom (jazykom). Požiadavky na ukotvenie sa v takomto systéme uplatňujú rôzne, podľa toho či ide o symbolový systém alebo ukotvený. Rozumnou požiadavkou pre ukotvenie významov (napr. slov) je, že musia byť prepojené na okolitý svet cez senzomotorický systém (napr. vnímanie červeného objektu spôsobí aktiváciu konceptu červenej farby a prípadne aktiváciu slova červená), a že takéto prepojenie s prostredím sa vytvorí agent/človek sám pomocou svojich interných mechanizmov ([141]). Tým sa stávajú pre neho jedinečné, aj keď samozrejme musia byť o rôznych agentov dosť podobné, aby mohli úspešne komunikovať. Podobne to robí dieťa (pozri časť 2.4), ktoré si v ranom štádiu vývinu vytvára vnútorné reprezentácie fyzických objektov pomocou ich percepčných charakteristík a tzv. afordancií, čiže možností, ako s objektmi manipulovať ([61]). O významoch objektov môžeme hovoriť aj bez nutnosti ich pomenovávať (čiže používať jazyk), ako napríklad u detí v predverbálnom štádiu vývinu alebo v prípade zvierat ([44]).

Treba si uvedomiť, že samotné symboly existujú aj v prostredí, nielen v našich hlavách, vieme ich externalizovať a ukladať v rôznych formách, či už na papieri, v počítači, v obrazovej forme alebo ako v podobe ľubovoľných znakov (samozrejme, interpretácia symbolu bez kognitívneho systému nie je možná, takže tú hlavu potrebujeme). Napríklad, pri čítaní textu symboly priamo spracovávame a snažíme sa ich mapovať na koncepty. Podobná situácia nastáva, keď sa dieťa učí nové výrazy v prítomnosti rodiča. To znamená, že proces ukotvenia kognícia pripúšťa účasť dizajnéra (napr. rodiča) ako súčasť prostredia v procese ukotvenia symbolov (čo však nepripúšťa spomínané kritériom nulového sémantického záväzku).

Ukotvenie významov je výzvou pre kognitívnu vedu. Konekcionistické modely majú dve výhody: (1) využívajú subsymbolové reprezentácie (ktoré môžu plniť úlohu perceptuálnych symbolov), a (2) vedia sa učiť.¹⁴ Zostávajúcou kritikou konekcionizmu môže byť, že sa trénujú na vopred pripravených dátach. To znamená, že neurónová sieť síce používa distribuované subsymbolové reprezentácie, ktoré sú ukotvené v prostredí, na základe ktorého vznikli, avšak tie boli predpripravené externe, čím bola určená ich interpretácia. Inými slovami, takýto systém si nezjednáva svoj svet, lebo je mu vopred daný v podobe podnetov a odpovedajúcich významov. Napriek tomu môžeme povedať, že významy má ukotvené, ak symboly už existujú v prostredí a treba si ich osvojiť (napr. matka učí svoje dieťa pomenovávať objekty). Tento proces zahŕňa dva kroky: vytvorenie konceptuálnych reprezentácií a ich prepojenie na jazykové symboly. Toto je iná situácia než tie, keď uvažujeme populáciu agentov, ktoré sa naučia vzájomne komunikovať tak, že si vytvoria nový spoločný jazyk ([134],[140]). Tam významy ani pomenovania nie sú vopred dané, ale emergujú počas interakcií medzi agentmi.

¹⁴ Prvú podmienku nespĺňajú symbolové systémy (teda drvivá väčšina z nich, ak pripustíme existujúce formy adaptačných mechanizmov v týchto systémoch, napr. v podobe aktualizácie bázy znalostí). S druhou podmienkou majú problém dynamické systémy, pretože sa neučia.

4.3 Učiace sa systémy

Aby sme to zhrnuli, z hľadiska typu výpočtov ako aj typu reprezentácií, konekcionalizmus tvorí akýsi **most medzi klasickou paradigmou a dynamickou paradigmou**, aj keď k dynamickým systémom má oveľa bližšie než ku klasickým systémom. Čo však jedine konekcionalizmus ponúka, a čo je kľúčové, sú biologicky inšpirované mechanizmy učenia. Vďaka existencii týchto mechanizmov sa dajú neuronové siete využiť v návrhu umelých systémov, ktoré na istej úrovni abstrakcie nielen simulujú ľudské správanie, ale aj vývin kognície na ontogenetickej škále.

Tu treba podotknúť, že metódy učenia používané v neuronových sieťach sú len podmnžinou **metód strojového učenia** (napr. [91]), ktoré sa búrlivo rozvíjajú aj v novej umelej inteligencii. V tomto kontexte niektorí autori hovoria o **výpočtovej inteligencii**, ktorá si našla miesto na zemi, a ktorá sa chce odlišovať od tradičnej umelej inteligencie práve dôrazom na algoritmy učenia ([26],[32]). Nie je prekvapujúce, že neuronové siete sú považované za hlavnú súčasť metód výpočtovej inteligencie.

V ďalšom texte budeme hovoriť o neuronových sieťach, pričom budeme mať na mysli umelé neuronové siete, ktoré sa líšia typmi aktivít i svojou dynamikou od impulzných neuronových sietí, prednostne používaných vo výpočtovej neurovede ([86]). Tie sú ešte vernejšou aproximáciou biologickej reality než umelé siete, a vzťahujú sa na ne všetky kategórie učiacich mechanizmov. Výber úrovne aproximácie závisí od konkrétneho cieľa, pričom aj umelé neuronové siete ponúkajú vysvetlenia, konzistentné s neurobiologickými princípmi na istej úrovni abstrakcie (napr. pomocou impulznej neuronovej siete možno reprezentovať mentálny stav spojený s rozpoznávaním nejakého objektu ako synchronne oscilácie v nejakej populácii neuronov, čomu v umelej neuronovej sieti odpovedá aktivita jedného alebo viacerých výstupných neuronov, podľa toho či uvažujeme lokalistické alebo distribuované reprezentácie).

Konekcionalizmus zažil prvý boom v druhej polovici 80-tych rokov, keď objavenie algoritmu učenia pomocou spätného šírenia chyby ([127]) pomohlo prekonať 15-ročný útlm súvisiaci s problémom naučiť vtedajšie modely (jednovrstvové siete) riešiť ťažšie (nelinéarne) klasifikačné úlohy.¹⁵ Niektorí priekopníci avizujú ďalšie možné oživenie v tejto výskumnej oblasti. Geoffrey Hinton z Univerzity v Toronte argumentuje úspešnosťou viacvrstvových **generatívnych** modelov neuronových sietí (napr. model DBN, [68]).¹⁶ Jürgen Schmidhuber z IDSIA, známeho výskumného inštitútu v Lugane vo Švajčiarsku, zase predpovedá možnú renesanciu neuronových sietí vďaka ich čerstvým úspechom s viacvrstvovými modelmi v klasifikačných úlohách (pozri časť 5.4). Súčasne však možno pozorovať, že od 90-tych rokov nastal v konekcionalistickom tábore čiastočný presun v orientácii na iné, štatisticky založené metódy učenia (nie primárne biologicky inšpirované), napr. bayesovské modely ([63]),

¹⁵ Tento algoritmus bol vynájdený už o viac ako dekádu rokov skôr, no na publicitu si zrejme musel počkať ([67]).

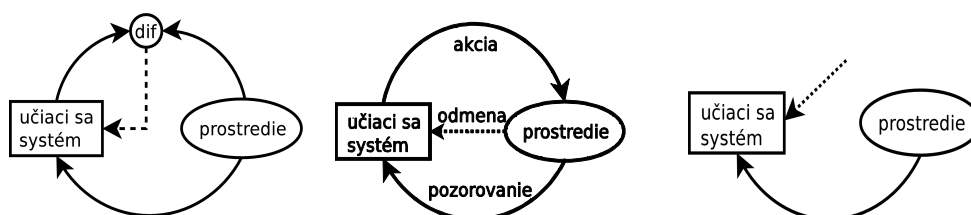
¹⁶ Generatívne a diskriminačné modely tvoria dve hlavné triedy učiacich sa klasifikátorov. Diskriminačné modely (väčšina modelov neuronových sietí) pasívne reagujú na daný vstup a generujú výstup, zatiaľ čo generatívne modely zahŕňajú aj koncept očakávaní (expectations), ktoré pôsobia zhora nadol prostredníctvom spätných váhových spojení.

ktoré tiež ponúkajú kognitívne relevantné vysvetlenia na vyššej úrovni abstrakcie. Je reálne predpokladať, že oba prúdy budú ďalej napredovať súbežne ([59]).

5 Neurónové siete v kognitívnej vede

Možno vyššie spomínané fakty čitateľa presvedčili o tom, že modelovanie ľudskej kognície spojenej s učením je principiálne ťažké v klasickej paradigme. Pochopiteľne, snahy o klasické prístupy budú pretrvávajúť a sú opodstatnené, napr. v oblasti budovania softvérových systémov, ktoré budú slúžiť človeku, ako napríklad znalostné/expertné systémy pre rôzne domény ([75]) alebo sémantický web ([28]). My sa budeme venovať neurónovým sieťam a ich využitiu v kontexte kognitívnej vedy.

Ako sme spomínali, kľúčovou vlastnosťou neurónových sietí, o ktorú sa kognitívna veda môže opierať, sú **mechanizmy učenia**. Tie môžeme rozdeliť do troch skupín (**Obr. 3**), a to v závislosti od toho, s akou mierou explicitnosti sa siete podáva informácia o chybe na výstupe. Pri **učení s učiteľom** (supervised learning) každý neurón má k dispozícii takúto informáciu, a mení svoje vstupné váhy tak, aby sa jeho chyba zmenšovala. Výstupné neuróny dostávajú túto informáciu priamo, a pre skryté neuróny sa dá táto chyba vypočítať, ako to bolo odvodené v známom pravidle učenia pomocou spätného šírenia chyby ([127]). Druhú skupinu tvorí **učenie s posilňovaním** (reinforcement learning), kde len sieť ako celok dostáva z prostredia informáciu (nie nevyhnutne v každom kroku) v podobe skalárnej veličiny o tom, „ako sa jej darí“ (tento typ učenia sa u nás tiež nazýva učením „pomocou odmeny alebo trestu,“ čo vystihuje význam spätnej väzby). Napokon, pri **učení bez učiteľa** (unsupervised) sieť nemá k dispozícii nijakú explicitnú informáciu o chybe, učí sa pomocou samoorganizácie. To znamená, že jej správanie závisí od konkrétneho učiaceho pravidla, ktorý vedie k nejakej forme extrakcie štatistických korelácií vo vstupných dátach. Všetky tri skupiny učiacich pravidiel sú pravdepodobne relevantné pre rôzne časti mozgu, pričom najväčší význam pre mozgovú kôru ako predpokladaný hlavný neurálny korelát kognitívnych funkcií, má asi samoorganizácia ([30]).



Obr. 3. Schémy troch paradigiem učenia: (a) s učiteľom, (b) s posilňovaním, (c) bez učiteľa.

Správna forma učenia vedie k schopnosti siete **zovšeobecňovať** (generalizovať) svoje znalosti a správne reagovať aj v nových situáciách. Neurónová sieť hľadá počas učenia štrukturálne vzťahy na (trénovacích) príkladoch a túto znalosť potom uplatňuje pri testovaní. Tu platia poznatky známe z výpočtovej teórie učenia (platné aj pre iné adaptívne modely), ktoré dávajú návod, ako sieť učiť a vyhnúť sa preučeniu (čo vedie k slabému zovšeobecňovaniu).

