Vznik inteligencie v umelých systémoch[1]

Ján Šefránek, Martin Takáč, Igor Farkaš

> „Namiesto sporov o tom, či je nejaké správanie inteligentné alebo nie (čo je vždy diskutabilné), pokúsime sa odpovedať na nasledujúcu otázku: Majme nejaké správanie, ktoré je pre nás niečím zaujímavé, či už je to správanie ľudské, slonie, mravčie, alebo správanie robota – ako toto správanie vzniká? Ak dokážeme dať dobré odpovede na túto otázku pre dosť širokú škálu správaní, môžeme povedať, že sme pochopili niečo o princípoch inteligencie."
>
> (Pfeifer, Scheier, 1999)

S pojmom inteligencie sa stretávame takmer každodenne, najčastejšie v kontexte ľudskej inteligencie, často aj v súvislosti s inými živými tvormi, a v odbornej či sci-fi literatúre čím ďalej tým častejšie v kontexte vznikajúcich umelých systémov, ktoré považujeme za inteligentné. Je koncept inteligencie a inteligentného správania vôbec jednoznačný? Pokúsime sa definovať tento koncept, a to najprv v doméne prirodzenej inteligencie a potom v doméne informatiky (umelých systémov), pretože obe domény poznania sa trochu líšia a navyše v oboch existuje viac názorov na povahu inteligencie a jej prejavov. Zatiaľ čo inteligenciou človeka (prípadne iných živých tvorov) sa psychológovia zaoberajú už od začiatku minulého storočia (a v ostatných dekádach rokov aj odborníci z iných vedných odborov), inteligencia v umelých systémoch je novším fenoménom, súvisiacim s vynájdením počítačov, ktorý vnáša svetlo aj do hľadania formálneho jednotiaceho pohľadu na inteligentné správanie systému (či už biologického alebo umelého). Prv než sa dostaneme k tomuto formálnejšiemu prístupu, povedzme si niečo o prirodzenej inteligencii z pohľadu psychológov.

Prirodzená inteligencia

Spektrum názorov na koncept prirodzenej inteligencie prešiel za posledné storočie svojím vývinom. Keď Journal of Educational Psychology v roku 1921 požiadal štrnásť špičkových expertov, aby definovali inteligenciu, dostal štrnásť rôznych

[1] Rukopis kapitoly do pripravovanej knihy Magdolen, D. (ed.): *Hmota, život, inteligencia: vznik*. VEDA.

odpovedí, z ktorých mnohé boli navzájom takmer disjunktné (Pfeifer, Scheier, 1999). Koncom minulého storočia, ako výsledok snahy o nájdenie nejakej spoločnej definície inteligencie vznikol v roku 1994 dokument „Mainstream Science on Intelligence". S jeho obsahom sa stotožňuje 52 signatárov-odborníkov, ktorí sa zhodli na tom, že (prirodzená) inteligencia je „...veľmi všeobecná mentálna vlastnosť, ktorá (okrem iného) zahŕňa schopnosť usudzovať, plánovať, riešiť problémy, myslieť abstraktne, rozumieť komplexným myšlienkam, ako aj schopnosť učiť sa rýchlo, a na základe skúseností... Táto vlastnosť odráža schopnosť porozumieť okoliu, schopnosť 'vhľadu' do vecí, 'zistenia' toho, čo treba (u)robiť v danej situácii..." (Gottfredson, 1997).

Títo odborníci sú presvedčení, že inteligencia človeka sa dá vedecky merať, a vyjadriť číselne (napr. pomocou skóre IQ, alebo faktora *g*), sú však aj experti, ktorí argumentujú o opaku. Takisto existujú teórie aj o viacerých typoch inteligencie, a nielen jednej, všeobecnej. Napríklad, Gardner (1983) navrhol 7 rôznych aspektov inteligencie (jazyková, matematicko-logická, priestorová atď.), zatiaľ čo triarchická teória inteligencie postuluje tri zložky – kreatívnu, analytickú a praktickú, ktoré sa podieľajú na inteligentnom správaní (Sternberg, 1985). Existuje dokonca aj teória, podľa ktorej v inteligencii človeka hrajú kľúčovú rolu jeho emocionálne vlastnosti (Goleman, 1995). Otázka prirodzenej inteligencie je teda stále otvorenou oblasťou výskumu (Gottfredson, 2005).

Z nášho pohľadu je spomínaná definícia inteligencie vágna, pretože sa opiera o niektoré pojmy, ktoré sú z formálneho či filozofického hľadiska samy o sebe nejasné. To, či nejaké správanie považujeme za inteligentné, bude závisieť aj od toho, akého živého tvora máme na mysli. Inými slovami, interpretácia nejakého správania závisí od našich očakávaní. Kto by napríklad uvidel mačku hrať dámu, určite by povedal, že je mimoriadne inteligentná. V prípade dospelého človeka by to tak zrejme nebolo, iba ak by šlo povedzme o dvojročné dieťa.

Čo vieme povedať o vzniku prirodzenej inteligencie človeka? Je zrejmé, že inteligencia človeka závisí od jeho mozgu, ktorý sa spolu so senzorickými orgánmi vyvinul v priebehu evolúcie tak, aby človek dokázal efektívne riešiť dôležité úlohy spoľahlivo a rýchlo. O vzniku prirodzenej inteligencie môžeme teda hovoriť vo *fylogenetickom* kontexte, v ktorom sa človek dostal na piedestál tvorstva za Zemi, pretože disponuje najvyšším stupňom inteligencie. Čím sa vlastne človek odlišuje od ostatných inteligentných zvierat na Zemi? Komunikujeme jazykom, vieme uvažovať, plánovať, ale aj vykonávať (generovať) zložité motorické pohyby (ručné práce, tanec, hudba, šport). Všetky tieto činnosti v sebe skrývajú schopnosť mozgu vykonávať zložité sekvenčné procesy, čo vystihuje hypotéza o „sekvenčnej mozgovej kôre"

(Calvin & Ojemann, 1994, kap. 16). Možno to bola z neuroanatomického hľadiska najmä expanzia frontálnej kôry, predpokladaného centra plánovania a exekutívy, ktorá umožnila vznik inteligentného správania (Calvin, 1996). Do akej miery je z pohľadu ontogenetického vývinu človeka táto schopnosť vrodená, a akú úlohu tu hrá skúsenosť, je však tiež zložitou otázkou (Neill, 2004). Niektorí vedci sa snažia hľadať rôzne neurálne koreláty inteligencie ako veľkosť mozgu (Jerison, 2000) alebo jeho elektrická aktivita. Zaujímavým je poznatok z oblasti moderných zobrazovacích techník mozgu, ktoré poukazujú na to, že získanie „vhľadu" do riešeného problému je v mozgu sprevádzané široko distribuovanou sieťou novoprepojených kortikálnych aktivít, ktoré jedincovi umožnia „vidieť" nové súvislosti a analógie (Jung-Beeman, 2004). Vplyvnou hypotézou je pohľad na mozog ako na komplexný distribuovaný biologický orgán so samoorganizáciou (t.j. bez centrálneho riadenia) na spracovanie informácie, ktorý sa svojou architektúrou a mechanizmami vo viacerých aspektoch kvalitatívne odlišuje od bežného počítača (Bell, 1999). Už pred takmer polstoročím bolo poukázané na to, ako procesy samoorganizácie v mozgu sa môžu podieľať na jeho činnosti (Ashby, 1962). V súčasnosti sa samoorganizácia považuje za kľúčovú súčasť spracovania informácie a adaptácie v mozgu (Kohonen, 1984; von der Malsburg, 2003). Samoorganizácia je prejavom plasticity mozgu, t.j. jeho schopnosti adaptovať sa, pričom táto schopnosť pravdepodobne koreluje s prejavom inteligencie u človeka – s jeho schopnosťou rýchlo sa učiť.

My sa v ďalšom texte budeme snažiť abstrahovať od týchto biologických korelátov, ktoré sú špecifické pre inteligenciu živých systémov (aj keď sa im v kontexte biologicky motivovaných formálnych prístupov nevyhneme). Je otvorenou otázkou, do akej miery inteligentné správanie závisí od implementácie. Príroda nám ponúka nejaké evolučné riešenia (ktorých fungovanie však zatiaľ poznáme len rámcovo), no je možné, že existuje viacero odlišných formálnych spôsobov ako opísať inteligentné správanie a jeho vznik. Najprv sa však musíme pokúsiť objasniť, čo je inteligentné správanie.

Inteligentné správanie

Takmer súčasne so vznikom počítačov sa začali objavovať otázky ako „Môžu stroje *myslieť*?", „Môže byť stroj *inteligentný*?", „Môže existovať *kognícia* na nebiologickej báze?" a či „Môže mať stroj *vedomie*?" Hoci sa pojmy myslenie, inteligencia, kognícia a vedomie navzájom líšia vo významových odtieňoch, kladené otázky majú spoločných niekoľko vecí: V prvom rade sa pýtajú, či možno neživým artefaktom pripísať isté atribúty, ktoré sa typicky používajú v súvislosti s ľuďmi. Ďalej to, že

atribúty nie sú dobre definované: ako bolo uvedené vyššie, je mnoho navzájom odlišných definícií týchto pojmov a medzi expertmi neexistuje konsenzus. A napokon, otázky majú pre nás emocionálny náboj, istým spôsobom sa dotýkajú našej existencie, čo je príčinou toho, že diskusie o nich často prekračujú vedeckú argumentáciu a ústia do zápalistých sporov.

Alan Turing v úvode svojho článku z roku 1950 správne poznamenáva, že odpoveď na otázku, či môžu stroje myslieť, závisí od toho, čo chápeme pod pojmami *stroj* a *myslieť*. Namiesto hľadania konvenčných vymedzení týchto pojmov navrhol nahradiť otázku operacionalizovaným testom, tzv. imitačnou hrou. Myšlienka hry, ktorá vošla do histórie ako Turingov test, spočíva v tom, že nejaký systém je inteligentný vtedy, ak sa nám ľuďom (ktorí sa bez dôkazu pokladáme za inteligentných ☺) nepodarí odhaliť, že taký nie je, resp. ak systém dokáže vykonávať činnosť, ktorá podľa nás vyžaduje inteligenciu, nerozoznateľne od človeka. Tou činnosťou v rôznych verziách Turingovho testu bolo viesť dialóg s človekom cez terminál ďalekopisu.