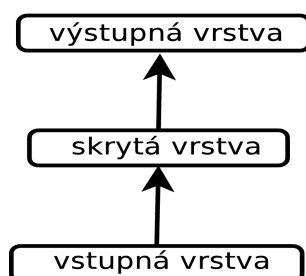
Prirodzenou aplikačnou oblasťou je vývinová kognitívna robotika (pozri prehľad napr. v [4]), ktorá využíva **konštruktivistický prístup** k budovaniu umelých agentov

(robotov), odlišný od tradičnej dizajnerskej umelej inteligencie, s cieľom vybaviť agentov vhodnými učiacimi mechanizmami, ktoré by im umožnili nadobúdať a využívať znalosti počas ontogenézy ([105]).¹⁷ V tejto oblasti nastáva oživenie, hnané víziou, že modelovanie kognitívnych procesov nemožno dosiahnuť bez prepojenia na nadobúdané senzomotorické zručnosti vyvíjajúceho sa vteleného agenta, situovaného v sociálnom prostredí ([104]). Pri návrhu mechanizmov učenia sú užitočné poznatky z kognitívnej neurovedy a vývinovej kognitívnej psychológie, ktoré ponúkajú empirické dáta. Táto oblasť je v podstate veľmi široká, pretože spektrum zručností, ktoré sa malé dieťa musí naučiť, je veľmi široké, spomeňme napríklad vnímanie objektov a ich rozpoznávanie, vizuálno-motorická koordinácia, referenčné rámce využívané pri priestorovej kognícii, zdieľanie pozornosti, ale aj tie vyššie ako prepojenie kognície na jazyk, ukotvenie významov, problém systematickosti v jazyku, a iné. My sa pozrieme konkrétnejšie na niektoré z nich.

5.1 Vizuálno-motorická koordinácia

Jednou z prvých zručností, ktoré si malé dieťa osvojuje, je senzomotorická koordinácia (pozri prehľad niekoľkých modelov v [41]). Jej skorou fázou je vizuálno-motorická koordinácia, kde možno rozlíšiť dva hlavné podciele: dosahovanie objektov (reaching) a uchopovanie objektov (grasping). Táto činnosť spočíva v snahe agenta dať do súvisu vizuálnu informáciu o cieľovom objekte (a prípadne koncového bodu ramena) s vnútornou proprioceptívnou informáciou o pozícii ramena. Existuje veľa behaviorálnych, neurofyziologických a zobrazovacích poznatkov o tom, že mozog si konštruje interné predikčné modely na tento účel ([74]). Tradičná teória vizuálno-motorického riadenia predpokladá koexistenciu dopredných a inverzných modelov, ktoré spolupracujú. **Dopredné** interné modely predpovedajú senzoricke dôsledky agentom vykonávaných akcií, t.j. ich vstupom je motorický signál a výstupom senzoricke reprezentácia („kde budem vidieť moju ruku, keď s ňou vykonám konkrétnu akciu?“). **Inverzné** interné modely fungujú naopak, vypočítavajú potrebné motorické pokyny na dosiahnutie cieľa, t.j. ich vstupom je senzoricke informácia a výstupom sú motorické akcie („aký pohyb musím vykonať, aby som sa dotkol rukou objektu?“). Oba modely sa ovplyvňujú a môže ich byť viac. Výstup inverzného modelu môže slúžiť ako vstup dopredného modelu, ktorého význam podľa súčasných teórií spočíva v napomáhaní pri potrebe korekcie trajektórie (napr. v prípade čiastočného prekrytia cieľového objektu, ktoré preruší vizuálny vstup) a v kompenzácii oneskorenia spôsobeného vizuálno-motorickou slučkou ([29]). Inverzný model využíva vonkajší oblúk cez prostredie, dopredný nie. Je prirodzené oba modely realizovať pomocou neurónových sietí, pretože vizuálno-motorická koordinácia sa deje v spojitom prostredí. Typickým prístupom je použitie dopredných sietí učených s učiteľom (**Obr. 4**), pretože v prípade inverzného modelu sa predpokladá dostupnosť chyby z prostredia, a v prípade dopredného modelu sa chyba vypočítava vnútorne, porovnaním senzorickej predikcie s tou skutočnou.

¹⁷ Je potešiteľné, že aj v bývalom Československu sa myšlienky kognitívnej robotiky objavovali už pred 30-imi rokmi (Havel, 1980).



Obr. 4. Dvojvrstvá¹⁸ dopredná neurónová sieť (s neurónmi typu perceptrón). Šípky označujú úplné prepojenie medzi susednými vrstvami, t.j. od každého neurónu nižšej vrstvy vedie váhované spojenie ku každému neurónu v nasledujúcej vrstve (čím dostávame dve matice parametrov). Počty vstupov a výstupných neurónov sú vo všeobecnosti rôzne, závisia od zvoleného kódovania použitých dát. Počtom skrytých neurónov určujeme zložitosť modelu. Takáto sieť je vhodná pre úlohy, ktoré sa dajú formálne opísať ako hľadanie nelineárneho vzťahu medzi vstupmi a výstupmi na základe predložených párov (vstup, požadovaný výstup). Perceptróny sa najčastejšie učia pomocou metódy spätného šírenia chyby, ktorá podľa gradientu chyby v priebehu učenia minimalizuje rozdiely medzi skutočnými a požadovanými výstupmi.

Pri voľbe učiacej paradigmy hrá rolu aj kognitívna prijateľnosť modelu, to znamená realistikosť predpokladov, na ktorých je model založený. Napríklad, v kontexte vizuálno-motorickej koordinácie môže byť predpoklad o dostupnosti chybového signálu otáznym, pretože to predpokladá znalosť o požadovanej motorickej aktivite (t.j. ako keby matka učila dieťa siahať na objekty tak, že by mu vzala ruku do svojej). Preto prijateľnejšou alternatívou (aj keď zložitejšou z hľadiska učenia) je učenie s posilňovaním, pretože tu sa predpokladá „mäkšia“ verzia spätnej väzby. Tou môže byť napríklad vzdialenosť medzi rukou a objektom (ktorú vizuálny systém odhaduje), ktorá nehovorí priamo, ako pohnúť rukou (systém to musí vypočítať; pozri časť 5.2). V kontexte kognitívnej robotiky je dôležitým krokom to, že boli navrhnuté rozšírenia konceptu učenia s posilňovaním (pôvodne navrhnutého pre diskretnú doménu) aj pre spojité domény, pre ktoré sú neurónové siete vhodné ([31]). V kontexte učenia s posilňovaním to znamená, že diskretné tabuľky sa nahradia aproximátormi funkcií (realizované napr. doprednými sieťami), ktoré umožňujú generalizáciu na nové vstupy. Príkladom tohto prístupu je nami simulovaný a vylepšovaný model trénovaný pomocou algoritmu CACLA ([146]), ktorý využíva dve dopredné siete: jednu na generovanie akcií (tzv. aktér) a druhú na odhad očakávanej odmeny v jednotlivých stavoch, konkrétne pozícií ruky od cieľa (tzv. kritik). Model sa nám podarilo nastaviť tak, že sa dokáže naučiť trajektórie z ľubovoľného počiatočného do ľubovoľného koncového bodu v pracovnom priestore s veľkou presnosťou ([78]).

Pochopenie mechanizmov spojených s predvídaním a ohodnocovaním motorických akcií je veľmi dôležité, nakoľko tieto typy mechanizmov sa môžu podieľať aj na vyššej kognitívnej činnosti (napr. plánovanie a jazyk) a pravdepodobne mozog využíva rovnaké alebo podobné mechanizmy aj na tejto úrovni abstrakcie. Napríklad, dopredné modely sa pravdepodobne podieľajú pri porozumení

¹⁸ Vstupnú vrstvu nepovažujeme za vrstvu neurónov. Sú to v podstate len receptory, a nemajú aktivačnú funkciu, ktorá realizuje nelineárnu funkciu medzi vstupom a výstupom neurónu.

pozorovaným akciám iných agentov v rámci systému tzv. zrkadliacich neurónov ([97]). Taktiež sa predpokladá, že slúžia ako neurálny korelát vlastnej vôľovej činnosti subjektu („sense of agency“), ktorý mu umožňuje odlíšiť seba od iného agenta ([27]).

5.2 Referenčné rámce

Použitie doprednej neurónovej siete priamo na prevod medzi vizuálnymi súradnicami (na vstupe) a motorickými súradnicami (na výstupe) je najjednoduchší spôsob, ako to dosiahnuť. Takéto modely ponúkajú vysvetlenie, „ako to funguje“ na pomerne vysokej úrovni abstrakcie, ak si uvedomíme, že v mozgu tomu odpovedá komunikácia medzi tromi lalokmi: spánkovým, kde sa spracovávajú vizuálne informácie, predným, kde sa spracovávajú motorické informácie, a temenným, kde sa realizuje prevod medzi **referenčnými rámcami** (súradnicovými systémami). Temenný lalok teda predstavuje oblasť, kde sa robia prevody medzi perceptuálnymi súradnicami a motorickými signálmi, ktoré treba vygenerovať, aby sa rameno (či noha) dostalo na požadované miesto v 3D priestore. Tento proces je však zložitejší, ak si uvedomíme, že mozog využíva viacero referenčných rámcov – napríklad retinocentrický (pozícia objektu na sietnici oka), hlavocentrický (pozícia objektu vzhľadom ku hlave) a i. – a prepočítava súradnice medzi nimi (napr. [114]). To je pri priestorovom vnímaní potrebné pri mnohých úlohách, napríklad na to, aby sme pri pohybovaní hlavy s pohľadom zafixovaným pre seba dopredu (t.j. bez pohybu očí) dokázali vnímať nejaký objekt v zábere (receptívnom poli) ako statický napriek tomu, že jeho súradnice na sietnici sa budú meniť.

Takéto prepočty súradníc medzi okom, hlavou a telom, experimentálne objavené u makakov v posteriornej časti temenného laloka (Brodmannova area 7a), sa dajú vysvetliť pomocou doprednej neurónovej siete s jednou skrytou vrstvou neurónov (**Obr. 4**). Prvý takýto model ([153]) bol trénovaný práve na takýto prevod súradníc medzi sietnicou a hlavou, a to tak, že na vstup dostával podnety s retinocentrickými súradnicami ako aj informáciu o polohe oka v rámci hlavy (t.j. smer pohľadu) a na výstupe generoval pozíciu objektu odozvy s hlavocentrickými súradnicami. Zaujímavé bolo zistenie, že sieť rieši úlohu na skrytej vrstve podobným spôsobom ako mozog. Skryté neuróny si vytvorili svoje vlastné receptívne polia, ktoré sa kvantitatívnym spôsobom líšili v citlivosti daného neurónu buď na jeden vstup (pozícia objektu na sietnici) alebo na druhý vstup (smer pohľadu oka vzhľadom ku hlave). Lineárnou kombináciou aktivít skrytých neurónov sieť dosiahne na výstupe pozíciu objektu v priestore, nezávislú (invariantnú) od natočenia hlavy i nasmerovania očí. Tieto rozmanité samoorganizované receptívne polia charakterizujú **priestorový zisk** (spatial gain) neurónu a zabezpečujú vlastne moduláciu odozvy neurónu (jeden vstup moduluje citlivosť neurónu na druhý vstup, resp. naopak). Neuróny s takýmito charakteristikami boli objavené už na viacerých miestach v mozgu a **ziskové polia** sú považované za univerzálny mechanizmus neurálnych výpočtov v mozgu ([129]).

Analogicky možno pomocou doprednej neurónovej siete modelovať prepočítavanie medzi perceptívnymi súradnicami a motorickými súradnicami (pozícia ruky) pri zámere siahnuť na objekt ([25]). Dokonca nedávno bol navrhnutý model opäť na báze dvojvrstvového perceptrónu, inšpirovaný novými neurovednými poznatkami,

ktorý implementuje **dvojitú moduláciu** ziskových polí (compound gain field), pretože pozícia cieľového objektu vzhľadom k pozícii ruky sa počíta z pozície tohto objektu vzhľadom k oku, modulovanej pozíciou oka a súčasne aktuálnou pozíciou ruky ([18]).

Človek na reprezentáciu priestorových vzťahov teda využíva rôzne referenčné rámce, všetky uvádzané príklady sa týkali vlastnej (ego) perspektívy. Dieťa najprv pozná len **egocentrický rámeč** , a až neskôr sa naučí chápať priestorové vzťahy aj z relatívnej perspektívy (**alocentrický rámeč**), napr. z pozície iného človeka alebo pozorovaného objektu. Pre úplnosť dodajme, že existuje aj absolútny referenčný rámeč, ktorý znamená priestorovú orientáciu človeka podľa svetových strán. Jednotlivé rámce v mozgu koexistujú, spolupracujú a subjekt medzi nimi dokáže prepínať ([14]). Okrem toho, používanie referenčných rámcov závisí od kultúry a je ovplyvnené aj jazykom ([88]). Neurálne modelovanie týchto procesov je len v začiatkoch. Napríklad, jeden model využíva atraktorové neurónové siete so spojenou dynamikou, kde dlhodobá pamäť sa modeluje pomocou alocentrických reprezentácií a krátkodobá pamäť pomocou egocentrických reprezentácií, modulovaných smerovaním vizuálnej pozornosti ([15]). Najnovšie poznatky o prepínaní medzi rámcami pochádzajú aj zo štúdie selektívnej vizuálnej pozornosti, ktorá sa bežne považuje za egocentrický mechanizmus. Frischen, Loach a Tipper ([55]) pomocou experimentov poukazujú na to, že pozorovateľ dokáže aktivovať aj ekvivalentný alocentrický mechanizmus, pomocou ktorého dokáže cielene zameriavať vizuálnu pozornosť na rôzne objekty z perspektívy iného subjektu.

5.3 Zrkadliace neuróny

S problematikou referenčných rámcov súvisí schopnosť človeka (alebo zvierat'a) rozumieť správaniu iného jedinca. Schopnosť predpovedať kroky pozorovaného pohybu (napr. človek pozoruje iného človeka, ako siaha po šálke) by sa dala vysvetliť pomocou „prenesenia sa do kože aktéra.“ Toto vysvetlenie pozostáva z dvoch krokov: najprv sa subjekt „prepne“ do perspektívy pozorovaného subjektu (aktivuje sa mu alocentrický rámeč), a potom simuluje sledovaný pohyb, namapovaný na mechanizmy vlastnej vizuálno-motorickej koordinácie, ktoré má už vyvinuté na základe predošlej vlastnej skúsenosti. V tomto kontexte treba spomenúť existenciu tzv. **zrkadliacich neurónov** , objavených najskôr u makakov ([121]) a len nedávno u človeka ([99]), ktoré sú považované za dôležitý komponent neurálnych obvodov umožňujúcich spektrum schopností človeka v oblasti sociálnej kognície, počnúc porozumenia pozorovaným akciám, schopnosťou imitovať, až po čítanie mysle a porozumenia emóciám pozorovaného subjektu ([56]). Je možné, že zrkadliace neuróny zohrali úlohu aj v evolúcii jazyka ([3]).

Tieto sociálne orientované schopnosti nie sú pravdepodobne človeku vrodené, musí sa ich naučiť, čo sa deje počas prvých rokov života. Tieto procesy musia vedieť vysvetliť aj výpočtové modely. Existujú viaceré modely na báze neurónových sietí, ktoré simulujú proces porozumenia akciám, no s využitím viacerých zjednodušení. Jedným z nich je napríklad to, že využívajú objektocentrický rámeč, vďaka ktorému predpokladajú, že vykonávanie pohybu a i jeho pozorovanie generujú tie isté súradnice (čiže nezávislé od agenta). Otázka modelovania porozumenia motorickým akciám nie

je v súčasnosti uspokojivo vyriešená, pretože existujú rôzne vysvetlenia. Všetky modely síce riešia prepojenie senzorických a motorických signálov (s využitím inverzného a prípadne i dopredného modelu), ale robia to iným spôsobom, za iných predpokladov a s využitím inej paradigmy učenia (všetky tri paradigmy majú svoje zastúpenie). Kritický prehľad konceptuálnych a výpočtových modelov, hlavne na báze umelých neurónových sietí možno nájsť v práci Farkaš, Malý a Rebrová, [43]. Aj keď to zatiaľ tak nie je, dá sa predpokladať, že modelovanie zrkadliacich neurónov sa stane súčasťou vývinovej kognitívnej robotiky.

5.4 Rozpoznávanie objektov

Túto schopnosť dieťa nadobúda pomerne skoro (počas prvého roku života). Vie sa, že vizuálny systém rieši úlohy typu „čo“ a „kde“ rôznymi časťami mozgu, ktoré sú však prepojené ([144]). Inými slovami, „where“ systém, situovaný v temenných lalokoch (v oboch hemisférach) rieši otázky pozície objektu (v rôznych referenčných rámcoch spomínaných vyššie) bez jeho identifikácie, a naopak, „what“ systém sa orientuje na identifikáciu objektu (na ktorý máme zameranú pozornosť), nezávisle od jeho presnej pozície¹⁹, veľkosti a natočenia. Úlohou určovania pozície objektu sme sa zaoberali v predchádzajúcej časti, teraz sa pozrime na úlohu rozpoznania. Invariantné rozpoznávanie je prirodzenou požiadavkou nielen z pohľadu potrebnej úspešnosti rozpoznania, ale aj z hľadiska kognitívnej relevancie: toto dokáže i dieťa. Jedným z prvých neurálnych modelov, ktorý čiastočne riešil otázku invariantnosti, a to hierarchickým spôsobom, bol štvorvrstvový perceptrón ([83]), trénovaný s učiteľom (štandardnou metódou spätného šírenia chyby) a testovaný na úlohe rozpoznávania rukou písaných číslíc (databáza MNIST). Riešenie problému spočíva vo využití dvoch princípov: architekturného a výpočtového. Architekturný princíp spočíva v tom, že každý neurón vyššej vrstvy v rámci svojej skupiny dostáva vstupy z časti receptívneho poľa neurónov nižšej vrstvy (t.j. časti vstupného priestoru, ktorú vidí) a susedné neuróny majú receptívne polia mierne posunuté (vidia iné časti vstupu), a to topografickým spôsobom. Druhým princípom je, že neuróny v rámci skupiny zdieľajú svoje dopredné váhy (weight-sharing), t.j. majú ich stále rovnaké, čím sa účelne redukuje počet voľných parametrov a podporuje schopnosť zovšeobecnenia.²⁰ Tieto dva princípy zabezpečia, že skupina neurónov začne fungovať ako detektor nejakého príznaku (napr. čiary nejakej orientácie) v rôznych pozíciách na sietnici. Ak máme viacero skupín na danej vrstve vedľa seba, tie sa naučia detegovať iné príznaky. Rôznorodosť príznakov dosiahne sieť sama v rámci učenia, čo je peknu ilustráciou toho, ako učiaci signál „naviguje“ detektory príznakov k ich špecializácii (ktorá závisí od počiatočných hodnôt váh, a tie bývajú náhodné). Ak nejaký neurón ešte vyššej vrstvy vidí časť neurónov nižšej vrstvy, stáva sa invariantným detektorom tohto

¹⁹ Tu mám na mysli relatívne malé posuny v rámci uhla pohľadu, keď sa na objekt pozeráme, no každý malý posun znamená veľké zmeny v aktiváciách receptorov na sietnici, preto predstavuje výpočtový problém.

²⁰ Toto je z výpočtového hľadiska šikovný krok, no z biologického hľadiska je takáto úprava ťažko prijateľná, aj keď susedné neuróny v kôre mávajú veľmi podobné charakteristiky (vrátane synaptických účinností), čím je zabezpečená redundancia a robustnosť.

príznaku. Tento proces sa opakuje aj vo vyšších skrytých vrstvách, pričom narastá miera invariantnosti, a výstupná vrstva predstavuje kategórie vstupov (10 číslíc).

Od pôvodného modelu ([83]) bolo navrhnutých viacero vylepšení na báze neurónových sietí alebo iných metód strojového učenia, no pozoruhodným je nedávny model ([22]), ktorý svojou presnosťou (no nie časom tréningu) prekonal všetky predchádzajúce modely na vyššie spomínanej štandardnej množine rukou písaných číslíc (MNIST). Model doprednej siete pozostáva z väčšieho množstva vrstiev (autori ich skúšali až 9), no využíva len štandardné pravidlo učenia pomocou spätného šírenia chyby ([127]), ktoré viacerí už považujú za „zastarané“. Ako vysvitlo, ak sieť má dostatok času na tréning (urýchlenie výpočtu asi 50x realizovali pomocou grafických procesorov), dokáže sa dosť presne naučiť úlohu, až na úrovni ľudských schopností. Z pohľadu potrebného času a počtu príkladov však strojové učenie veľmi zaostáva za ľudskou kogníciou.

Ukazuje sa však, že návrh hlbokých hierarchických architektúr je asi cestou, ako dosiahnuť invariantné robustné rozpoznanie komplexných dát ([9]). Koniec koncov, hierarchické spracovanie informácie je inherentnou vlastnosťou aj architektúry mozgu, čo sa týka nielen percepcie ([46]), ale aj konania ([62]).

5.5 Akvizícia jazyka – lexikálna úroveň

Oblasť akvizície jazyka je zaujímavou doménou z toho dôvodu, že v sebe zahŕňa široké spektrum procesov, počnúc fonologickou analýzou, cez morfológickú a lexikálnu úroveň až po syntax a sémantiku viet.²¹ Okrem toho, niektoré aspekty jazyka (napr. gramatika) priamo nabádajú na symbolové prístupy k modelovaniu. Nie náhodou boli spomedzi počítačových modelov jazyka donedávna dominantnými práve symbolové modely. Tie dokážu transparentne opísať niektoré aspekty (v podstate všetky tie, ktoré sa dajú zachytiť pravidlami), no na druhej strane nevedia vysvetliť niektoré kognitívne javy vznikajúce pri používaní jazyka. Preto sa popri klasických modeloch začala v 90-tych rokoch rozvíjať **konekcionistická psycholingvistika** ([20]) v oblasti modelovania jazyka, ale aj v kontexte ostatných kognitívnych procesov ([119],[96]). Prirodzený jazyk pritom vzbudil najviac záujmu, ale aj kontroverzie ([108]). Podľa niektorých kritikov sa totižto (minimálne) niektoré aspekty jazyka vôbec nedajú vysvetliť konekcionisticky, ale len pomocou symbolových modelov. Jadro kontroverzie súvisí s už spomínanou otázkou, aká je povaha mentálnych procesov prebiehajúcich v mozgu pri vyšších kognitívnych procesoch ([108]): či dokážu konekcionistické modely iba implementovať symbolové modely (**implementačný konekcionizmus**), alebo či mentálne procesy možno zredukovať na subsymbolovú úroveň, a v tom prípade konekcionizmus bude znamenať alternatívnu paradigmu k vysvetleniu povahy jazyka (**eliminačný konekcionizmus**). Odpoveď na túto otázku je pri súčasnej úrovni poznania zložitá, pretože pre viaceré jazykové fenomény existuje konekcionistické i symbolové vysvetlenie. V ďalšej časti budem ilustrovať oblasť konekcionistickej psycholingvistiky na pár príkladoch.

²¹ Mohli by sme pokračovať nadvetnou úrovňou smerom ku pragmatike jazyka a diskurzu, ale pre naše prezentačné účely postačí, ak skončíme pri vetnej úrovni.

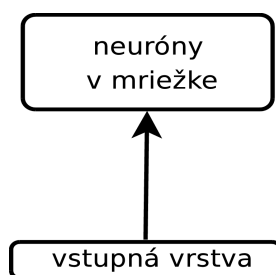
Detailnejší a obsažnejší prehľad modelov môže ctený čitateľ nájsť v prehľadových prácach ([19],[36],[123]).