Odporcovia Turingovho testu často namietali v štýle tautológie „Myslieť môže iba človek, pretože myslieť môže iba človek“ (resp. jej sofistikovanejšími verziami s vmedzerenými premennými, napr. k mysleniu je potrebné vedomie a vedomie má iba človek, teda myslieť môže iba človek). Tautológie nemožno vyvrátiť, tak sa o to ani nebudeme pokúšať. Zaujímavejšie argumenty voči Turingovmu testu a strojovej inteligencii s dôsledkami pre štúdium kognície ako takej rozoberieme v ďalšom texte. Osvojíme si prístup S. Paperta: „Je zrejmé, že stroje nemyslia presne rovnakým spôsobom ako ľudia... Otázkou je, či by sme radi rozšírili pojem myslenie tak, aby pokrýval i to, čo môžu robiť stroje. Toto je jediný zmysluplný význam otázky, či stroje myslia.“ (podľa Kelemen, 1994).

Fenomén, ktorého štúdium nás zaujíma, budeme v ďalšom texte najčastejšie označovať výrazom *inteligentné správanie*. Vybrali sme si najneutrálnejší výraz, keďže pojmy vedomie a myslenie majú príliš veľa vžitých konotácií. Otázke subjektívneho vedomia u strojov sa teda nebudeme venovať. Aj u zvoleného výrazu neprijmeme nejaké uzuálne definície, ale jeho sémantiku vybudujeme „zdola“ v ďalšom texte. Otázka teda nestojí „Vykazujú stroje inteligentné správanie?“ – budeme skôr skúmať, čo možno za inteligentné správanie považovať, a kde, za akých podmienok a akým spôsobom vzniká. Pravdepodobne zistíme, že tento pojem má neostré hranice a namiesto binárnych odpovedí bude skôr namieste hovoriť o *stupni*

resp. *miere* inteligentného správania u skúmaných systémov. Pokúsime sa čo najviac vyhýbať antropomorfným definíciám a kritériám, ktoré by tautologicky spĺňal iba človek. Mieru inteligentného správania budeme potom môcť skúmať u ľudí a umelých artefaktov, ako aj u iných biologických systémov, živých organizmov či ich spoločenstiev na rôznom stupni fylogenetického vývoja a ontogenetického vývinu.

Idea hľadať inteligentné správanie vo fylogenetickom strome hlboko pred človekom nie je nová, napr. (de Chardin, 1956, Goodwin, 1978, Kováč, 1986). Zakladateľ kognitívnej biológie Ladislav Kováč postuluje elementárne formy kognície (v našom jazyku: systémov schopných zaujímavého resp. inteligentného správania) už na bakteriálnej, bunkovej, ba aj molekulárnej úrovni (Kováč, 2006). Zo systémov, ktoré sa objavili v priebehu evolúcie, niektoré pretrvali, čo znamená, že ich štruktúra lepšie odrážala vlastnosti prostredia. Biologická evolúcia spočíva v generovaní hypotéz o povahe prostredia (v podobe štruktúr, ktoré v ňom môžu/majú prežiť), vyvracaní hypotéz a udržovaní tých hypotéz, ktoré neboli vyvrátené (Kováč, 1986). Evolučnú adaptáciu možno teda vnímať ako akýsi druh fylogenetického učenia, pričom poznatky sa zabudovávajú do štruktúry systému (vrátane vrodených inštinktov u zvierat, či predprogramovaného softvéru u umelých systémov). Šanca systémov pretrvať sa zvyšuje, keď aktívne spoznávajú prostredie a adaptujú sa naň alebo ho svojím správaním adjustujú. Systém spoznáva tie aspekty prostredia, ktoré sú pre neho významné. Rozlišovanie istých aspektov či stavov prostredia (prostredie zahŕňa aj samotný systém) môžeme nazvať prvotnou tvorbou významov – *signifakciou* (Kováč, 2003). Najjednoduchší kognitívny systém pozostáva z dvoch mechanizmov – senzie (prijatie a rozlíšenie signálu z prostredia) a akcie. Procesom interkalárnej evolúcie sa medzi existujúce poznávacie mechanizmy postupne vmedzerujú ďalšie stupne: percepcia, afekcia, kogitácia (Kováč, 2003). Na najvyšších stupňoch evolúcie sa objavuje myslenie ako abstraktná akcia – testovanie rôznych motorických aktov bez zapojenia svalov. Schopnosť systému mentálne simulovať rôzne scenáre a vyhodnotiť ich dôsledky (tzv. „what-if“ myslenie) bez nutnosti odskúšať si všetky varianty „naostro“ zvyšuje jeho šance na prežitie.

Searlova čínska izba

Vráťme sa späť k Turingovmu testu. Starostlivá analýza jeho výpovednej sily má kardinálny význam z hľadiska našej centrálnej témy, ktorou je vznik inteligencie

v umelých systémoch (špeciálnejšie, počítačových programoch). Treba sa pýtať, či má vôbec zmysel hovoriť o inteligentnom správaní počítačových programov; ak áno, ako charakterizovať takéto správanie a vďaka, akým črtám, kvalitám sa môže objaviť (vzniknúť).

Searlova metafora čínskej izby (Searle 1980, 1984, 1990, 1999) a jej zamýšľaná interpretácia odpovedá na túto fundamentálnu otázku negatívne: počítačovým programom nemožno pripísať inteligenciu, myslenie, schopnosť pochopiť. Cieľom Searlovho argumentu, myšlienkového experimentu bolo ukázať, že „umelá inteligencia“ (v silnom zmysle slova) nie je možná. Searlov argument bol formulovaný takto (vychádzame hlavne z formulácie v (Searle, 1999)):

Predpokladajme, že v nejakej miestnosti je zavretá osoba, ktorej materinským jazykom je angličtina; nerozumie však čínsky, nepozná čínske písmo. V miestnosti sú krabice s čínskymi znakmi (*databáza*) a kniha s inštrukciami – pravidlami, určujúcimi ako manipulovať s čínskymi znakmi (*program*). Tieto pravidlá sú formálne v tom zmysle, že berú do úvahy iba formu, tvar výrazov, poskladaných z čínskych znakov. Predstavme si ďalej, že ľudia, ktorí sú mimo miestnosti, posielajú do miestnosti postupnosti čínskych znakov (*vstup*), tie sú otázkami v čínštine. Osoba, ktorá je v miestnosti, nevie o tom, že to sú otázky v čínštine.

Predpokladajme teraz, že na základe spomínaných inštrukcií osoba v miestnosti dokáže posielať von z miestnosti postupnosti čínskych znakov, ktoré sú korektnými odpoveďami na otázky (*výstup*).

Pripusťme, že pravidlá predstavujú úplnú množinu inštrukcií, implementovateľných na počítači a že generujú korektnú konverzáciu v čínštine. Osoba v miestnosti nevie, čo generujú pravidlá a ani nerozumie obsahu konverzácie, ktorej sa zúčastňuje. Zato tí, čo sú mimo miestnosti, majú dojem, že vnútri sedí niekto, kto rozumie čínsky.

To znamená, že *program* umožňuje osobe v miestnosti úspešne prejsť Turingovým testom (aplikovaným na schopnosť rozumieť čínsky), i keď nerozumie ani slovo.

Metaforu čínskej izby možno podľa Colea (2004) interpretovať dvojako – ako užší a širší argument. Zmyslom užšieho argumentu je, že žiadny počítačový program neumožňuje v pravom zmysle slova *rozumieť* (napr. čínštine).

Hlavná myšlienka (užšieho) argumentu je nasledovná: ak osoba v miestnosti nerozumie čínsky vďaka implementácii programu na pochopenie čínštiny, potom to

na tomto základe nedokáže ani žiadny počítač. Pretože žiadny počítač nedisponuje ničím navyše v porovnaní s osobou z čínskej izby.

Teda, manipulácia čínskych znakov na základe formálnych pravidiel nestačí na to, aby osoba alebo počítač rozumeli čínsky. Počítač (a umelá inteligencia) vykonáva iba zmyslu zbavenú manipuláciu symbolov.

Užší argument je zameraný proti koncepcii, ktorú Searle nazýva silnou umelou inteligenciou. Podľa tejto koncepcie je možné skonštruovať mysliaci počítač. Vhodne naprogramovaný počítač (resp. samotný program) má mentálne schopnosti podobné ľudským. Implementácia takéhoto programu je postačujúca pre myslenie. Počítač môže hrať šach inteligentne alebo rozumieť prirodzenému jazyku.

Podľa Searla však postup riadený nejakým algoritmom sám o sebe nemožno nazvať myslením (i vtedy, keď presne zodpovedá ľudským procesom myslenia). Nie je možné, aby počítače rozumeli prirodzenému jazyku alebo mysleli. K tomu, aby sme mohli hovoriť o porozumení, myslení, treba niečo navyše.

Užší argument proti silnej umelej inteligencii možno zhrnúť takto:

(1) ak silná umelá inteligencia je pravdivá, tak existuje program, ktorý umožňuje výpočtovému systému rozumieť čínsky,

(2) ja môžem postupovať podľa toho programu bez toho, že by som rozumel čínsky,

(3) teda silná umelá inteligencia je nepravdivá.

Premisa (2) sa opiera o užší argument čínskej izby. Podľa užšieho argumentu teda nemožno dosiahnuť pochopenie tým, že vykonávame nejaký program. Formálne výpočty na symboloch nemôžu produkovať myslenie.

Searlov argument nie je zameraný proti slabej umelej inteligencii. Slabá umelá inteligencia v Searlovom chápaní pripúšťa (iba), že programy (počítače) môžu realizovať zložité úlohy, ktoré obvykle vykonávajú inteligentné bytosti. Vďaka tomu programy/počítače umožňujú študovať kognitívne procesy tým, že ich simulujú. Netvrdí však, že počítače skutočne rozumejú alebo sú inteligentné.

Prejdime teraz k „širšiemu argumentu“ (k širšej interpretácii Searlovej metafory). Cieľom je ukázať, že sémantiku nemožno získať zo syntaxe (z formálnych manipulácií symbolov). Počítačový program, implementujúci konverzáciu v čínštine principiálne nie je schopný pochopiť významy, sémantiku. Searle to vyjadril takto:

Axióma 1: Počítačové programy sú formálne (syntaktické).

Axióma 2: Ľudské mysle majú mentálne obsahy (sémantiku).

Axióma 3: Syntax sama o sebe nie je konštitutívna, ani postačujúca pre sémantiku.
Záver: Programy nie sú ani konštitutívne, ani postačujúce pre myseľ.

K širšiemu argumentu sa vrátime v častiach *Programy a sémantika* a *Evolúcia a akvizícia významov*. Tu iba upozorňujeme na nedostatočne jasný obsah axióm 2 a 3 a na to, že metafora čínskej izby neukázala, že sémantiku prirodzeného jazyka nemožno špecifikovať sadou formálnych pravidiel – tie pravidlá, ktoré načrtol Searle tú ambíciu nemali.

Metafora čínskej izby je relevantná vzhľadom na viaceré filozofické otázky, ale aj pre diskusie o základoch kognitívnej psychológie a kognitívnej vedy. Vyvolala mimoriadnu pozornosť, rozprúdila množstvo diskusií; reagovali na ňu stovky článkov; podľa všetkého ide o najviac diskutovaný problém/prácu, týkajúcu sa filozofických základov umelej inteligencie i kognitívnej vedy

Na tomto mieste je vhodné stručne charakterizovať našu pozíciu. Netrápia nás ambície silnej umelej inteligencie – nezáleží nám na tom, či programy sú alebo nie sú mysľou. Naša pozícia je pozíciou slabej umelej inteligencie. Napriek tomu považujeme za zmysluplné pýtať sa, vďaka čomu (a v akom zmysle) možno hovoriť o inteligentnom správaní programov, o ich schopnosti rozumieť (najmä, keď sa na to pýta editor ☺).

Poznamenáme ešte, že keď v tejto kapitole hovoríme o (po)rozumení, nemáme na mysli iba rozumenie prirodzenému jazyku. Kontexty ako „rozumie, prečo to Robert urobil“, „nerozumie teórii relativity“, „títo dvaja futbalisti si na ihrisku rozumejú“ naznačujú, že je primerané chápať predikát *rozumieť* veľmi všeobecne. Nebudeme čitateľa unúvať pokusom o definíciu pomerne jasného pojmu pomocou pojmov nie oveľa jasnejších (poznať, vedieť, za akých podmienok a ohraničení). Keď to však bude dôležité, budeme sa usilovať o dostatočnú jasnosť.

Iná čínska izba

Searle si nerobí žiadny problém s významom slova rozumieť. Ten je však pre pochopenie (rozumenie) príbehu o čínskej izbe kľúčový. Na základe niektorých jeho formulácií sa možno domnievať, že rozumenie stotožnil s poznaním významov (vyjadril to predovšetkým negatívne – napr. syntax nie je konštitutívna, ani postačujúca pre sémantiku). Takéto vymedzenie však problém odsúva iba do inej roviny – čo znamená, že niekto pozná nejaký význam a ako sa o tom dá presvedčiť?

V časti *Evolúcia a akvizícia významov* uvidíme, ako možno (syntaktickými) prostriedkami, ktorými disponujú počítačové programy, dosiahnuť „spoznávanie významov“.

Ďalšia možnosť odhadnúť Searlove intuície sa opiera o jeho zdôrazňovanie úlohy ľudského biologického „hardvéru“ ako „kauzálnej sily“, ktorá generuje myslenie. Pokiaľ sa jasne nepovie, aký iný hardvér môže ešte generovať myslenie, dostávame tautologický argument typu „počítač nemôže rozumieť, myslieť, lebo počítač nemôže generovať myslenie“.

Jeden zo spôsobov, ako ľudia zisťujú, či niečomu rozumejú, je introspekcia. Ak (mlčky alebo nahlas) predpokladáme, že čínsky môže rozumieť iba entita, ktorá introspekciou zažíva evidenciu toho, že rozumie a takouto entitou je iba človek, nie je čo dokazovať a Searlova metafora nič nehovorí – nemá žiadny empirický obsah (tak ako každá tautológia). „Dokazuje“ iba to, čo predpokladá.

Searlov argument, že porozumenie a inteligenciu nemožno posúdiť behaviorálnym testom, ale len introspekciou, smeruje k tautologickým záverom typu „k mysleniu treba subjektívne vedomie a to má iba človek“. S ním sa ľahko vysporiadame otázkou: „Na základe čoho vieme, že Searle rozumie po anglicky?“ (Pfeifer, Scheier, 1999).

Vráťme sa však k otázke, ako rozumieť slovu „rozumieť“?

Odpoveď budeme motivovať metaforou (inej) čínskej izby. Predpokladajme, že v nejakej miestnosti je zamknutá a strážená skupina Číňanov. Spolu s nimi je v miestnosti Armén. Číňania zamýšľajú naplánovať a uskutočniť útek, nie sú si však istí, či Armén nerozumie čínsky a či ich nevyzradí. Inú možnosť, ako *pozorovanie jeho správania* a *interpretáciu pozorovaní* nemajú.

Preto sa rozhodnú preveriť hypotézu, že Armén rozumie čínsky a že to tají. Preverujú ju tak, že sa opierajú o nejakú „teóriu“, ktorá s vysokou pravdepodobnosťou dovolí predvídať nejaké (čo i len jemne rozlíšiteľné) mimické alebo motorické reakcie počúvajúceho na nejaké čínske vety, ak by im rozumel. O pravdepodobnosti hovoríme preto, lebo ani u ľudí nemusí byť isté, či v danej chvíli rozumejú (situácii, vete, textu, významu, problému, riešeniu, ...).

Teda: o tom, či niekto rozumie, sa môžeme presvedčiť tak, že pozorujeme *správanie*, ktoré zodpovedá hypotéze o pochopení (na základe nejakej dobre zdôvodnenej „teórie“). Zdôrazňujeme, že nejde o *čisto* behaviorálne kritérium: správanie samo o sebe nestačí, musí korešpondovať solídnemu poznaniu relevantného

kontextu a táto korešpondencia sa potvrdzuje/vyvracia nejakými usudzovacími operáciami.

Počítačový program ELIZA (Weizenbaum, 1966), ktorý imituje rolu psychoterapeuta v dialógu s klientom, by možno obstál v čisto behaviorálnom teste. Jeho šablónovité reakcie vyvolávajú zdanie, že počítač rozumie tomu, čo pacient hovorí. Bystrý pozorovateľ by však pravdepodobne rýchlo spozoroval povrchnú šablónovitosť, prijal zodpovedajúce hypotézy a našiel vstup do dialógu, ktorým by program „nachytal“. Uvedieme príklad takéhoto minidialógu s jednou z implementácií ELIZY (ide o preklad, pôvodný dialóg prebehol v angličtine). Čitateľ si to môže sám interaktívne vyskúšať na stránke http://www-ai.ijs.si/eliza.

E Ako ti môžem pomôcť?

U Som posadnutý dievčatami.

E Ako dlho si už posadnutý dievčatami?

U Niekoľko rokov.

E Prosím, pokračuj.

U Toto si povedala preto, že nerozumieš vete, ktorú som povedal?

E Och, toto som povedala preto, že nerozumiem vete, ktorú si povedal.

U Dávaš hlúpe odpovede, Eliza.

E Och, dávam hlúpe odpovede, Eliza.

Od čias ELIZY bolo naprogramovaných mnoho umelých konverzujúcich systémov, tzv. chatterbotov (http://en.wikipedia.org/wiki/List_of_Chatterbots). Niektoré z nich sa dokážu z dialógov učiť a postupne svoju komunikačnú schopnosť zdokonaľujú.

Uveďme iný príklad k problému rozumenia. To, či nejaký študent rozumie pojmu limita sa nedá preveriť iba tým, že odrapká definíciu. Ak nie je schopný *vypočítať* limity konkrétnych funkcií v konkrétnych bodoch a *zdôvodniť* tento výpočet (jeho korektnosť), nerozumie tomu pojmu („teória“, o ktorú sa pritom opierame umožňuje odfiltrovať napríklad numerické chyby pri našom rozhodovaní o porozumení). Opäť zdôrazníme úlohu usudzovacích operácií – tie sú nevyhnutné pri zdôvodnení korektnosti výpočtu. Tento príklad a táto analýza ukazujú, že realistický prístup k pojmu rozumenia akceptuje jeho stupňovitosť (rôzni študenti môžu rozumieť v rôznej miere, nie buď rozumieť alebo nerozumieť). V prírode je inteligencia (a kognitívne schopnosti vo všeobecnosti) distribuovaná stupňovito a mnohotvárne.

Upozorňujeme na už spomínanú kognitívnu biológiu Ladislava Kováča, podľa ktorého evolúcia je vlastne evolúciou poznania.

Argument, založený na inej čínskej izbe nie je principiálne nový. Podobné argumenty už sformulovali viacerí kritici Searlovej pozície. Sú to tie, ktoré Searle pomenoval ako The Other Minds Reply. Možnou formuláciou je napríklad: „Ako vieme, že iní ľudia rozumejú čínsky (alebo čomukoľvek inému)? Iba na základe ich správania. Počítač môže úspešne absolvovať behaviorálne testy podobne ako ľudia. Teda, ak sme ochotní pripísať kogníciu iným ľuďom, musíme byť ochotní pripísať ju aj počítačom."

Searlova odpoveď je v kocke taká, že predpoklady, ktoré sú akceptovateľné vzhľadom na ľudí, nemožno automaticky prenášať na počítače. Predpoklady, ktoré prijímame o našom vlastnom biologickom druhu, nemusia byť relevantné mimo neho. Presvedčenie, že sa to týka počítačov a rozumenia možno oprieť o to, že programy spracúvajú informácie syntakticky, kým ľudia (aj) sémanticky. K vzťahu syntaxe a sémantiky sa ešte vrátime.

Podobné črty ako má argument inej čínskej izby, možno nájsť aj v triede argumentov, pomenovaných The Intuition Reply. Podľa nich je Searlov argument založený na *intuícii*, že počítač (alebo spomínaná osoba v miestnosti) nemôže rozumieť. Intuície však môžu byť nespoľahlivé, zavádzajúce. Navyše, môže byť užitočné/potrebné zosúladiť náš koncept pochopenia so svetom, v ktorom by roboty boli z toho istého prírodného druhu ako ľudia (Block, 1980).

Simon a Eisenstadt (2002) tvrdia, že Searlov argument nevyvracia (nenapadá) tézu, podľa ktorej je možné naprogramovať počítač tak, že presvedčivo spĺňa obvyklé kritériá rozumenia.

Zhrňme: Nechceme pripísať kogníciu programom, počítačom. Budeme sa usilovať o to, aby sme ukázali, že správanie programov možno opísať a hodnotiť na základe obvyklých kritérií rozumenia (usudzovania a ďalších schopností vlastných inteligentnému správaniu). V prípade rozumenia máme na mysli toto: ak počítač úspešne absolvuje behaviorálne testy porozumenia (opreté o nejakú prijateľnú teóriu alebo hypotézu) podobne ako ľudia, tak sme ochotní toto jeho správanie považovať za správanie rovnocenné správaniu, založenému na pochopení/porozumení (v doméne, pre ktorú je daný test relevantný).