V oblasti morfológického spracovania je najznámejším príkladom problém tvorenia minulého času slovíes v angličtine. Ako vieme, väčšina anglických slovíes získava v minulom čase koncovku /-ed/, avšak pri nepravidelných slovíesách sú minulé tvary rôzne. Pre tento jav existujú dve vysvetlenia ([110]). Klasické vysvetlenie predpokladá existenciu **dvoch ciest** v mozgu, jednu pre pravidelné slovíesá (aplikácia pravidla pridania koncovky /-ed/) a druhú pre nepravidelné slovíesá (tabuľka párov neurčitok - minulý čas). Algoritmus pre rozhodovanie je takýto: mám neurčitok, ak nájdem tvar v tabuľke, použijem odpovedajúci minulý čas, ak nie, aplikujem pravidlo pre pravidelné slovíesá. Na druhej strane, konekcionistické vysvetlenie predpokladá **len jednu** cestu, pretože dopredná neurónová sieť sa dokáže naučiť tvoriť minulý čas všetkých slovíes a zovšeobecňovať svoju znalosť pri predložení nových tvarov. Neurobiologické poznatky o mozgu sa zdajú byť skôr v prospech existencie dvoch ciest ([110]), no symbolový model nevie uspokojivo vysvetliť **proces akvizície** minulého času slovíes u dieťaťa, zatiaľ čo neurónová sieť áno (napr. [112]). Proces akvizície má tri typické fázy: najprv dieťa správne tvorí minulé tvary pravidelných i tých niekoľkých známych nepravidelných slovíes, potom však začne nesprávne aplikovať pravidlo /-ed/ aj na nové nepravidelné slovíesá, no napokon sa minulé tvary všetkých nepravidelných slovíes naučí správne.²²

Táto debata poukazuje na zaujímavý, všadeprítomný fenomén, s ktorým sa učiaci sa agent stretáva. Veľa dát, vrátane tých lingvistických, má vlastnosť tzv. **kváziregulárnosti**, čo znamená, že medzi dátami (či už medzi vstupmi, alebo v rámci vstupno-výstupných párov na tréning) existuje systematickosť (zákonitosť, pravidelnosť), ale aj viaceré výnimky (ako sme videli pri slovíesách). Na vysvetlenie takýchto vzťahov sa v symbolovom prístupe predpokladá, že systematickosť je reprezentovaná pomocou explicitných **pravidiel**, zatiaľ čo **výnimky** sa zabezpečujú asociačným mechanizmom ([110]). Pri konekcionistickom vysvetlení potrebujeme len jeden systém a všetky vstupné dáta sa navzájom líšia iba stupňom konzistentnosti z pohľadu vstupno-výstupnej transformácie. To znamená, že aj podobnosť (systematickosť) má gradovaný a nie binárny charakter (pravidlo/výnimka). V prípade nepravidelných anglických slovíes tiež možno nájsť ich vnútorné členenie na jednotlivé podskupiny (keep/kept, sleep/slept) s rôznymi veľkosťami až po veľmi nepravidelné slovíesá (napr. go/went). Táto kváziregulárnosť odzrkadľuje rôzne nároky na (nelineárnu) transformáciu pri generovaní požadovaného tvaru, od čoho potom závisí aj poradie, v ktorom sa sieť minulý čas jednotlivých slovíes naučí. Z pohľadu neurónovej siete teda ide o úlohu nelineárnej transformácie s rôznymi stupňami pravidelnosti. Podobný prípad kváziregulárnosti možno pozorovať i pri čítaní slov (systematickosť závisí od jazyka), spracovaní rečového signálu alebo pri artikulácii slov ([20]).

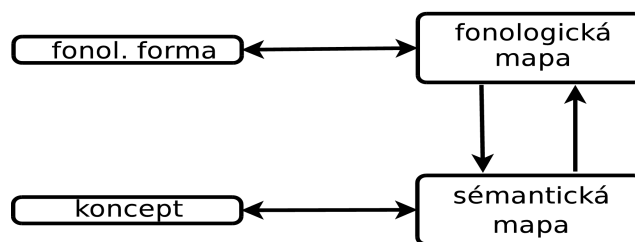
²² Symbolové vysvetlenie tohto procesu by sa dalo asi formulovať takto: v prvej fáze dieťa len memoruje minulé tvary, v druhej fáze začne aplikovať pravidlo /-ed/ a až v tretej fáze nájde rovnováhu medzi oboma cestami. Debata okolo tvorby minulého času má svoj vývoj a celá problematika je o čosi zložitejšia, než sme tu uviedli (viac možno nájsť v [36]).

Analógiou kváziregulárnosti je známa teória prototypov a základných kategórií ([124]), pomocou ktorej možno vysvetliť kategorizáciu objektov u človeka (a iných živočíchov). Stupeň príslušnosti objektu do danej triedy možno charakterizovať ako spojité (teda nie binárne ako v symbolovom systéme), čo závisí od prototypickosti objektu (príklad: vrabec je lepším prvkom kategórie vták než tučniak). Toto vysvetlenie má aj psychologickú podporu, pretože človek dokáže verbálne zhodnotiť, nakoľko sa mu pozdáva daný objekt ako prvok danej kategórie.



Obr. 5. Samoorganizujúca sa mapa (SOM). Vstupy sú úplne prepojené adaptovateľnými váhami so všetkými neurónmi v mape, usporiadanými do pravidelnej 2D mriežky. SOM zobrazí distribuovane reprezentované vstupy na lokalistické odpovede v mape (zobrazenie redukujúce dimenziu). Požadovanou vlastnosťou tohto zobrazenia je vlastnosť zachovania topológie: blízke vstupy sa zobrazia na blízke neuróny v mape. Takéto zobrazenie možno učením dosiahnuť práve vďaka existencii neurálnej štruktúry v mape (každý neurón „vie“, akých má susedov), čo sa využije pri mechanizmoch súťaženia a kooperácie medzi neurónmi. Reakcie neurónov SOM sa organizujú podľa štatistických vlastností vstupov.

Ako ďalší príklad spomeňme modelovanie tzv. lexikálnych máp, kde možno využiť neurónové siete so samoorganizáciou. Známy model samoorganizujúcej sa mapy ([77],[35]) umožňuje modelovať rôzne lexikálne efekty a je biologicky zaujímavý. SOM (**Obr. 5**) realizuje zobrazenie, ktoré zachováva vzťahy podobnosti medzi vstupnými dátami, čo je vlastnosť niektorých častí mozgovej kôry, kde máme napríklad rôzne senzorické mapy (napr. v doméne vizuálnej alebo akustickej), alebo motorické mapy, ktoré si mozog interne vytvára. Predpokladá sa, že v mozgu môžu existovať i mapy vyšších, abstraktnejších entít, napríklad konceptov ([120]), čo sa ponúka ako vysvetlenie kognitívnych porúch pozorovaných u pacientov, napr. neschopnosť pomenovať predmety patriace do nejakej kategórie. Vysvetlenie spočíva práve v predpoklade o topografickom usporiadaní odpovedajúcich neurónov v mape: simulácia lokálnej poruchy v mape s veľkou pravdepodobnosťou poškodí len neuróny v jednej časti mapy, zodpovedajúce nejakej kategórii objektov.



Obr. 6. DevLex model lexikálneho vývinu. Dve rastúce SOM sú vzájomne asociatívne prepojené. Fonologická mapa sa organizuje na fonologických symboloch a sémantická mapa na reprezentáciách konceptov. Všetky spojenia sa aktualizujú pomocou samoorganizácie.

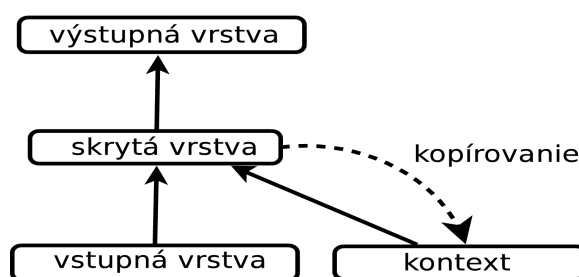
Niektoré modely založené na SOM majú komplexnejšiu architektúru, čo rozširuje ich použiteľnosť. Napríklad Dislex ([98]) predstavuje systém troch prepojených máp (fonologická a ortografická mapa sú prepojené so sémantickou mapou), čo umožňuje ich vzájomné aktivovanie. Tým sa dajú modelovať jazykové procesy, ako napríklad čítanie (ortograficko-sémantická linka) alebo artikulácia (sémanticko-fonologická linka), ale aj rôzne poruchy ako dyslexia alebo špecifické afázie (napr. poruchy pomenovania vecí), a to pomocou simulovaných lokálnych porúch v modeli. Na základe inšpirácie Dislexom navrhli Li, Farkaš a McWhinney ([84]) model DevLex (**Obr. 6**), pozostávajúci z dvoch navzájom prepojených rastúcich máp (fonologickej a sémantickej), ktoré majú tú výhodu, že nepotrebujeme vopred poznať vhodnú veľkosť siete, pretože tá sa upraví podľa dát. Rastúce siete sú ilustráciou maturačných procesov (funkčných a anatomických), ktoré sa predpokladajú v mozgoch cicavcov, a ktoré sa môžu týkať aj oblastí, spracovávajúcich jazyk ([118]). DevLex bol navrhnutý ako psychologicky prijateľný model vznikajúceho lexikónu u malých detí. Napríklad ukazuje, ako nielen gramatické kategórie (podstatné mená, slovesá), ale aj sémantické kategórie (napr. zvieratá, jedlo, atď.) postupne vznikajú v sémantickej časti modelu. DevLex taktiež vysvetľuje javy rôznych chýb pomenovania objektu, a to v závislosti od frekvencie slov a od sémantických podobností medzi slovami. Na druhej strane, DevLex je príkladom typického použitia neurónovej siete, kde vstupy z prostredia sú predpripravené, takže nemožno hovoriť o spätnoväzbovej slučke, ako je to v prípade učiaceho sa dieťaťa.

5.6 Akvizícia jazyka – vetná úroveň

Viaceré konekcionistické modely boli aplikované i na vetnej úrovni, kde je to už viac zaujímavé, lebo ku slovu sa dostáva syntax a sémantika²³. Prvé modely sa zaoberali akvizíciou gramatiky, a tu kľúčovú úlohu zohrala priekopnícka práca Jeffreyho Elmana ([34]), ktorá spustila „lavínu“ používania **rekurentných neurónových sietí** pri spracovaní sekvenčných dát v jazyku i mimo neho. Elman zaviedol novú paradigmu tréningu nasledujúcej položky (item) v danej vstupnej sekvencii symbolov, napr. slov v prípade viet. Elmanova jednoduchá rekurentná sieť (simple recurrent network, SRN) sa trénuje na vetách nejakého jazyka, vždy slovo po slove, a snaží sa predikovať

²³ O lexikálnej sémantike sme hovorili v prípade lexikálnych máp, kde sme predpokladali existenciu sémantických príznakov slov, vytvorených dizajnérom.

nasledujúce slovo. Slová boli kódované lokalisticky, t.j. každý neurón svojou aktivitou reprezentoval práve jedno slovo. Napríklad vo vete John broke glass, ak je aktuálnym vstupom John, sieť má predikovať slovo broke. V ďalšom kroku sa broke stáva vstupom, a požadovaným výstupom je glass, atď. Túto schopnosť má nepochybne aj človek, ktorý pri počúvaní reči dokáže s istou presnosťou predikovať, čo rečník povie. Je tomu tak vďaka syntaktickým, sémantickým a pragmatickým obmedzeniam, ktoré determinujú povrchovú štruktúru viet.



Obr. 7. Schéma jednoduchého rekurentného siete (SRN). V porovnaní s viacvrstvovým perceptrónom je obohatená o spätnú väzbu od skrytých neurónov. Plné šípky označujú úplné prepojenie medzi susednými vrstvami neurónov pomocou adaptovateľných váh. Čiarkovaná šípka označuje kopírovanie aktivít skrytých neurónov na kontextové vstupy, čím sa aktivity skrytých neurónov v aktuálnom kroku stanú kontextovými vstupmi v nasledujúcom kroku. Túto sieť možno trénovať aj pomocou štandardnej metódy spätného šírenia chýb, ale aj pomocou efektívnejších algoritmov. SRN je vhodná na niektoré časopriestorové úlohy, kde kontextové vstupy slúžia ako pamäť (s exponenciálnym zabúdaním do minulosti) a umožňujú sieti generovať výstup nielen podľa aktuálneho vstupu, ale aj podľa ľavého kontextu.

Aj keď predikcia nasledujúceho slova nie je finálnym produktom pri spracovaní jazyka (cieľom je porozumieť vete), predikcia je jedným z užitočných komponentov jazykového systému v ľudskom mozgu (človek takisto robí predikcie rôzneho typu, viz časť 5.1). Preto táto úloha vzbudila v 90-tych rokoch pozornosť i v konekcionistickej komunite. Z pohľadu paradigmy učenia ide o špeciálny typ učenia s učiteľom (**self-supervised**), pretože požadované výstupy pochádzajú zo vstupnej sekvencie. Elman (1990) aplikoval svoj model SRN na dve úlohy so sekvenčnými dátami. Prvou bola predikcia nasledujúceho (tlačeného) písmena v súvislom texte. Sieť teda v každom kroku prečítala aktuálne písmeno (symbol) a jeho ľavý kontext, a mala predikovať nasledujúcu písmeno (bez medzier medzi slovami). Elman ukázal, že predikcie začiatku nového slova boli vždy najhoršie. Tento prístup implikuje, že schopnosť predikcie nasledujúcej fonémy (v prípade hovorenej reči) môže byť primárnym mechanizmom, ktorý novorodenci používajú pri segmentácii slov ([128]).

Druhou úlohou bola predikcia nasledujúcich slov (tiež lokalisticky kódovaných) v korpuse jednoduchých viet (napr. John runs, Mary loves John, John broke glass, atď.). SRN sa naučila pomerne presne predikovať vhodných kandidátov v každom kroku (v jazyku danom stochastickou bezkontextovou gramatikou bola neurčitost', takže v každom mohlo nasledovať viacero slov vhodnej kategórie), čím sa teda naučila gramatiku jazyka. Naučila sa ju však implicitne, pretože nezískala explicitnú znalosť o pravidlách (na rozdiel od symbolových modelov, kde pravidlá gramatiky sú

explicitne dané). Schopnosť naučiť sa gramatiku sa Elmanovi podarilo vysvetliť, keď analyzoval interné reprezentácie po natrénovaní, ktoré si SRN vybudovala vo svojej skrytej vrstve.²⁴ Zaujímavým zistením bolo, že tieto interné distribuované reprezentácie mali svoju štruktúru, takže podobné slová (vyskytujúce sa v podobných kontextoch) mali podobné interné reprezentácie, čo pri ich zhľukovaní viedlo k odhaleniu jednotlivých syntaktických i sémantických kategórií slov. Naučená sieť sa teda naučila gramatiku, čo dosiahla vytvorením sémantických reprezentácií na skrytej vrstve slov na základe minimalizácie výstupnej predikčnej chyby (viac o sémantických reprezentáciách a významoch pojednávam v časti 5.7).

Ako bolo povedané, Elmanove práce zohrali v histórii rekurentných neurónových sietí dôležitú úlohu a SRN je dodnes najčastejšie používaným modelom rekurentnej siete²⁵. Predikčné modely sekvencií slúžili aj ako stimulácia pre empirické skúmanie a záujem o štatistické učenie, pretože sa ukázalo, že ľudia dokážu zachytávať štatistické závislosti na rôznych úrovniach. Táto schopnosť je často podvedomá (subjekt nevie, že sa dané závislosti učí, no jeho výsledky v teste sú toho dôkazom), a preto ide o **implicitné učenie** ([24]). To sa samozrejme netýka len jazyka, ale aj iných sekvenčných činností. V kontexte je jazyka je zaujímavé, že človek dokáže sledovať aj tzv. nesusedné závislosti ([101]), ako napríklad vo vzťažnej vete Janko, ktorý sedí vzadu, vidí mačku, kde v hlavnej potrebujeme naviazať sloveso vidí na podmet Janko. Táto schopnosť je nevyhnutná pri používaní jazyka, jej vysvetlenie závisí od pohľadu (zástanca symbolovej paradigmy by to vysvetlil tak, že subjekt využíva explicitnú gramatiku), no z hľadiska štatistického učenia nie je vôbec triviálna, pretože slovný materiál medzi cieľovými slovami je premenlivý a môže tak pôsobiť „rušivo“. Bolo však ukázané, že aj SRN sa dokáže za istých podmienok naučiť s istou presnosťou nesusedné závislosti v umelom jazyku (pozri [40] a tamojšie referencie).

Druhá línia konekcionistického výskumu súvisiaca so spracovaním jazykových sekvencií sa týka tzv. **syntaktickej systematickosti** jazyka (a mysle). Fenomén systematickosti ľudskej mysle sa priamo dotýka spomínaných paradigiem a možno ho vysvetliť pomocou analógie so schopnosťou človeka systematicky rozumieť jazyku. Ak niekto vie povedať alebo rozumie vete Janko ľúbi Aničku, bez problémov vie vytvoriť vetu Anička ľúbi Janka a porozumieť jej, pretože vie, že jazyk sa skladá z častí a vie, ako ich skladať pri tvorbe viet. Systematickosť v jazyku implikuje produktívnosť (schopnosť generovať teoreticky neobmedzené množstvo viet) a kompozičnosť (skladanie atomických znakov do zložených pomocou rekombinácie; viz. definícia symbolového systému v časti 2.1). V symbolových modeloch je problém systematickosti automaticky vyriešený, pretože používané reprezentácie a syntaktické operácie nad nimi (podľa explicitných pravidiel) priamo zabezpečujú

²⁴ Postup získania interných reprezentácií slov je nasledovný: po natrénovaní siete raz prejdeme tréningovou množinou a zaznamenáme aktivácie skrytej vrstvy pri každom vstupnom slove (v rôznych kontextoch). Potom aktivácie pre každé slovo spriemerujeme, čím dostaneme kontextovo-nezávislé reprezentácie slov (niečo ako prototypy pri kategorizácii), ktoré môžeme znázorniť v hierarchickom diagrame a vidieť tak podobnosti medzi slovami.

²⁵ V ostatných rokoch sa začal častejšie skúmať aj jednoduchší model, rekurentná sieť s echo stavmi (pozri prehľad v [85]), ktorá sa dá rýchlejšie trénovať a v niektorých prípadoch postačí ako efektívna náhrada SRN.

produktívnosť symbolového systému. V neurónových sieťach takéto vlastnosti vôbec nie sú zrejmé (sieť si nemôže explicitne uložiť gramatiku vo svojej pamäti), a podľa niektorých kritikov ani principiálne možné. Fodor a Pylyshyn ([50]) vo svojom vplyvnom kritickom článku sformulovali argumentáciu zhruba nasledovne: „Ľudská myseľ je systematická, a preto jej interné reprezentácie sú štruktúrované. Keďže v konekcionistických systémoch neexistujú štruktúrované reprezentácie, nie sú (tie systémy) dobrými modelmi ľudskej mysle.“

Vysvetlenie systematického správania stále predstavuje pre neurónové siete výzvu ([2]), aj keď nejaké lastovičky sa už objavili. Hadley ([64]) navrhol rozlíšenie medzi slabou a silnou syntaktickou systematickosťou. Neurónová sieť je slabo systematická, ak dokáže spracovať vety (t.j. správne predikovať) s novými kombináciami slov, pričom všetky slová sú na syntaktických pozíciách, na ktorých sa objavili už aj počas tréningu (napr. sieť tréningovaná na vetách chlapec ľúbi dievča a pes naháňa mačku dokáže spracovať aj vetu pes naháňa dievča). Na druhej strane, silná systematickosť už vyžaduje, aby sieť vedela generalizovať aj v nových syntaktických pozíciách (napr. spracovať vetu pes naháňa chlapca, za predpokladu že podstatné meno chlapec sa nikdy nevyskytlo počas tréningu v úlohe objektu len subjektu). Bolo ukázané, že vlastnosť tzv. slabej syntaktickej systematickosti dokážu nadobudnúť rekurentné siete s rôznymi architektúrami a algoritmami učenia: SRN, ESN (Frank, 2006a) a rekurentná sieť so samoorganizáciou ([42]). Vlastnosť silnej systematickosti bola preukázaná v prípade ESN ([52]) a SRN ([11]). Tieto pozitívne výsledky eliminačného konekcionizmu oslabujú kritiku, podľa ktorej aj keď sa podarí natréningovať nejaký konekcionistický model so systematickým správaním, to nestačí, pretože konekcionizmus „ako taký“ (vo všeobecnosti) by mal byť systematický; len vtedy ho môžeme považovať za kognitívne relevantný model ([50]). Treba však priznať, že dosiahnuté výsledky sa týkajú jednoduchších foriem jazyka, takže problém konekcionistickej syntaktickej systematickosti ešte nie je uspokojivo vyriešený.

5.7 Porozumenie vetám

Otázka významov a ich reprezentácie sa netýka len samotného jazyka, ale ide o širší koncept. Spomínali sme, že Elmanova SRN vytvárala (kontextovo-závislé) sémantické reprezentácie slov na skrytej vrstve, čo znamená, že jednotlivým aktivitám skrytých neurónov by sme mohli priradiť význam. Takýto spôsob reprezentovania významom je príkladom prístupu, ktorý môžeme nazvať štatistickým symbolizmom. Ide v podstate o učiace metódy, ktoré využívajú štatistické závislosti vo výskytoch slov (symbolov), teda ako slová po sebe nasledujú v nejakom korpuse (kovariácia). Inými slovami, význam slova, t.j. jeho sémantická reprezentácia, sa odvodí na základe výskytov tohto slova v rôznych kontextoch²⁶ (v prípade SRN ide o ľavý kontext, t.j. predchádzajúce slová). Tu však ostáva problémom ukotvenie symbolov (časť 4.2), pretože takéto významy sa vytvárajú súvzťažne, u každého slova pomocou ostatných slov, a preto tieto významy nie sú ukotvené v prostredí. Je to akoby sa niekto snažil učiť slová

²⁶ Kontext slova môže mať rôzne formy, napríklad môže ísť aj o počítanie výskytov slov v rôznych dokumentoch ([81]).

cudzieho jazyka len pomocou výkladového slovníka toho jazyka. Jediným zdrojom informácie je tu spomínaná kovariácia. Tým sa však tento spôsob tvorby významov odlišuje od tradičných symbolových systémov, kde sémantiku slov vkladá dizajner a kde kovariácia slov ako zdroj informácie nehrá úlohu.

V prípade konekcionistických modelov hrá štatistika kľúčovú úlohu. Prvé modely využívali vopred pripravené sémantické reprezentácie ako požadované výstupy, ktoré mala sieť priradiť vetám na vstupe. Napríklad, model McClellanda a Kawamota ([94]) je realizovaný ako štandardná dopredná sieť, ktorá sa učí priradiť sémantické roly (agens, paciens, inštrument, modifikátor) jednotlivým konštituentom vety (podmet, prísudok, predmet, predložková fráza „s“), čo možno považovať za dôležitý krok pri komprehenzii (porozumení). Hlavným obmedzením modelu je to, že významy viet sú vopred dané a predpokladajú sa len vety generované podľa niekoľkých rámcov (templates) s požadovanými sémantickými rolami. Napriek tomu úloha mapovania viet na významy nie je triviálna, lebo priradenie rolí závisí od kontextu (vo vetách *Ball hit girl* a *Ball moved* je ten istý podmet raz inštrumentom a raz paciensom), alebo závisí od sémantiky slovesa (vo vetách *Man ate pasta (with) spoon* a *Man ate pasta (with) cheese* v prvom prípade predložková fráza „s“ modifikuje paciensa, zatiaľ čo v druhom prípade vyjadruje inštrument). V niektorých prípadoch ostáva nejednoznačnosť nerozlúštená (napr. vo vete *Boy hit girl with ball* nevieme, či ball je inštrumentom alebo modifikátorom paciensa girl). Model však dokázal tieto nejednoznačnosti zohľadniť a správne zovšeobecňovať (aj keď len na malej množine viet). Tento model sme predstavili na ilustráciu raného prístupu k reprezentácii významov, ktoré nevyužívajú prvky situovanosti ani vtelenia, a naopak zdieľajú charakteristiky s tradičnými symbolovými systémami (napr. rámce). Na tento model nadviazali ďalšie vylepšenia smerom ku zložitejším konštrukciám a architektúram. Súčasnú „špičku“ konekcionistického modelovania komprehenzie viet z hľadiska zložitosti architektúry a používaného jazyka predstavujú pravdepodobne dva modely: CSCP ([122]) a INSOMNet ([92]).