Inteligentné programy

Pozrime sa teda bližšie na programy, ktorých správanie by nás mohlo zaujímať. Začneme pomerne skromnými predstavami o inteligentnom správaní (programov).

Správanie programu možno jednoducho opísať tak, že zaznamenávame, akými výstupmi reaguje na aké vstupy. V súlade s mottom tejto kapitoly nebudeme trvať na tom, že správanie nejakých programov je inteligentné (i keď pre jednoduchosť vyjadrovania hovoríme o inteligentných programoch a ich inteligentnom správaní). Zdôrazňovať však budeme vysvetlenie, ako také správanie vzniká (vďaka čomu sa objavuje).

Predstavme si, že sme sa obrátili na poradcu v otázkach zdravotného poistenia (urobili sme tak v prostredí, kde je veľmi bohatá ponuka rozmanitých poistiek s rôznou mierou spoluúčasti klienta). Ak nám poradca bude schopný poradiť *iba* vtedy, ak mu povieme kód zdravotnej služby (napr. kód CT vyšetrenia krčnej chrbtice) a kód našej diagnózy a jeho odpoveďou bude, či máme našou poistkou pokrytú danú službu (prípadne s akou spoluúčasťou), pomyslíme si, že poradca príliš nerozumie oblasti zdravotného poistenia. Podobný názor budeme mať vtedy, keď poradca otvorí nejakú príručku, nájde nejakú pasáž a poskytne nám ju miesto odpovede na naše otázky.

Expertný systém, navrhnutý práve pre poradenstvo v oblasti zdravotného poistenia (Morgenstern, Singh, 1997) prišiel s ambíciou prekonať poradenské systémy, ktorých schopnosti boli presne také, ako schopnosti fiktívnych poradcov z predchádzajúceho odstavca: vyhľadávali podľa kódov v nejakej tabuľke alebo z nejakého textu vybrali časť, obsahujúcu informácie o nejakej podtéme (na základe výskytu nejakých kľúčových slov).

Pozrime sa detailnejšie na program, ktorý hľadá na základe vstupných kódov odpoveď v nejakej tabuľke. Prvá nepríjemná nevýhoda takéhoto programu je, že sa od užívateľa vyžaduje znalosť kódu. *Reprezentácia* domény a problému v prirodzenejších pojmoch je podmienkou flexibilnejšieho a vhodnejšieho - z ľudského hľadiska - správania programu. Druhou nevýhodou je pevne určená miera detailnosti (špecifickosti) poskytovaných informácií – možno sa pýtať iba na kategórie, ktoré sú zakódované; nemožno sa pýtať na (nezakódované) kategórie, ktoré sú všeobecnejšie (resp. špecifickejšie, sumárne). Po tretie, klienta zaujímajú otázky, často nepredvídateľné, na zodpovedanie ktorých treba bohatšie informácie a istú schopnosť *usudzovať*.

Citovaný expertný systém reprezentoval (vyjadril) doménu zdravotného poistenia v prirodzených pojmoch, zaznamenal (reprezentoval) aj pravidlá, ktoré

upresňovali, zjemňovali informácie o zdravotných službách, poistných produktoch a ich vzťahoch. Používané pojmy boli organizované v hierarchii. Špecifickejšie úrovne hierarchie dedia informáciu zo všeobecnejších úrovní. V doméne zdravotného poistenia sú rozmanité výnimky zo všeobecných pravidiel. Tie sú tiež zaznamenané (reprezentované). Usudzovanie o tejto hierarchii je *nemonotónne* v tom zmysle, že niektoré závery, opierajúce sa o isté predpoklady možno neskôr odmietnuť (*revidovať*), ak sa informácia, z ktorej vychádzame doplní (napríklad tak, že služba, o ktorú sa v danom kontexte zaujímame, predstavuje výnimku z pravidla, ktoré sme uplatnili).

Autori vytvorili program, ktorý sa správal uspokojujúcejšie ako jeho predchodcovia. Dokázal odpovedať na väčší rozsah otázok, aj na neočakávané otázky. Užívateľ mohol oprávnene získať pocit, že program lepšie „rozumie“ zdravotnému poisteniu ako jeho predchodcovia.

Teraz sa vrátime k otázke, čo nás zaujíma: vďaka akému vybaveniu, vďaka akým črtám/schopnostiam môžeme identifikovať v správaní nejakého programu čosi ako inteligenciu (rozumenie)? Alebo inak: vďaka čomu sa v správaní programu môže objaviť (vzniknúť) inteligencia (rozumenie)?

V našom príklade sa tak stalo vďaka uvedenej reprezentácii znalostí o doméne zdravotného poistenia (hierarchicky usporiadané poznatky doplnené o pravidlá) a vďaka implementácii schopnosti usudzovať s takto reprezentovanými znalosťami. Keďže usudzovacie schopnosti tohto programu sú nemonotónne (program je schopný revidovať svoje predchádzajúce úsudky po získaní nových informácií), môžeme hovoriť o tom, že jeho „inteligencia“ je blízka ľudskej inteligencii. Totiž, pre ľudskú inteligenciu je charakteristické to, že sa dokáže adaptovať na zmenenú situáciu, pretože dokáže revidovať závery svojich úsudkov.

Otázka, či takéto programy *skutočne* rozumejú alebo sú skutočne inteligentné, nás nezaujíma. Odlišnosti programov od živej prírody sú nesporné, ale nezaujímavé, triviálne. Podobnosti na živú prírodu, na človeka, sú netriviálne, zaujímavé. Tou podobnosťou je schopnosť pracovať s (nejako vyjadrenými, reprezentovanými) poznatkami, usudzovať. Zámerne sme vybrali program s pomerne skromným cieľom – implementovať nejakú reprezentáciu znalostí a usudzovanie. Oblasťou najintenzívnejšieho výskumu v umelej inteligencii v posledných dvoch desaťročiach je práve reprezentácia znalostí a usudzovanie, ak za mieru vezmeme počty článkov v najrešpektovanejšom časopise Artificial Intelligence Journal alebo zborníky

bienálne usporiadavaných vrcholných svetových konferencií IJCAI, či európskych ECAI. Napríklad na konferencii ECAI 2006 bolo 10 sekcií z celkového počtu 40 venovaných tejto oblasti.

Schopnosť usudzovať je jednou zo základných ľudských kognitívnych schopností. Viaceré spôsoby usudzovania sú dobre definované (sémanticky špecifikované). Ak nejaký program robí presne to, čo vyžaduje takáto špecifikácia, môžeme v presne definovanom zmysle slova povedať, že usudzuje.

Programy a sémantika

Zaujímavým príkladom inteligentného programu je SAPFO (Páleš 1994). Tento program dostane na vstupe nejakú slovenskú vetu a k nej vyprodukuje parafrázy – iné vety, ktoré majú približne zhodný význam. Správanie tohto programu považujeme za inteligentné z prostého dôvodu: povedať to isté inými slovami je tradičný spôsob, ako demonštrovať pochopenie u ľudí (na rozdiel od bezmyšlienkovitého zopakovania nejakej formulácie, napríklad odstavca učebnice). Neraz učiteľ žiada žiaka, aby to povedal „vlastnými slovami“. Ďalej, nepovažujeme za primerané takéto správanie považovať za simuláciu ľudského správania (tak ako to robí napríklad Searle). Pravdepodobne žiaden človek nedokáže vychrliť tak pohotovo také množstvo parafráz ako SAPFO. Napokon, chceme zdôrazniť, že ide o program, ktorého výstupy možno charakterizovať sémanticky (napriek tomu, že ich dosahuje formálnymi manipuláciami).

Vďaka čomu toto správanie „vzniklo“? Páleš odviedol pri návrhu programu veľký kus práce. Modeloval a reprezentoval všetky subsystémy slovenského jazyka. Kooperáciou týchto subsystémov dosiahol výsledné správanie programu. Kľúčová však bola charakterizácia sémantiky slovies slovenčiny pomocou tzv. intenčných rámcov. Jeho program si vyžiadal oveľa väčšiu mieru detailnosti (v tomto prípade poznania sémantiky slovenských slovies) než je obvyklé pri tradičných metódach (v tomto prípade lingvistického opisu). Vo všeobecnosti, programovanie vyžaduje extrémnu mieru detailnosti a explicitnosti reprezentácie pravidiel a faktov príslušnej domény práve preto, že program vykonáva čisto formálne manipulácie s reprezentáciami a nemá možnosť odvolávať sa na intuície, na samozrejmé vedomosti, na tzv. zdravý rozum. V prípade SAPFO tieto formálne manipulácie na formálnych reprezentáciách vyprodukovali parafrázovanie viet slovenčiny.

Samozrejme, človek dokáže parafrázovať bez toho, že by poznal Pálešove rámce a formálne pravidlá. Teda, rámce a pravidlá nie sú nevyhnutné pre dosiahnutie tejto schopnosti. To však nič nemení na tom, že je možné túto schopnosť na nich založiť.

Na záver – nevidíme nijaký dôvod správane programu SAPFO charakterizovať inými slovami ako *parafrázuje*. Nezáleží nám na tom, aby sme programu SAPFO pripísali myseľ, inteligenciu. Vďaka tomu, že SAPFO disponuje reprezentáciou sémantiky slovenských slovies a ďalších lingvistických znalostí a vďaka operáciám na týchto reprezentáciách môže parafrázovať. Podobne Deep Blue, Deep Fritz alebo iný šachový program *skutočne hrá šach*. Nepotrebujeme sa sporiť, či má myseľ, nepotrebujeme hľadať kontraargumenty proti tvrdeniam, že nemá introspekciu, intencionalitu alebo vetry, že vykonáva iba formálne manipulácie symbolov. Treba si však uvedomiť, že Deep Blue nedokáže parafrázovať alebo SAPFO nedokáže hrať šach či robiť stručné zhrnutia z rozsiahlejších textov.

Uvedieme ďalší príklad „inteligentného“ programu. Ide o systém, ktorý bol implementovaný pre potreby NASA a podporuje rozhodovanie o riešení problémov so Space Shuttle. V prípade, že boli namerané (zistené) nejaké poruchy, systém vytvoril plán na ich odstránenie (Balduccini et al., 2002, Nogueira, 2003). Samozrejme, na konštrukciu plánu je potrebné nejaké usudzovanie. Tu sa chceme sústrediť na vzťah usudzovania a sémantiky. Každá dobre opísaná formalizácia (implementácia) usudzovania vychádza zo sémantickej špecifikácie. S formalizáciou môžeme byť spokojní až vtedy, keď je korektná a úplná vzhľadom na sémantickú špecifikáciu. Korektnosť znamená, že to, čo odvodíme syntaktickými manipuláciami, vyhovuje aj sémantickým kritériám. Úplnosť sme dosiahli vtedy, keď všetko, čo na základe sémantických kritérií vyplýva z nejakých predpokladov, odvodíme z týchto predpokladov aj našou formalizáciou.

Program, o ktorom hovoríme, patrí do triedy logických programov. To znamená, že je skonštruovaný z pravidiel tvaru: „ak platia predpoklady …, potom platí záver ...“. Výpočet prebiehal podľa paradigmy „answer set programming“. Vypočítajú sa stabilné modely programu, to znamená množiny všetkých pravdivých a nejakými pravdivými dôvodmi podopretých elementárnych formúl (atomárnych, ďalej neanalyzovateľných výrokov). Potom sa na otázky kladené programu, (pravdivo) odpovedá na základe vypočítaných modelov.

Problém, ktorý nás tu zaujíma, je problém, či s pomocou programu (syntaktického objektu spracovávaného na základe syntaktických pravidiel) možno

vytvoriť (zrekonštruovať, získať) sémantiku. Pojmy pravdivosti a modelu sú sémantické, takže (pozitívne) riešenie tohto problému je zrejmé.

Možná výhrada proti sémantike stabilných modelov je tá, že je definovaná iba v nejakom formálnom jazyku a že chýba ukotvenie pojmov z formálneho jazyka do sféry reálnych kognícií (*symbol grounding problem).* Tento problém sformuloval Harnad (1990) a týka sa vzťahu symbolov k reálnemu prostrediu a toho, odkiaľ a ako naberú symboly spracúvané umelým systémom sémantiku. K problému sa vrátime v časti *Evolúcia a akvizícia významov*.

Konekcionistická alternatíva

Pre porovnanie spomenieme aj príklad inteligentného systému spomedzi konekcionistických výpočtových architektúr, ktoré predstavujú alternatívnu výpočtovú paradigmu ku „klasickej" (symbolovej) výpočtovej paradigme. Najprv si však stručne povedzme o čo ide. Počiatky konekcionizmu siahajú do obdobia vzniku počítačov, no reálny vplyv tejto paradigmy nastal až po prekonaní niektorých zásadných problémov, a to v druhej polovici 80-tych rokov minulého storočia (Rumelhart, a spol., 1986). Typický konekcionistický model pozostáva z množiny vzájomne prepojených uzlov (umelých neurónov), ktoré sú zvyčajne usporiadaných do vrstiev, a ktoré paralelne šíria aktivitu v sieti prostredníctvom adaptovateľných spojení (takému systému bežne hovoríme aj neurónová sieť). Neurónová sieť je príkladom subsymbolovej paradigmy, pretože (zväčša) využíva distribuované reprezentácie elementov pomocou ich príznakov (t.j. v podobe vektora čísel). Napríklad v jazykovej doméne môže byť takým elementom počuté slovo, ktoré bude reprezentované pomocou svojich fonologických charakteristík. S výpočtovou paradigmou súvisia nielen používané reprezentácie, ale aj možné operácie nad nimi. Na subsymbolové reprezentácie v neurónovej sieti aplikujeme maticové operácie, ako aj rôzne nelineárne transformácie, zatiaľ čo so symbolmi narábame podľa (prepisovacích) pravidiel, pričom štruktúra ich reprezentácie je irelevantná, pretože aplikácia pravidiel to nevyžaduje. So symbolmi možno pracovať v neurónovej sieti, tam im však musíme priradiť nejakú numerickú reprezentáciu. Napríklad, keď máme k dispozícii 5 vstupných neurónov, distribuovaný tvar môže vyzerať ako (0.8 0.1 0.3 0.7 1.0), zatiaľ čo lokalistický ako (0 0 0 1 0).

Používanie subsymbolových reprezentácií je prístupom ku (biologicky prijateľnejšiemu) kognitívnemu modelovaniu „zdola“, pretože tie sú odvodené z prostredia. V podstate všetky údaje prichádzajúce do mozgu existujú v numerickej, t.j. subsymbolovej forme. Napríklad, písmeno napísané na papieri sa zobrazí na sietnici ako obraz, počuté slovo je sluchovým nervom prenášané do mozgu v podobe frekvenčných a časových charakteristík. Až po sérii zložitých transformácii v asociačných oblastiach kôry z týchto numerických reprezentácií vzniká (nejakých spôsobom) reprezentácia entity, s ktorou vieme narábať ako so symbolom (podľa nejakých stanovených pravidiel). Napriek subsymbolovému charakteru neurálnych reprezentácií je symbolový pohľad na (vyššiu) kogníciu užitočnou abstrakciou, pretože symbolová reprezentácia znalosti a použiteľné metódy v súčasnosti umožňujú efektívnejšie riešenie niektorých problémov UI než neurónové siete (napr. šachový program DeepBlue už niekoľkokrát porazil Kasparova).

Neurónové siete sú svojou povahou „šité na mieru“ pre úlohy nižšej kognície (ako napr. rozpoznávanie vzorov), kde bol urobený veľký pokrok (napr. Bishop, 1995). Ich hlavnou výzvou je oblasť problémov „vyššej“ kognície, a preto zámerne uvedieme príklad konekcionistického modelu z lingvistickej domény (iné, jednoduchšie modely možno nájsť napr. vo Farkaš, 2005). Model INSOMNet (Incremental Non-monotonic Self-organizing Network; Mayberry, 2003) s modulárnou architektúrou bol navrhnutý ako interpretátor anglických viet, ktorý predloženej vete v písanej forme (prezentovanej slovo za slovom) priradí jej sémantickú reprezentáciu. Reprezentácie viet s tzv. plochou sémantikou (použité na trénovanie výstupného modulu s externým učiteľom) boli zobraté z lingvisticky anotovaného korpusu, v ktorom každá veta má podobu grafu (pomocou uzlov a orientovaných hrán). Takýto graf vyjadruje syntaktické a sémantické vzťahy medzi vetnými členmi, pričom každý vetný člen je reprezentovaný 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 modelov), a (2) podobne ako symbolové programy využívajú smerníky (ukazovatele), no tie sú implementované tiež pomocou aktivácií viacerých neurónov. Obtiažnosť problému spočíva v tom, že aj smerníky sa sieť musí naučiť ako súčasť reprezentácie vety. Model sa musí tiež naučiť gramatiku, pretože tá nie je explicitne daná (na rozdiel od systému SAPFO). Natrénovaný systém je

psychologicky prijateľný, lebo (podobne ako človek) sa učí na príkladoch, generuje predikcie, nemonotónne reviduje interpretáciu vety pri jej inkrementálnej analýze (podobne ako vyššie spomínaný expertný systém), správa sa robustne (voči chybám), dokáže paralelne vytvárať viacnásobné interpretácie v prípade nejednoznačných viet, a vie modelovať rôzne jazykové poruchy. V skratke, INSOMNet je psychologicky prijateľným modelom, pretože jeho silné a slabé stánky dosť korešpondujú s jazykovými schopnosťami ľudí. A môžeme povedať, že systém „rozumie“ vetám? Ak za prejav porozumenia považujeme schopnosť zovšeobecňovať, správne reagovať na nové podnety, tak áno, pretože systém dokáže väčšine nových viet priradiť správnu sémantickú reprezentáciu. Preto mu v tomto zmysle môžeme prisúdiť inteligentné správanie. Model INSOMNet sa trénuje na pomerne rozsiahlej doméne, aj keď veľmi pracne: trénovanie trvá asi dva mesiace na súčasnom výkonnom stolovom počítači. Z pohľadu komplexnosti predstavuje aktuálny stav konekcionistického výskumu v lingvistike. Sem by sme ešte zaradili i trochu odlišný, ale tiež psycholingvisticky prijateľný modulárny model CSCP (Rohde, 2002), ktorý si dokáže vytvárať sémantické reprezentácie anglických viet a vie aj v istom zmysle odpovedať na nové otázky (tak, že doplní chýbajuce slovo vo vete-otázke), čiže ide o „porozumenie“ v podobnom slova zmysle ako v modeli INSOMNet.

Avšak ani konekcionistické modely samy o sebe neriešia vyššie spomínaný problém ukotvenia symbolov, pretože neinteragujú aktívne s prostredím (sú to pasívne skrinky na spracovanie informácie). Systém INSOMNet síce disponuje aj samoorganizáciou (pri klasterizácii sémantických grafov), ktorá vytvára predpoklady pre ukotvenie symbolov, pretože umožňuje vytváranie prirodzených kategórií (t.j. nestanovených dizajnérom systému). Tie však nemajú pre systém žiaden význam (ak nevedú k nejakej zaujímavej odozve) a musia byť interpretované vonkajším pozorovateľom. Ukotvený systém musí generovať výstupy, ktoré budú ovplyvňovať jeho správanie. To sa dá dosiahnuť vnorením systému do prostredia, čo by si v kontexte INSOMNet-u vyžadovalo pridanie produkčného modulu, ktorý by vytváral vety a tým by ich sémantiku ukotvil priamo v komunikácii s prostredím. Model CSCP síce produkčný modul má, ale ten nezasahuje priamo do prostredia.

Evolúcia a akvizícia významov

Autonómny systém vnorený v prostredí a dosahujúci v ňom určité ciele pomocou vykonávania akcií budeme ďalej označovať slovom *agent* (Kelemen, 1994). Teraz predstavíme model (Takáč, 2006a), v ktorom je sémantika ukotvená v senzomotorických interakciách agenta s prostredím. Prostredie je simulované – agent nie je robot vybavený kamerou a motorčekmi, napriek tomu vníma svoje prostredie a dokáže meniť jeho charakteristiky vykonávaním akcií. Prostredie je dvojrozmerná mriežka, na ktorej sú náhodne rozmiestnené objekty 4 typov: ovocie, hračky, nábytok a sám agent. Agent na začiatku žiadne typy nerozlišuje, dostáva iba informáciu o vnímateľných vlastnostiach objektov vo svojom okolí v podobe rámcov – množín párov <atribút : číselná hodnota> (atribútmi sú napr. poloha v priestore, veľkosť, hmotnosť, farba, tvar, atď.). V modeli beží čas v diskrétnych krokoch a v každom kroku agent dostane jeden rámec pre každý vnímaný objekt. Agent môže v každom kroku vykonať nejakú akciu zo svojho repertoára, konkrétne zdvihnúť alebo položiť nejaký objekt. Akcie sú parametrizované – agent si vyberá s akou silou bude zdvíhať a do akej výšky zdvihne resp. spustí rameno (maximálna sila a výška sú obmedzené „konštrukciou“ agenta).