Rohde ([122]) navrhol integrovaný model komprehenzie a produkcie viet (CSCP), ktorý sa trénuje na vetách (rôznej zložitosti) generovaných podľa gramatiky zjednodušeného anglického jazyka (ale pozoruhodnej zložitosti). Význam vety pozostáva z množiny propozícií, ktoré sú reprezentované distribuovane a sú explicitne prezentované sieti počas učenia. Jedna časť systému (sémantický systém) sa učí skomprimovať sekvenciu týchto propozícií do jednej, statickej reprezentácie významu celej vety. Model CSCP je postavený na báze modulov SRN ([34]), ktoré sa naučia zobrazit' vstupnú vetu (sekvenciu slov) na jej celkový význam (komprehenzia) ako aj opačným spôsobom, t.j. význam na povrchovú formu vety (produkcia). Dôležité je, že CSCP je schopný simulovať rôzne psychologické aspekty spracovania viet u človeka a poukazuje na dôležitosť prepojenia medzi komprehenziou a produkciou počas učenia jazyka. Diskutabilnou stránkou modelu je však práve trénovanie pomocou explicitných propozíčných reprezentácií, ktoré prebieha pomocou tzv. dotazovania (query) na jednotlivé konštituenty propozícií, čo je z kognitívneho hľadiska otáznave (tento prvok figuruje aj v starších, jednoduchších modeloch komprehenzie viet, zrejme ako nutná pomôcka na to, aby sa model dokázal vstupno-výstupnú transformáciu naučiť [36]).

Flexibilnejší model INSOMNet ([92]) nevyžaduje fixné rámce propozícií ako CSCP. Tento model s modulárnou architektúrou bol navrhnutý ako interpretátor anglických viet, ktorý predloženej vete v textovej forme (slovo za slovom) priradí jej sémantickú reprezentáciu. Reprezentácie viet s tzv. plytkou sémantikou, použité na tréning výstupného modulu s externým učiteľom, boli vzaté z lingvisticky anotovaného korpusu, v ktorom sémantika každej vety má formálne podobu orientovaného grafu. Takýto graf vyjadruje syntaktické a sémantické vzťahy medzi vetnými členmi, reprezentovanými distribuovane. Reprezentácie viet založené na tejto lingvistickej konceptuálnej štruktúre majú dve zaujímavé vlastnosti: (1) vďaka grafovej štruktúre sa vyhýbajú problému reprezentovania syntaktických stromov s veľkou hĺbkou (typických pre prirodzený jazyk v kontexte symbolových a štatistických modelov), a (2) podobne ako symbolové programy tiež využívajú smerníky, no tie sú implementované pomocou aktivácií viacerých neurónov a sieť sa ich musí naučiť ako súčasť reprezentácie vety. Model sa musí tiež naučiť gramatiku, pretože tá nie je explicitne daná. INSOMNet je psychologicky prijateľný model, lebo (podobne ako človek) sa učí na príkladoch, generuje predikcie, nemonotónne reviduje interpretáciu vety pri jej inkrementálnej analýze, správa sa robustne, dokáže paralelne vytvárať viacnásobné interpretácie v prípade nejednoznačných viet, a vie modelovať rôzne jazykové poruchy pozorované u ľudí.

V oboch spomínaných modeloch sa používajú sémantické reprezentácie viet vo formáte propozícií, čo znamená, že vieme lokalizovať jednotlivé konštituenty vety. Teórie porozumenia textu ([47]), podporené psychologickými experimentmi, potvrdzujú, že okrem propozíčnej úrovne človek využíva aj tzv. situačnú úroveň reprezentácie, ktorá už je jazykovo nezávislá a formuje sa nad propozíčnou úrovňou (textová báza). Tento poznatok slúžil ako inšpirácia pre model distribuovaného situačného priestoru (DSS,[53]), ktorý bol navrhnutý na reprezentáciu významov (situácií) na situačnej úrovni. Situačné vektory teda nemajú propozíčný charakter, pretože neuróny v DSS sa podieľajú na reprezentácií viacerých situácií, a reprezentácia každej situácie je distribuovaná. DSS reprezentácie situácií sú pripravené vopred pomocou samoorganizácie. Úloha neurónovej siete (SRN) spočíva v tom, že jednotlivým vstupným vetám, opisujúcim nejakú situáciu v uvažovanom mikrosvete má SRN priradiť ich situačný vektor, čo sa interpretuje ako pochopenie významu vety. SRN teda predpovedá danú situáciu, preskakujúc propozíčnú úroveň reprezentácie vety. SRN číta vetu slovo po slove, a ku koncu má vygenerovať situačný vektor. Samozrejmom požiadavkou je, aby sieť dokázala zovšeobecňovať, t.j. aby porozumela nielen novým vetám opisujúcim známe situácie, ale aj novým vetám opisujúcim nové situácie. Napríklad, v modeli existujú tri postavičky, ktoré môžu robiť rôzne činnosti na rôznych miestach (samostatne alebo spolu). Ak sa systém naučí porozumieť vetám Janko hrá šach vnútri a Janko sa hrá na schovávačku vonku, musí porozumieť aj vetám Janko hrá šach vonku a Janko sa hrá na schovávačku vonku (ak sú tieto situácie konzistentné s vlastnosťami mikrosveta). Táto schopnosť zovšeobecnovania vyjadruje vlastnosť tzv. **sémantickej systematickosti** ([50]). Situačné vektory v modeli DSS majú pekné vlastnosti, sú subsymbolové (forma vektora a jeho význam sú neoddeliteľné) a teda analógové, čím sú konzistentné s perceptuálnymi symbolmi ([6]). Tieto situačné vektory však nie sú modálne, a preto ani ukotvené. Okrem toho,

v situačných vektoroch neexistujú koncepty ani vzťahy medzi nimi (čo je typickou vlastnosťou propozičných vektorov). Model DSS od tohto abstrahuje, no ťažko si predstaviť mentálne reprezentácie situácií, kde nefigurujú koncepty.

5.8 Integrácia modalít vo vývinovej kognitívnej robotike

Ukazuje sa, že vo výpočtovom modelovaní v kognitívnej vede nastáva v súčasnosti oživenie a snaha o integrované ukotvené prístupy pri učení kognitívnych schopností s čo možno najväčšou mierou autonómnosti učiaceho sa agenta. Z dôvodu zložitosti realizácie týchto cieľov sa používajú zväčša jednoduchšie architektúry sietí a spôsoby učenia, nielen v snahe o vytváranie multimodálnych perceptuálnych reprezentácií, ale najmä v úlohách prepojenia jazykových schopností na ostatné kognitívne funkcie (prehľad modelov a výzvy pre neurorobotiku možno nájsť v [16]). Dôraz sa teda kladie na to, aby jazykový vývin šiel „ruka v ruke“ s kogníciou a nie izolovane, ako to bolo v modeloch uvedených v predchádzajúcich častiach o modelovaní akvizície jazyka (5.5-5.7). Typickým elementom týchto modelov je napríklad prepojenie motorickej aktivity s učením sa pohybových slovies, ktoré takto priamo nadobúdajú ukotvené významy (napr. anglické slovesá kick, pick, lick sa týkajú rozličných častí tela a vieme, že aj ich neurálne koreláty v mozgu sú odlišné, [116]). K tomu sa dá v najjednoduchšom prípade pristupovať pomocou jednej doprednej neurónovej siete, ktorá dostáva na vstup aktuálnu proprioceptívnu informáciu o svojom tele, napr. ramene, a snaží sa urobiť relevantnú akciu (natočenie ramena), pričom spracováva aj jazykový vstup, ktorý danú akciu špecifikuje ([89]). Takáto sieť sa dá trénovať aj pomocou bežného algoritmu spätného šírenia chyby. Zložitejším prístupom je prepojenie dvoch rekurentných sietí, senzomotorickej a jazykovej, ktoré sa súčasne trénujú na sekvenciách s využitím predikcií na požadovaných hodnotách v ďalšom kroku (uhly natočenia ramena robota, resp. ďalšie slovo vety) ([135]). Natrénovaná sieť potom dokáže napríklad vykonať požadovanú akciu na základe jej pomenovania, a to aj v prípade nových akcií – táto schopnosť zovšeobecnenia sa opiera o vlastnosť kompozičnosti (spomínanej v časti 5.6). Tieto prístupy však využívajú nerealistické predpoklady pri učení, ako napríklad existenciu požadovaných hodnôt pre pohyb ramena. Realistickejším prístupom je napríklad učenie s posilňovaním, kde spätná väzba z prostredia nie je taká explicitná (pozri časť 5.1) a teda agent musí prísť na to, aké akcie má vykonávať aby viedli k cieľu. Niektoré prístupy zase viac využívajú samoorganizáciu v senzorickej, motorickej i jazykovej doméne. Napríklad model Wernter a Willshaw ([150]) má pekné topografické vlastnosti (podobne ako senzorické i motorické časti mozgovej kôry) a môže slúžiť ako asociátor akcií a ich jazykového opisu, no chýbajú mu predikčné vlastnosti (spätná väzba). Možno povedať, že integračné prístupy s využitím neurónových sietí najmä v oblasti kognitívnej robotiky sú vo svojich začiatkoch, a až budúcnosť ukáže, ako sa podarí napredovať smerom k realistickejšej zložitosti (kognitívnych architektúr a prostredia).

6 Záver

V príspevku sme sa venovali výpočtovým prístupom v kognitívnej vede, so zameraním na konekcionizmus. Výpočtové modelovanie vo všeobecnosti možno považovať na nevyhnutný metodologický komponent, pretože ponúka explicitné mechanizmy, ktoré opisujú ako môže kognícia fungovať ([87]). Pri prechádzke paradigmami kognitívnej vedy sme sa snažili objasniť najmä rozdiely medzi symbolovou a konekcionistickou paradigmou, pričom sme spomenuli, že koncept počítania musíme chápať širšie než ponúka počítačová metafora mysle, pretože mozog nie je počítač a neurónové tiež počítajú, no nemusia pritom vykonávať programy. S využitím teoretickej argumentácie ako aj empirickej evidencie sme dospeli k záveru, že konekcionistické prístupy ponúkajú detailnejšie vysvetlenie kognitívnych mechanizmov, konzistentné s neurobiologickými princípmi na istej úrovni abstrakcie (ale značne nižšej než v prípade symbolových modelov). Zaujímavou vlastnosťou neurónových sietí je aj to, že tvoria akýsi most medzi symbolovou a dynamickou paradigmou (ktoré stoja v úplnom kontraste), pretože obsahujú niektoré charakteristiky oboch. Za hlavnú výhodu neurónových sietí považujeme ich paradigmy učenia s rôznym stupňom spätnej väzby z prostredia (učenie s učiteľom, s posilňovaním a so samoorganizáciou).