Agent tiež dokáže pozorovať dôsledky svojich akcií (porovnaním hodnôt v perceptuálnych rámcoch manipulovaného objektu v aktuálnom a predchádzajúcom časovom kroku). Skutočné dôsledky agentových akcií reguluje prostredie na základe simulovaných (veľmi primitívnych) fyzikálnych zákonov, napr. ak sa agent pokúša zodvihnúť objekt príliš malou silou vzhľadom ku hmotnosti objektu, objekt sa nepohne. Agent vykonáva akcie náhodne a postupne sa učí rozlišovať vlastnosti prostredia a združuje do kategórií entity, ktoré sú z nejakého hľadiska podobné.

Každá kategória je reprezentovaná tzv. rozlišovacím kritériom - aktivačnou funkciou, ktorej hodnota vyjadruje stupeň príslušnosti vstupu funkcie k reprezentovanej kategórií. Vstupom funkcie môže byť rámec objektu (kritérium objektov a vlastností), viacerých objektov (kritérium vzťahu), rámcov toho istého objektu v rôznych časoch (kritérium zmeny) aj výsledky iných rozlišovacích kritérií (kompozičné kritériá situácií a udalostí). Na začiatku agent nemá žiadne kritériá – tie sú budované na základe príkladov extrakciou ich spoločných štatistických vlastností. Ak agent vykoná nejakú akciu na rôznych objektoch s dostatočne podobným výsledkom, považuje tieto objekty za príklady tej istej kategórie (vzhľadom na danú akciu a výsledok). Podobným spôsobom združuje do kategórií aj rámce s parametrami vykonávaných akcií.

Komplexnejšie (propozičné) znalosti agenta o kauzálnych vzťahoch sú reprezentované ako asociácie rozlišovacích kritérií kategórií objektov, akcií a ich dôsledkov. Tieto znalosti umožňujú agentovi vyberať si vhodné akcie a predikovať ich dôsledky, pričom kategoriálna reprezentácia je efektívnejšia ako keby si pamätal každý konkrétny objekt a akciu.

Takáto reprezentácia je v súlade s poznatkami o ľudskej organizácii reprezentácie znalostí. Podľa Gibsona (1979) dokážeme v prostredí priamo vnímať možnosti akcií a manipulovania s objektmi: po zemi sa dá chodiť, dvere sa dajú otvoriť, jablko sa dá zjesť atď. Vnímateľné interakčné vlastnosti predmetov Gibson nazval *afordanciami*. V našom modeli im zodpovedajú dôsledky asociované s objektovými kategóriami. Podľa Tomasella (1992) je mentálny lexikón dieťaťa spočiatku organizovaný okolo konkrétnych slovies s konkrétnymi scenármi vykonávania. Takéto štruktúry Tomasello nazval *slovesnými ostrovmi*. V našom modeli slovesné ostrovy tvoria objekty a dôsledky asociované s akciami.

V implementovanom simulačnom experimente si agent vytvoril štyri kategórie objektov, ktorých sémantiku by sme vzhľadom na asociované kauzálne znalosti mohli opísať ako „veci príliš ťažké, aby sa dali zdvihnúť“, „veci, ktoré sa nedajú položiť nižšie, lebo sú už na zemi“. Do tretej kategórie patrila väčšina ovocia a do štvrtej väčšina hračiek. Experiment modeluje senzomotorické štádium (Piaget, Inhelder, 1966), v ktorom dieťa vykonáva množstvo nekoordinovaných a náhodných pohybov, čím spoznáva objekty okolo seba aj svoje možnosti interagovania s nimi. Dôležitým prínosom experimentu je, že agent je vybavený len učiacimi mechanizmami, ale znalosti o konkrétnom prostredí získava, až keď je mu vystavený. To mu umožňuje flexibilné prispôsobenie sa rôznym prostrediam, ktoré nie sú známe vopred a navyše sa môžu v čase meniť (predstavme si úlohu naprogramovať robota, ktorý bude vyslaný na vzdialenú planétu s neznámymi podmienkami).

Druhou dôležitou vlastnosťou je, že model možno priamočiaro rozšíriť o jazykovú komunikáciu s ukotvenou sémantikou. Ak by niekto (učiteľ, rodič, iný agent) komentoval to, čo agent vníma a robí, ten by si mohol vznikajúce kategórie asociovať s počutými jazykovými výrazmi – symbolmi (Takáč, 2006b). Sémantika symbolov by tak bola ukotvená v interakčnej skúsenosti agenta s prostredím a inými agentmi.

Systém, ktorý sme práve predstavili, mal zjednodušenú úlohu v tom, že zo simulovaného prostredia dostával vstupy akoby predspracované – v podobe rámcov

s atribútmi a hodnotami. Stelesnené systémy operujúce v reálnom prostredí, napr. roboty, dostávajú zo senzorov analógové a navyše zašumené signály a úlohy spojené s percepciou a spracovaním takýchto signálov sú veľmi netriviálne. Rovnako je netriviálne analógové riadenie efektorov agenta (napr. motorčekov mechanickej ruky, či pohybovacieho zariadenia).

Metódy výpočtovej inteligencie

Na to, aby bol agent schopný extrahovať z komplexného a dynamického prostredia užitočné informácie (ak teda neuvažujeme zjednodušenie, že vstupy bude mať „predžuté"), musí v prvom rade disponovať percepčným aparátom, ktorý mu to umožní, a až potom uplatniť svoje „programy" či neurónové siete pri interakcii s prostredím. Tieto rôzne „inteligentné" metódy spracovania informácie by sa dali označiť termínom *výpočtová inteligencia* (VI). Aj keď v súčasnosti neexistuje zhoda v tom, čo by sa malo skrývať pod pláštikom VI, sympatickou sa javí snaha stanoviť jej definíciu nie podľa používaných metód, ale podľa riešených problémov. Ak za cieľové problémy budeme považovať kognitívne schopnosti, ktorými človek alebo iný živý tvor disponuje, vytvára sa priestor pre zjednotenie všetkých formálnych prístupov v rámci VI, pokrývajúcich všetky metódy umelej inteligencie, výpočtové metódy inšpirované prírodou (evolučné algoritmy, umelé neurónové siete a fuzzy systémy), ale aj rôzne iné matematické a štatistické metódy (Duch, 2007a). Cieľom VI by teda mohli byť výpočtovo riešiteľné problémy, obzvlášť tie, pre ktoré neexistujú efektívne riešenia, a to buď kvôli ich výpočtovej obtiažnosti alebo kvôli obtiažnosti ich algoritmizácie. Mnohé takéto problémy človek zvláda, a teda potrebuje ich zvládať aj umelý agent v komplexnom prostredí.

Návrh nových metód VI rýchle napreduje, no je tu priestor pre ďalšie zlepšovanie (Duch, 2007b). Vo väčšine metód VI sa nastavujú parametre s cieľom optimalizovať správanie systému. Metódy optimalizácie samozrejme závisia od typu metódy, no v súčasnosti sa už na riešenie daného problému aplikujú súčasne viaceré metódy, pričom bránový modul vyberá tú najlepšiu (zmes expertov). Kombinácia rôznych metód v rámci systému smeruje ku heterogénnym adaptívnym systémom. Túto snahu podporuje aj „no-free-lunch" teoréma, z ktorej vyplýva, že neexistuje univerzálny učiaci algoritmus, ktorý by bol lepší vo všetkých prípadoch ako nejaký iný algoritmus (Duda a spol., 2001). Preto by inteligentný systém mal disponovať rôznymi metódami a optimálne vyberať medzi nimi podľa potreby.

Hľadanie spoločných čŕt

Na začiatku tejto kapitoly sme sa zaoberali ľudskou inteligenciou a jej fylogenetickými počiatkami. Potom sme predstavili niekoľko príkladov umelých systémov, ktorých správanie by sme mohli nazvať inteligentným. Zistili sme pritom, že niektoré črty robia správanie systému či programu viac inteligentným od programov, ktoré tieto črty nemajú. V tejto časti sa pokúsime zosumarizovať a čiastočne usporiadať črty potrebné k (stupňovitému) inteligentnému správaniu.

Všetky predstavené systémy, rovnako ako všetky živé organizmy, sú vnorené v nejakom prostredí (pojem prostredie používame v širšom zmysle, takže nezahŕňa len reálne fyzické prostredia, ale napr. aj informačné prostredie v pamäti počítačov, či na internete) a majú nejaký účel (napr. prežiť, alebo odpovedať na otázky používateľa). Prostredie je dynamické (jeho stav sa mení v čase) a vzhľadom na systém externé (v prostredí sa menia aj veci, ktoré systém nemá pod kontrolou, napr. správanie užívateľa, ktorý interaguje s počítačovým programom). Systém je s prostredím spriahnutý pomocou vstupu (senzie) a výstupu (akcie). Prvým záverom teda je, že inteligentné správanie môžeme hľadať (iba) u systémov, ktoré na základe vnímania stavu prostredia a pomocou správania ovplyvňujúceho jeho budúci stav dosahujú určité ciele. Takéto systémy sme nazvali agentmi. Keďže pojmy senzia, akcia a prostredie používame v širšom význame, ako je obvyklé, táto definícia je dosť všeobecná na to, aby pokryla baktériu Escherichia Coli pohybujúcu sa v smere gradientov teploty a kyslosti prostredia, softvérový filter monitorujúci internet a blokujúci niektoré prichádzajúce pakety, umelú neurónovú sieť verifikujúcu elektronické podpisy, program odpovedajúci na otázky používateľa, aj ľudské správanie. Schopnosť rozlišovať a reagovať je teda nutnou podmienkou aj toho najelementárnejšieho inteligentného správania.

Medzi reaktívnym správaním baktérie či predprogramovaného internetového filtra a správaním človeka je však dôležitý rozdiel. Cyklus *senzia, akcia a následná senzia* poskytuje systému spätnú väzbu o účinkoch jeho akcií na prostredie. Agent, ktorý pri detekcii neúspechu vie adaptívne zmeniť svoje správanie, má šancu byť v dosahovaní svojich cieľov úspešnejší ako agent s fixným správaním (a podľa príslovia o somárovi, ktorý sa nevie poučiť z vlastných chýb, prisúdime adaptívnemu agentu viac inteligencie ako neadaptívnemu). V perfektne stabilnom prostredí, ktorého vlastnosti sú dané vopred, má šancu na úspech aj neadaptívny agent, čo je príklad väčšiny „klasických“ softvérových programov. Dokonca ich výhodou môže byť väčšia rýchlosť. Ak sa však vlastnosti prostredia radikálne zmenia, doteraz

úspešné správanie nemusí byť viac adekvátne. Vrodené správanie biologických organizmov je výsledkom adaptácie v priebehu evolúcie. Analógom fylogenetickej adaptácie u umelých systémov je konštruktérsky vývojový cyklus: ak sa naprogramovaný systém neosvedčí, návrhár skonštruuje/naprogramuje novú verziu a „vypustí“ ju do prostredia. Pre úspešnosť takýchto systémov je dôležitá relatívna stabilita prostredia voči rýchlosti „fylogenetickej“ adaptácie. Ak sa prostredie mení príliš rýchlo, takáto adaptácia nestačí. Systém musí byť schopný zmeniť repertoár svojich správaní za behu (počas ontogenézy). Takúto adaptáciu správania nazveme učením.

Prvoradou podmienkou učenia je exploratívnosť, resp. flexibilita – schopnosť meniť správanie. Schopnosť reagovať úplne novým spôsobom na situáciu, na ktorú systém nebol skonštruovaný, môžeme nazvať invenciou. Invencia (aj v umelých systémoch) často v sebe obsahuje prvok náhodnosti, nedeterminizmus a metódu pokus - omyl. Sama osebe ale k učeniu nestačí: systém, ktorý by permanentne generoval náhodné správanie, má invenciu, ale neučí sa.

Ďalšou podmienkou je schopnosť udržať a neskôr reprodukovať správania, ktoré sa osvedčili. Okrem už spomínaného spätnoväzobného cyklu s prostredím na to systém potrebuje pamäť. Na veľmi všeobecnej úrovni môžeme pamäť definovať ako časť štruktúry systému, ktorá sa môže meniť počas ontogenézy (behu systému) a ktorej zmena má kauzálny vplyv na správanie systému. Výskum formálnych modelov výpočtových zariadení ukázal, že ich principiálna výpočtová sila závisí práve od typu dostupnej pamäte. Na najnižšom stupni hierarchie sú konečnostavové automaty – zariadenia s konečným počtom preddefinovaných stavov zodpovedajúce reaktívnym systémom; nasledujú automaty s jedným zásobníkom, s dvoma zásobníkmi alebo zapisovacou páskou atď. (Hopcroft, Ullman, 1969).

Príklad: Malé dieťa pri hre v kolíske vykoná tisíce náhodných pohybov rúčkou, kým sa mu podarí uchopiť hrkálku. Postupným opakovaním sa vydarené uchopenie fixuje a stáva sa použiteľnou súčasťou repertoára správania dieťaťa. Avšak učenie nie je iba osvojovaním nových zručností. Zasahuje aj percepčné procesy. Agent môže potrebovať znovurozpoznať už videný objekt a adekvátne naň zareagovať. Alebo môže existovať celá trieda objektov, stavov prostredia či situácii, na ktoré treba zareagovať rovnakým spôsobom (a odlišne od iných tried objektov či situácií). Faktorizácia (rozdelenie celého nekonečného spojitého spektra možných vstupov na konečný počet tried) je ekonomickejšia a znamená menšie zaťaženie

pamäte. Zároveň dáva agentu väčšiu robustnosť voči šumu (drobné rozdiely medzi členmi kategórie zväčša nie sú podstatné), schopnosť aproximovať a zovšeobecňovať (predpovedať u objektu ďalšie vlastnosti na základe členstva v kategórii a tiež možnosť zaradiť úplne nový vstup do najpodobnejšej kategórie a podľa toho naň zareagovať) a zvyšuje stabilitu komunikácie (pri odlišnosti vnímania a skúseností sa aktéri komunikácie skôr zhodnú na úrovni diskrétnych kategórií ako na úrovni jednotlivých perceptov). Zhrňme: pre agenty je šikovnejšie namiesto nekonečného spojitého spektra manipulovať s diskrétnymi entitami – kategóriami. Kategóriu, ktorá pre agent zastupuje nejakú triedu vstupov, nazveme *reprezentáciou* týchto vstupov a vstupy reprezentované kategóriou nazveme *referentmi* kategórie (kategórie môžu mať aj prázdnu, či jednoprvkovú množinu referentov). Pojem reprezentácie neskôr rozšírime.

Združovanie do kategórií môže prebiehať rôznymi spôsobmi, podľa dostupnosti spätnej väzby z prostredia. Ak agent dostáva iba vstupy z prostredia, ale nemá spätnoväzobnú slučku, môže prirodzeným spôsobom vytvoriť kategórie tak, aby odrážali štatistické rozdelenie vlastností prostredia. Prirodzená kategorizácia maximalizuje medzikategoriálne rozdiely a minimalizuje rozdiely v rámci kategórie (teda priemerné množstvo rozdielov medzi členmi tej istej kategórie má byt čo najmenšie a medzi členmi rôznych kategórií čo najväčšie). V umelých systémoch sa vytváranie prirodzených kategórií nazýva učenie pozorovaním (observational learning), učenie bez učiteľa (unsupervised learning), alebo klasterizácia.

Prirodzená kategorizácia nemusí byť pre ciele systému postačujúca: napr. živočích, ktorý sa živí hubami (a chce prežiť), sa musí naučiť vnímať jemné odlišnosti vo vzhľade aj veľmi podobných húb, pokiaľ niektoré sú jedlé a iné jedovaté. Tu je vhodným typom kategorizácia na základe pragmatickej spätnej väzby. Systém interaguje s entitami v prostredí a kategórie vytvára na základe výsledkov interakcie (pôvodné vstupy sa teda zobrazia do priestoru vnímaných výsledkov, na ktorých sa vykoná prirodzená kategorizácia: do tej istej kategórie padnú objekty, rovnaká interakcia s ktorými vedie k podobným výsledkom).

Tretím typom je kategorizácia s pomocou externého učiteľa. Napríklad dieťa si môže vyformovať kategóriu objektov, ktoré jeho rodič pomenúva tým istým menom – takto sa deti učia väčšinu nadradených kategórií ako zviera, či nábytok (Waxman, Braun, 2005).

Okrem kategórií priamo založených na perceptuálnej skúsenosti sú niektoré typy agentov schopné vytvárať ešte ďalšie kategórie pomocou modifikujúcich transformácií a zobrazení, napr. zúžením, zlúčením, nahradením časti, kompozíciou, abstrakciou, metaforou, metonýmiou (Lakoff, 1987).

Kategórie sú základnými prvkami reprezentácie. Z príkladu s rozlišovaním jedovatých a jedlých húb je zrejmé, že agent si potrebuje pamätať aj propozičné znalosti, teda tvrdenia o kategóriách, ktoré môžu byť pravdivé alebo nepravdivé v agentovom prostredí. Propozičné znalosti sú budované nad kategóriami a reprezentujú agentove viery, domnienky či poznatky o vzťahoch (hierarchických, kauzálnych, ...) a vlastnostiach prostredia. S nimi agent (s dostatočne inteligentným správaním) uskutočňuje rôzne usudzovacie operácie umožňujúce predikciu vlastností, odvodzovanie dôsledkov a v neposlednom rade revíziu platných tvrdení na základe nových informácií.

Peter Gärdenfors (1996) rozdeľuje reprezentácie na vyvolané (cued, online) a oddelené (detached, offline). Vyvolaná reprezentácia sa vyskytuje vždy spolu so svojím referentom, ktorý musí byť aktuálne prítomný v externej situácii systému. Živočích disponujúci vyvolanou reprezentáciou svojej potravy ju dokáže rozlíšiť od nejedlých objektov a adekvátne zareagovať – správanie sa však spustí iba v prítomnosti potravy.

Významným mechanizmom, ktorý nahrádza limitovanú pamäť organizmu, je ukladanie značiek do prostredia. Príkladom môžu byť pachové stopy, ktoré zviera zanecháva v prostredí a neskôr sa podľa nich orientuje, alebo u ľudí známy „uzlík na vreckovke“. V oboch prípadoch je do prostredia uložená značka, ktorá neskôr spustí – *vyvolá* príslušnú reprezentáciu.

Oddelená reprezentácia môže ovplyvniť správanie systému aj v neprítomnosti svojho referentu. Napríklad šimpanz, ktorý nevie dočiahnuť banán, odíde hľadať palicu (ktorú naokolo nevidí), keď ju nájde, vráti sa a strhne pomocou nej banán, musí mať oddelenú reprezentáciu palice aj možností jej použitia. Existencia oddelených reprezentácií je nevyhnutným predpokladom vzniku vyšších kognitívnych funkcií ako sú plánovanie, lesť, sebauvedomenie a komunikácia (Gärdenfors, 1996). Zo všetkých živočíchov zašiel človek v externalizovaní svojich reprezentácií najďalej: kultúra, umenie, knihy, počítače, internet, ... Je nesporné, že tieto výdobytky akcelerujú ďalší rozvoj ľudských schopností.

Plánovanie spočíva v schopnosti systému „mentálne“ (t. j. na úrovni oddelených reprezentácií) vyhodnotiť predpokladané dôsledky rôznych variantov správania a vybrať sekvenciu akcií najlepšie vyhovujúcu stanovenému cieľu. Dobré plánovanie musí zvažovať aj dôsledky akcií iných systémov. Schopnosť klamať predpokladá reprezentáciu iných kognitívnych systémov nie ako konajúcich vecí, ale ako systémov s vlastnými reprezentáciami, plánmi, atď. (teda akúsi „teóriu mysle“). Klamár musí mať aj reprezentáciu toho, ako ho pravdepodobne bude klamaný vnímať, čo znamená reprezentáciu reprezentácie samého seba. Takáto metareprezentácia je nutným predpokladom sebauvedomenia.