Využitelnosť umelých neurónových sietí sme ilustrovali na príkladoch zameraných na vybrané procesy spojené s raným kognitívnym vývinom. Pohľadom zdola sme sa snažili pokryť nielen úlohy nižšej kognície (napr. senzomotorická koordinácia, rozpoznávanie objektov) ale aj vyššej (jazyk, otázka syntaktickej a sémantickej systematickosti). Konekcionistický prístup v kognitívnom modelovaní je založený na predpoklade, že ľudská myseľ/mozog nie je hybridný systém, na ktorý potrebujeme nutne použiť dva opisy ([39]), ale že z evolučného hľadiska ide o homogénny systém, čo sa týka mechanizmov a mentálnych reprezentácií. Clark ([23], s. 135) hovorí v tejto súvislosti o otázke tzv. kognitívnej inkrementálnosti: „Aký je vzťah medzi stratégiami používanými pri základných problémoch vnímania a konania, a tými, ktoré používame na riešenie abstraktnejších problémov na vyššej úrovni?“ Ohľadne tejto otázky však stále neexistuje v kognitívnej vede konsenzus, pretože myšlienka hybridnej kognitívnej architektúry je stále aktuálna, a je realizovaná aj v návrhoch hybridných systémov (napr. [72]).

Principiálnou motiváciou pre orientáciu na neurónové siete pri kognitívnom modelovaní je aj predpoklad, že svet je spojitý (t.j. fyzikálne charakteristiky vnímané senzormi sú spojité), akcie agenta sú spojité (uhly natočenia), a preto aj myseľ agenta/človeka by prirodzene mohla byť spojitá ([133]). Kľúčová výhoda spojitých priestorov spočíva v možnosti zovšeobecňovania, čo sa javí ako kognitívne relevantný faktor a požiadavka na učiace sa systémy.

Pochopiteľne, svet naokolo má aj svoje diskkrétne prvky, napríklad počty objektov bývajú diskkrétne, počet akcií vykonateľných agentom v danom kontexte býva diskrétny vzhľadom na ich efekt (napr. odbočiť vľavo alebo vpravo), a koniec koncov logické úlohy rozmanitého typu ktoré človek dokáže riešiť, bývajú formulované pomocou symbolov. Celé toto spektrum je nutné vedieť vysvetliť konekcionisticky, ak táto paradigma má ambíciu tvoriť komponenty ideálnej kognitívnej architektúry.

V diskrétnom svete symbolových systémov sú aplikovateľné kvalitatívne odlišné mechanizmy, no tie sú tiež užitočné, napr. pri návrhu expertných systémov alebo systémov s ontológiami, ktoré slúžia človeku. Voľba prístupu a úrovne abstrakcie teda závisí od cieľa a kritérií. Umelá inteligencia a výpočtová kognitívna veda majú veľa spoločného, no je medzi nimi jeden dôležitý rozdiel. Umelá inteligencia sa zaujíma o efektívne riešenia využiteľné v inteligentných systémoch, bez ohľadu na ich biologickú relevantnosť. V kognitívnej vede na biologickej prijateľnosti záleží.

PodĎakovanie: Tento príspevok vznikol z podpory grantovej agentúry VEGA v rámci grantových úloh 1/0439/11 a 1/0602/10, KEGA K-09-016-00.

Literatúra

- [1] Abrahamsen A., Bechtel W.: Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective. In R. Stainton (Ed.), *Contemporary Debates in Cognitive Science*. Oxford: Basil Blackwell, 2006.
- [2] Aizawa K.: *The Systematicity Arguments*. Kluwer Academic, Dordrecht, 2003.
- [3] Arbib M.: From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105-168, 2005.
- [4] Asada M. a spol.: Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development* 1(1), 12-34, 2009.
- [5] Bell A.: Levels and loops: The future of artificial intelligence and neuroscience. *Philosophical Transactions of the Royal Society London B*, 354, 2013-2020, 1999.
- [6] Barsalou L.: Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577-660.
- [7] Barsalou L.: Grounded cognition. *Annual Reviews of Psychology*, 59, 617-645, 2008.
- [8] Bechtel W., Abrahamsen A.: *Connectionism and the Mind: Parallel Processing, Dynamics and Evolution in Networks*. Malden, MA: Blackwell, 2002.
- [9] Bengio Y., LeCun Y.: Scaling learning algorithms towards AI. V knihe Bottou L. et al. (eds.), *Large-Scale Kernel Machines*, MIT Press, 2007.
- [10] Boroditsky L., Ramscar M.: The roles of body and mind in abstract thought. *Psychological Science*, 13, 185-88, 2002.
- [11] Brakel P., Frank S.: Strong systematicity in sentence processing by simple recurrent networks. In: *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 1599-1604, 2009.
- [12] Bressler S., Kelso J.S.: Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, 5, 26-36, 2001.

-
- [13] Brooks R.: A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1), 14-23, 1986.
- [14] Burgess N.: Spatial memory: how egocentric and allocentric combine. *Trends in Cognitive Sciences*, 10(12), 551-557, 2006.
- [15] Byrne P., Becker S., Burgess N.: Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological Review*, 114, 340-375, 2007.
- [16] Cangelosi A. a spol.: Integration of action and language knowledge: a roadmap for developmental robotics, *IEEE Transactions on Autonomous Mental Development*, 2(3), 167-195, 2010.
- [17] Cangelosi A., Tikhanoff V., Fontanari J., Hourdakis E.: Integrating language and cognition: A cognitive robotics approach. *IEEE Computational Intelligence Magazine*, 2(3), 65-7, 2007.
- [18] Chang S., Papadimitriou C., Snyder L.: Using a compound gain field to compute a reach plan. *Neuron*, 64, 744-755, 2009.
- [19] Christiansen M., Chater N.: Connectionist natural language processing: The state of the art. *Cognitive Science*, 23(4), 417-437, 1999.
- [20] Christiansen M., Chater N. (eds.): *Connectionist psycholinguistics*. Ablex Publishing, 2001.
- [21] Churchland P., Sejnowski, T.: *The Computational Brain*. Cambridge, MA: MIT Press, 1992.
- [22] Ciresan D., Meier U., Gambardella L., Schmidhuber J.: Deep big simple neural nets for handwritten digit recognition. *Neural Computation*, 22(12), 3207-3220, 2010.
- [23] Clark A: *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford University Press, 2001.
- [24] Cleeremans A., Destrebecqz A., Boyer M.: Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2, 406-416, 1998.
- [25] Cohen Y., Andersen R.: A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience*, 3, 553-562, 2002.
- [26] Craenen B., Eiben A.: Computational intelligence. In: *Encyclopedia of Life Support Sciences*, EOLSS Publishers Co. Ltd., 2003.
- [27] David N., Newen A., Vogeley K.: The „sense of agency“ and its underlying cognitive and neural mechanisms. *Consciousness and Cognition*, 17, 523-534, 2008.
- [28] Davies J.: *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. John Wiley & Sons, 2006.
- [29] Desmurget M., Grafton S.: Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4(11): 423-431, 2000.
- [30] Doya K.: What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex? *Neural Networks*, 12, 961-974, 1999.

-
- [31] Doya K.: Reinforcement learning in continuous time and space. *Neural Computation*, 12(1), 219-245, 2000.
- [32] Duch W.: Towards comprehensive foundations of computational intelligence. *Lecture Notes in Computer Science*, Springer, Vol. 63, 261-316, 2007.
- [33] Edelman G.: Bright Air, Brilliant Fire: On the Matter of the Mind. New York: Basic Books, 1992.
- [34] Elman J.: Finding structure in time. *Cognitive Science*, 14(2), 179-211, 1990.
- [35] Farkaš I.: Samoorganizujúce sa mapy. Kapitola v knihe V. Kvasnička a kol., *Úvod do teórie neurónových sietí*, IRIS, Bratislava, 142-189, 1997.
- [36] Farkaš I.: Konekcionistické modelovanie jazyka. V knihe: Rybár J., Kvasnička V., Farkaš I. (zost.), *Jazyk a kognícia*, Kalligram, Bratislava, 262-305, 2005.
- [37] Farkaš I.: Samoorganizácia ako hybná sila dynamických vzorcov aktivít v mozgu a mysli. V zborníku *Kognice a umělý život VI*, Opava: Slezská univerzita, 143-148, 2006.
- [38] Farkaš I.: Hľadanie kauzálnych vzťahov v probléme mysle a tela z pohľadu neredukcionistického fyzikalizmu. V knihe: Kvasnička V. a spol. (zost.), *Mysel, inteligencia a život*. Vydavateľstvo STU v Bratislave, 3-16, 2007.
- [39] Farkaš I.: Konceptuálne východiská pre model stelesnenej mysle. V knihe: Kvasnička V., Kelemen J., Pospíchal J. (zost.), *Modely mysle*, Vydavateľstvo Európa, Bratislava, 35-64, 2008.
- [40] Farkaš I.: Learning nonadjacent dependencies with a recurrent neural network. In: Koeppen M. et al. (zost.), *Int. Conference on Neural Information Processing Systems (ICONIP'2008)*, Springer, Lecture Notes in Computer Science, 292-299, 2009.
- [41] Farkaš I.: Neurálne modely v kognitívnej robotike: vizuálno-motorická interakcia. V zborníku Kelemen J., Kvasnička V.: *Kognice a umělý život X*, Ostravice, ČR, 93-99, 2010.
- [42] Farkaš I., Crocker M.: Syntactic systematicity in sentence processing with a recurrent self-organizing network. *Neurocomputing*, 71, 1172-1179, 2008.
- [43] Farkaš I., Malý M., Rebrová K.: Mirror neurons: theoretical and computational issues. *Technická správa TR-2011-28*. FMFI Univerzita Komenského, 2011.
- [44] Farkaš I., Rebrová K.: Meaning and its representation from the perspective of grounded cognition. *Kognice*, Praha, 2010.
- [45] Feldman J., Ballard D.: Connectionist models and their properties. *Cognitive Science*, 6, 205-254, 1982.
- [46] Felleman D., van Essen D.: Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1-47, 1991.
- [47] Fletcher C., Chrysler S.: Surface forms, text bases, and situation models: recognition memory for three types of textual information. *Discourse Processes*, 13, 175-190, 1990.
- [48] Fodor J.: *The Modularity of Mind*. Cambridge, MA: MIT Press, 1983.

-
- [49] Fodor J.: *The Mind Doesn't Work That Way*. MIT Press, 2000.
- [50] Fodor J., Pylyshyn Z.: Connectionism and cognitive architecture. *Cognition*, 28, 3-71, 1988.
- [51] Frank S.: Learn more by training less: Systematicity in sentence processing by recurrent networks. *Connection Science*, 18, 287-302, 2006a.
- [52] Frank S.: Strong systematicity in sentence processing by an echo-state network. In: *Proceedings of ICANN*, Part I, Lecture Notes in Computer Science, vol. 4131, Springer, 2006b.
- [53] Frank, S., Haselager W., Van Rooij I.: Connectionist semantic systematicity. *Cognition*, 110, 358-379, 2009.
- [54] Freeman W.: *How Brains Make Up Their Minds*. New York: Columbia University Press, 2001.
- [55] Frischen A., Loach D., Tipper S.: Seeing the world through another person's eyes: simulating selective attention via action observation. *Cognition*, 111(2), 212-8, 2009.
- [56] Gallese V., Keysers C. and Rizzolatti G.: A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396-403, 2004.
- [57] Gardner H.: *The Minds New Science: A History of the Cognitive Revolution*. New York: Basic Books, 1987.
- [58] Gärdenfors P.: Cued and detached representations in animal cognition. *Behavioural Processes*, 36, 263-273, 1996.
- [59] Gentner D.: Psychology in cognitive science: 1978-2038. *Topics in Cognitive Science*, 2, 328-344, 2010.
- [60] Gibbs R.: *Embodiment and Cognitive Science*. New York: Cambridge University Press, 2006.
- [61] Gibson J.: *Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin, 1970.
- [62] Greenfield J.: Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14, 531-595, 1991.
- [63] Griffiths T., Kemp C., Tenenbaum J.: Bayesian models of cognition. In R. Sun (Ed.), *Cambridge Handbook of Computational Psychology*, New York: Cambridge University Press, 59-100, 2008.
- [64] Hadley R.: Systematicity in connectionist language learning. *Mind and Language*, 9(3), 247-272, 1994.
- [65] Harnad S.: The symbol grounding problem. *Physica D*, 42, 335-346, 1990.
- [66] Havel, I.M.: *Robotika – úvod do teorie kognitivních robotů*. SNTL, Praha, 1980.
- [67] Haykin S.: *Neural Networks and Learning Machines*, Prentice Hall, 2008.
- [68] Hinton G., Osindero S., Teh Y.: A fast learning algorithm for deep belief nets. *Neural Computation*, 18, 1527-1554, 2006.

-
- [69] Hornik K., Stinchcombe M., White H.: Multilayer feedforward networks are universal approximations. *Neural Networks*, 2, 359-366, 1989.
- [70] Hutchins E.: *Cognition in the Wild*. MIT Press, 1996.
- [71] Izhikevich E.: Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15, 1063-1070, 2004.
- [72] Jilk D., Lebiere C., O'Reilly R., Anderson J.: SAL: An explicitly pluralistic cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 20, 197-218, 2008.
- [73] Kaczmarek L., Levitan I. (zost.): *Neuromodulation: The Biochemical Control of Neuronal Excitability*. NY: Oxford University Press, 1987.
- [74] Kawato M.: Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9(6): 718-727, 1999.
- [75] Kelemen J., Liday M.: *Expertné systémy pre prax*. Bratislava, Sofa, 1996.
- [76] Kelso J.: *Dynamic Patterns: The Self-Organization of Brain and Behavior*. The MIT Press, Cambridge, MA, 1995.
- [77] Kohonen T.: *Self-Organizing Maps*. Springer, (3rd, extended ed.), 2001.
- [78] Korenčiak R.: Aproximácia motorického priestoru ramena simulovaného robota. *Diplomová práca*, FMFI Univerzity Komenského, 2010.
- [79] Kosslyn S.: *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press, 1994.
- [80] Lakoff G., Johnson M.: *Philosophy in the Flesh: The Embodied Mind and It's Challenge to Western Thought*. New York: Basic Books, 1999.
- [81] Landauer T., Foltz P., Laham D.: Introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284, 1998.
- [82] Lave J.: *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. New York: Cambridge University Press, 1988.
- [83] LeCun Y. a spol.: Backpropagation applied to handwritten zip code recognition, *Neural Computation*, 1(4), 541-551, 1989.
- [84] Li P., Farkaš I., MacWhinney B.: Early lexical development in a self-organizing neural network. *Neural Networks*, 17(8-9), 1345-1362, 2004.
- [85] Lukosevicius M., Jaeger H.: Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3), 127-149, 2009.
- [86] Maass W., Bishop C. (eds.): *Pulsed Neural Networks*. MIT Press, 1999.
- [87] McClelland J.: The place of modeling in cognitive science. *Topics in Cognitive Science*, 1(1), 11-38, 2009.
- [88] Majid A., Bowerman M., Kita S., Haun D., Levinson, S.: Can language restructure cognition? The case for space. *Trends in Cognitive Science*, 8, 108-114, 2004.

-
- [89] Marocco D., Cangelosi A., Fischer K., Belpaeme T.: Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Frontiers in Neurorobotics*, 4(1), 1-15, 2010.
- [90] Marr D.: *Vision*. San Francisco, CA: W.H. Freeman, 1982.
- [91] Marsland S.: *Machine Learning: An Algorithmic Introduction*. CRC Press, New Jersey, USA, 2009.
- [92] Mayberry M.: Incremental nonmonotonic parsing through semantic self-organization. *Doctoral thesis*. Department of Computer Sciences, University of Texas at Austin, 2004.
- [93] McClelland J.: The place of modeling in cognitive science. *Topics in Cognitive Science*, 1(1), 11-38, 2009.
- [94] McClelland J., Kawamoto A.: Mechanisms of sentence processing: Assigning roles to constituents of sentences. In *Parallel Distributed Processing*, Vol. 2. MIT Press, Cambridge, MA, 318-362, 1986.
- [95] McCulloch W., Pitts W.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7, 115-133, 1943.
- [96] McLeod P., Plunkett K., Rolls, E.: *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford, UK: Oxford University Press, 1998.
- [97] Miall R.: Connecting mirror neurons and forward models. *Neuroreport*, 14(17), 2135-7, 2003.
- [98] Miikkulainen R.: Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language*, 59(2), 334-366, 1997.
- [99] Mukamel R., Ekstrom A., Kaplan J., Iacoboni M., Fried I.: Single-neuron responses in humans during execution and observation of actions. *Current Biology*, 20(8), 750-756, 2010.
- [100] Newell A., Simon H.: Computer science as empirical enquiry. *Communications of the ACM*, 19 (1976) 113-126. In: *Mind Design II*, J. Haugeland (ed.), 81-110. Cambridge, MA: MIT Press, 1997.
- [101] Newport E., Aslin R.: Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162, 2004.
- [102] O'Brien G., Opie J.: How do connectionist networks compute? *Cognitive Processing*, 7, 30-41, 2006.
- [103] O'Reilly R., Munakata Y.: *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press, 2000.
- [104] Pfeifer R., Lungarella M., Iida F.: Self-organization, embodiment, and biologically inspired robotics. *Science*, 318(5853), 1088-93, 2007.
- [105] Pfeifer R., Scheier C.: *Understanding Intelligence*. MIT Press, Cambridge, MA, 1999.

- [106] Piccinini G.: Some neural networks compute, others don't. *Neural Networks*, 21(2-3), 311-321, 2008.
- [107] Pinker S.: *How the Mind Works*. W.W. Norton and co., 2009.
- [108] Pinker S., Mehler J. (zost.): *Connections and Symbols*. Cambridge, MA: MIT Press, 1988.
- [109] Pinker S., Prince A.: On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193, 1988.
- [110] Pinker S., Ullman M.: The past tense debate. *Trends in Cognitive Sciences*, 6(11), 456-474, 2002.
- [111] Pitt D.: Mental representation, *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/fall2008/entries/mental-representation>, 2008.
- [112] Plunkett K., Juola P.: A connectionist model of English past tense and plural morphology. *Cognitive Science*, 23, 463-490, 1999.
- [113] Port R., van Gelder T. (Eds.): *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press, 1995.
- [114] Pouget A., Snyder L.: Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3, (suppl.), 1192-1198, 2000.
- [115] Pstružina K.: Abstrakce, komparace a generalizace jako základní operace lidského myšlení. V kniže: *Mysel, inteligencia a život*. Vydavateľstvo STU v Bratislave, 111-117, 2007.
- [116] Pulvermüller F.: Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576-582, 2005.
- [117] Pylyshyn Z.: *Computation and Cognition*. MIT Press, 1984.
- [118] Quartz S., Sejnowski T.: The neural basis of cognitive development. *Behavioral and Brain Sciences*, 20, 537-596, 1997.
- [119] Quinlan P.: *Connectionism and Psychology: A Psychological Perspective on New Connectionist Research*. Chicago, IL: University of Chicago Press, 1991.
- [120] Ritter H., Kohonen T.: Self-organizing semantic maps. *Biological Cybernetics*, 61, 241-254, 1989.
- [121] Rizzolatti G., Fadiga L., Fogassi L., Gallese V.: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131-41, 1996.
- [122] Rohde D.: A connectionist model of sentence comprehension and production. *Doctoral thesis*. Carnegie Mellon University, Pittsburg, 2002.
- [123] Rohde D., Plaut D.: Connectionist models of language processing. *Cognitive Studies*, 10, 10-28, 2003.
- [124] Rosch E.: Principles of categorization. In: Rosch E., Lloyd B. (zost.), *Cognition and Categorization*. Lawrence Erlbaum Associates, Hillsdale, NJ, 27-48, 1978.

-
- [125] Rosenblatt, F.: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington, DC: Spartan, 1962.
- [126] Roy D.: Grounding words in perception and action: computational insights. *Trends in Cognitive Science*, 9(8), 389-96, 2005.
- [127] Rumelhart D., McClelland J. and the PDP research group: *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, Vol. 1. Cambridge, MA: MIT Press, 1986.
- [128] Saffran J., Newport E., Aslin R.: Word segmentation: the role of distributional cues. *Journal of Memory and Language*, 35, 606-621, 1996.
- [129] Salinas E., Sejnowski, T.: Gain modulation in the central nervous system: Where behavior, neurophysiology and computation meet. *The Neuroscientist*, 7(5), 430-440, 2001.
- [130] Searle J.: Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3, 417-458, 1980.
- [131] Siegelmann H., Sontag E.: Turing computability with neural nets. *Applied Mathematics Letters*, 4, 77-80, 1991.
- [132] Smith L.: Cognition as a dynamic system: principles from embodiment. *Developmental Review*, 25, 278-98, 2005.
- [133] Spivey M.: *The Continuity of Mind*. Oxford: Oxford University Press, 2007.
- [134] Steels L.: The symbol grounding problem has been solved, so what's next? In: de Vega M., Glenberg A. & Graesser A. (eds.), *Symbols and Embodiment: Debates on Meaning and Cognition*, Oxford University Press, 223-244, 2008.
- [135] Sugita Y., Tani J.: Learning semantic combinatoriality from the interaction between linguistic and behavioral processes, *Adaptive Behavior*, 13(1), 33-52, 2005.
- [136] Šefránek J.: *Inteligencia ako výpočet*. IRIS Bratislava 2002.
- [137] Šefránek J., Takáč M., Farkaš I.: Vznik inteligencie v umelých systémoch. V knihe Magdolen, D. (zost.): *Hmota, život, inteligencia: Vznik*. VEDA, Bratislava, 245-270, 2008.
- [138] Šíma J., Orponen P.: General-purpose computation with neural networks: A survey of complexity theoretic results. *Neural Computation*, 15, 2727-2778, 2003.
- [139] Tadeo M., Floridi L.: The symbol grounding problem: a critical review of fifteen years of research. *Journal of Experimental and Theoretical Artificial Intelligence*, 419-445, 2005.
- [140] Takáč M.: Construction of Meanings in Living and Artificial Agents. *Dizertačná práca*, FMFI UK v Bratislave, 2007.
- [141] Takáč M.: Autonomous construction of ecologically and socially relevant semantics. *Cognitive Systems Research*, 9, 293-311, 2008.
- [142] Thelen E., Smith L.: A Dynamic Systems Approach to the Development of Cognition and Action. Cambridge, MA: MIT Press, 1994.

-
- [143] Turing A.: Computing machinery and intelligence. *Mind*, 59, 433-460, 1950.
- [144] Ungerleider L., Mishkin M.: Two cortical visual systems. V kniže D. Ingle, M. Goodale, & R. Mansfield (Eds.), *Analysis of visual behavior*, Cambridge, MA: MIT Press, 549-586, 1982.
- [145] Van Gelder T., Port R.: It's about time: An overview of dynamical approaches to cognition. In: *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press, 1-44, 1995.
- [146] Van Hasselt H., Wiering M.: Reinforcement learning in continuous action spaces. *IEEE Int. Symposium on Approximate Dynamic Programming and Reinforcement Learning*, 272-279, 2007.
- [147] Varela F., Thompson E., Rosch, E.: *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 1991.
- [148] Vavrečka M.: Ukotvení symbolů v kontextu nulového sémantického závazku. V zborníku *Kognícia a umělý život VI*, Opava: Slezská univerzita, 401-411, 2006.
- [149] Vavrečka M.: Kognitivní sémantika a její aplikace v modelu reprezentace prostorových vztahů. *Dizertačná práce*, Masarykova Univerzita v Brně, 2008.
- [150] Wermter S., Elshaw M.: Learning robot actions based on self-organising language memory. *Neural Networks*, 16, 691-699, 2003.
- [151] Wilson M.: Six views of embodied cognition. *Psychonomics Bulletin Review*, 9(4), 625-636, 2002.
- [152] Ziemke T.: Rethinking grounding. In: Riegler A. et al. (eds.), *Understanding Representation in the Cognitive Sciences*, New York: Plenum Press, 177-190, 1999.
- [153] Zipser D., Andersen R.: A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331, 679-684, 1988