Napokon, medzi najvyššie kognitívne schopnosti založené na oddelených reprezentáciách patrí jazyková komunikácia. Jazyk je symbolový systém umožňujúci externalizáciu vnútorných reprezentácií a ich komunikovanie. Podľa základnej premisy kognitívnej sémantiky nie sú symboly jazyka mapované priamo na externé prostredie, ale prostredníctvom vnútorných reprezentácií (Lakoff, 1987, Gärdenfors, 2000). Práve oddelenosť vnútorných reprezentácií umožňuje rozšíriť komunikáciu nad rámec „tu a teraz“, na druhej strane vytvára paradox: ako sa môžu agenty navzájom dorozumieť, ak ich individuálne systémy reprezentácií nie sú rovnaké? Ak v rozhovore pod tým istým slovom myslí každý partner niečo iné, najpravdepodobnejším dôsledkom bude nedorozumenie. Kľúč je opäť vo väzbe na prostredie: Reprezentácie agentov nie sú úplne arbitrárne, ale kódujú vlastnosti prostredia, v ktorom sú agenty vnorené. Samotná podoba a realizácia reprezentácií nie je podstatná, úspešnosť porozumenia sa v konečnom dôsledku prejaví pragmaticky v správaní aktérov komunikácie. Porozumenie, že symboly jazyka majú nejaký sémantický obsah, spočíva v schopnosti správať sa k týmto symbolom spôsobom primeraným ich obsahu (van Gulick, 1988).

Záver

Čitateľovi je iste jasné, že formulácia problému z motta tejto kapitoly je nám blízka. Skúsme teda na záver zhrnúť, čo považujeme za dobré odpovede na otázku, ako vzniká správanie, ktoré je pre nás (v tomto kontexte ☺) zaujímavé. Začnime kontrastom. Tu nás nezaujíma správanie balíka štatistických programov alebo pomyselný program, ktorý by hral štandardné šachové otvorenia dovtedy, kým by sa jeho súper od štandardného otvorenia neodchýlil. Takéto programy (a milióny ďalších podobných) robia – z ľudského hľadiska – rutinnú prácu. Buď pomáhajú

odbremeniť človeka od nej alebo ju realizujú aj vtedy, keď človek naráža na svoje limitované kapacity.

Nás zaujímali systémy, ktoré dokázali hrať šach lepšie ako majstri sveta, hľadať sémantickú reprezentáciu anglických viet, vytvárať významy na základe interakcie s prostredím alebo plánovať (a mnohé podobné ďalšie). Správanie takýchto systémov buď priamo alebo nepriamo má na výber z veľkého množstva alternatív. Podobne ako človek, ktorý je často odkázaný skúšať rôzne možnosti a následne korigovať omyly. Takéto – v istom zmysle slova nedeterministické – správanie bolo pre nás zaujímavé. Počítač má tú výhodu, že to môže robiť s obrovskou rýchlosťou. Treba však povedať, že mnohé úlohy umelej inteligencie sú pre vstupy väčšieho rozsahu na hranici súčasných výpočtových možností alebo za ňou. Aj takéto negatívne zistenia však podstatne prispievajú k nášmu poznaniu.

Aké odpovede o vzniku takéhoto správania považovať za dobré? Medzi ne pravdepodobne patria aj také, o ktoré sme sa tu pokúšali. Odpovede, ktoré nahliadnu do „vnútra“ takýchto systémov a ukážu, aká reprezentácia (problému, prostredia, znalostí) umožňuje takéto správanie.

Odporúčaná literatúra

Balduccini, M., Gelfond, M., Noguiera, M., Watson, R. (2002): Planning with the USA-Advisor. In *3rd International NASA Workshop on Planning and Scheduling*

Bell, A. (1999): Levels and loops: The future of artificial intelligence and neuroscience. *Philosophical Trans-actions of the Royal Society London B,* 354, 2013-2020.

Bishop, C. M. (1995): *Neural Networks for Pattern Recognition.* Oxford University Press.

Block, N. (1980): komentáre k Searle (1980). In: *Behavioral and Brain Sciences* 3, 417-457.

Calvin, W. H. (1996): *How Brains Think: Evolving Intelligence, Then and Now*. Basic Books.

Calvin, W. H., Ojemann, G. A. (1994): *Conversations with Neil's Brain. The Neural* Nature *of Thought and Language,* Addison-Wesley.

de Chardin, P. T. (1956): Le phénomène humain. Les Éditions du Seuil, Paris. (Preklad: de Chardin, P. T.: *Vesmír a lidstvo.* Vyšehrad, Praha, 1990.)

Cole, D. (2004): heslo "The Chinese Room Argument". In: Zalta, E. N. (zost.) *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), URL = <http://plato.stanford.edu/archives/fall2004/entries/chinese-room/>.

Denett, D. (1995): *Darwin's Dangerous Idea.* The Pinguin Press, Hammondsworth.

Duch, W. (2007a): Towards comprehensive foundations of computational intelligence. *Lecture Notes in Computer Science,* Springer.

Duch, W. (2007b): What is computational intelligence and what would it become? Lecture *Notes in Computer Science*, Springer.

Duda, R. O., Hart, P. E., Stork, D. G. (2001): *Pattern Classification.* J. Wiley & Sons, New York.

Farkaš, I. (2005): Konekcionistické modelovanie jazyka. Kapitola v knihe *Jazyk a kognícia*, Kalligram, Bratislava, str. 262-305.

Gärdenfors, P. (1996): Cued and detached representations in animal cognition. *Behavioural Processes,* 36, 263-273.

Gärdenfors, P. (2000): *Conceptual Spaces,* MIT Press, Cambridge MA.

Gibson, J. J. (1979): *The Ecological Approach to Visual Peception.* Houghton Mifflin, Boston.

Goodwin, B. C. (1978): A cognitive view of biological process. *Journal of Social and Biological Structures*, 1, 117-125.

Harnad, S. (1990): The symbol grounding problem. *Physica D* 42, 335–346.

Hopcroft, J. E., Ullman, J. D. (1969): *Formal languages and their relation to automata*. Addison-Wesley. (Preklad: Hopcroft, J. E., Ullman, J. D.: Formálne jazyky a automaty. Alfa, Bratislava, 1978.)

Jung-Beeman, M. a spol. (2004): Neural activity when people solve verbal problems with insight. *PLoS Biology*, 2(4): e97.

Kelemen, J. (1994): *Strojovia a agenty*. Archa, Bratislava.

Kohonen, T. (1984): *Self-organization and Associative Memory*. Springer.

Kováč, L. (1986): Úvod do kognitívnej biológie. *Biologické listy* 51 (3), 172-190.

Kováč, L. (2003): Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu. In: Kelemen, J. (zost.): *Kognice a umělý život. III.* Slezská univerzita, Opava.

Kováč, L. (2006): Princípy molekulárnej kognície. In: Kelemen, J., Kvasnička, V. (zost.): *Kognice a umělý život VI.* Slezská univerzita, Opava.

Lakoff, G. (1987): *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press. (Preklad: Lakoff, G.: Ženy, oheň a nebezpečné veci. Triáda, Praha, 2006.)

Mayberry, M. R. (2003): *Incremental Nonmonotonic Parsing through Semantic Self-Organization.* Doktorská dizertačná práca, The University of Texas at Austin.

Morgenstern, L., Singh, M. (1997): An Expert System Using Nonmonotonic Techniques for Benefits Inquiry in the Insurance Industry. *Proc. of 15th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann.

Neill, J. T. (2004): *Nature vs nurture in intelligence*. Wilderdom. http://www.wilderdom.com/personality/L4-1IntelligenceNatureVsNurture.html

Nogueira, M. (2003): *Building Knowledge Systems in A-Prolog*. PhD thesis, University of Texas at El Paso.

Páleš, E. (1994): *SAPFO. Parafrázovač slovenčiny. Počítačový nástroj na modelovanie v jazykovede*. VEDA, vydavateľstvo SAV. Bratislava.

Pfeifer, R., Scheier, Ch. (1999): *Understanding Intelligence*, Bradford Books.

Piaget, J., Inhelder, B. (1966): La psychologie de l'enfant. PUF, Paris. (Preklad: Piaget, J., Inhelder, B.: Psychológia dieťaťa. SOFA, Bratislava, 1997.)

Rohde, D. (2002): *A Connectionist Model of Sentence Comprehension and Production.* Doktorská dizertačná práca, Carnegie Mellon University, Pittsburgh, PA.

Rumelhart, D. E., McClelland, J. L. a výskumná skupina PDP (1986): *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume I. Cambridge, MA: MIT Press.

Searle, J. R. (1980): Minds, Brains, and Programs. *Behavioral and Brain Sciences* 3, 417-424.

Searle, J. R. (1984): *Minds, Brains and Science*. Harvard University Press.

Searle, J. R. (1990): Is the brain's mind a computer program? In: *Scientific American* 262, 26-31.

Searle, J. R. (1999): The Chinese Room. In: Wilson, R.A. and F. Keil (zost.), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge: MIT Press.

Simon, H., Eisenstadt, S. (2002): 'A Chinese Room that Understands'. In: Preston, Bishop (zost.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. New York: Oxford University Press.

Šefránek, J. (2002): Kognícia bez mentálnych procesov. In: Ľ. Beňušková a kol. (zost.): *Kognitívne vedy*. Kaligram, Bratislava, str. 200-256.

Takáč, M. (2006a): Categorization by Sensory-Motor Interaction in Artificial Agents. In: Fum, D., Del Missier, F., Stocco, A. (zost.): *Proceedings of the 7th International Conference on Cognitive Modeling*, Edizioni Goliardiche, Trieste.

Takáč, M. (2006b): Kognitívna sémantika rozlišovacích kritérií. In: J. Kelemen, V. Kvasnička (zost.): *Kognice a umělý život VI*, Slezská univerzita, Opava.

Tomasello, M. (1992): *First verbs: A case study of early grammatical development*. CUP, Cambridge.

Turing, A. M. (1950): Computing Machinery and Intelligence. Mind LIX. (Preklad: Turing, A. M.: Počítacie stroje a inteligencia. In: Gál, E., Kelemen, J. (zost.): *Myseľ, telo, stroj*. Bradlo, Bratislava, 1992.)

van Gulick, R. (1988): Consciousness, intrinsic intentionality and self-understanding machines. In: Marcel, A. J., Bisiach, E. (zost.): *Consciousness in Contemporary Science*. Clarendon Press, Oxford. (Preklad: van Gulick, R.: Vedomie, vlastná intencionalita a stroje, ktoré rozumejú samy sebe. In: Gál, E., Kelemen, J. (zost.): *Myseľ, telo, stroj*. Bradlo, Bratislava, 1992.)

von der Malsburg, C. (2003): Self-organization and the brain. In: *Handbook of Brain Theory and Neural Networks* (M. Arbib, ed.), MIT Press, 1002-1005.

Waxman, S. R., Braun, I. (2005): Consistent (but not variable) names as invitations to form object categories: new evidence from 12-month-old infants. *Cognition* 95, B59-B68.

Weizenbaum, J. (1966): ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.