

Konstruktivistický prístup k štúdiu kognície

Martin TAKÁČ¹

Abstrakt. Schopnosť rozumieť – pripisovať význam – je kľúčovým aspektom kognície. Pre štúdium rôznych foriem kognície – živých aj umelých – je užitočné disponovať takým pojmovým a teoretickým aparátom, ktorý nie je založený na výlučnosti ľudského rozumenia. V tejto kapitole sa pokúsime takýto pojmový aparát vybudovať a budeme ho aplikovať na analýzu širokej škály kognitívnych systémov od jednoduchých predverbálnych organizmov až po ľudí a softvérové agenty, či roboty. Ukážeme, že nutnou podmienkou rozumenia je situovanosť v prostredí a schopnosť učiť sa. Vysvetlíme mechanizmy takéhoto učenia a uvedieme ich konkrétne výpočtové modely.

1 Úvod

Kľúčovou vlastnosťou kognitívnych resp. inteligentných systémov je schopnosť rozumieť, pripisovať význam [79]. Preto bude táto kapitola najmä o významoch. Štúdiom významov v rôznych typoch kognitívnych systémov – v živočíšnej ríši, u ľudí aj u umelých systémov sa môžeme naučiť veľa o kognícii ako takej. Aj preto, že významy sa často spájajú s komunikáciou a ľudský jazyk je najdokonalejší komunikačný systém, ktorého ovládanie je späté s mnohými kognitívnymi funkciami. Spôsob, akým budeme kogníciu a významy študovať, je konstruktivistický – aby sme pochopili, ako niektoré kognitívne mechanizmy fungujú, pokúsime sa ich (alebo aspoň ich zjednodušenú podobu) replikovať v počítačových simuláciách. Pôjde teda o „chápanie tvorením“ [60].

Slovo konstruktivistický v názve tejto kapitoly má však aj ďalší dôvod: ako ukážeme, schopnosť učiť sa a samostatne konstruovať významy na základe skúseností je pre kogníciu kľúčovou – významy, ktoré sú predprogramované resp. dané zvonka nikdy nebudú kognitívnemu systému vlastné a teda ťažko hovoriť o nejakom rozumení.

Rôzne typy agentov – baktérie, zvieratá, ľudia, niektoré počítačové programy a roboty – majú niečo spoločné: dosahujú ciele pomocou vnímania a konania vo svojom (reálnom či virtuálnom) prostredí [42]. Niektoré z nich dokážu komunikovať medzi sebou či dokonca s ľuďmi. Do akej miery rozumejú tomu, čo robia? Pokiaľ pripisujú

¹ Katedra aplikovanej informatiky FMFI UK, Mlynská dolina, 848 48 Bratislava, E-mail: takac@ii.fmph.uniba.sk

Martin Takáč

nejaký význam situáciám a udalostiam v prostredí, aká je povaha tohto významu? Používajú rovnaké významy aj keď komunikujú? Odkiaľ sa berú významy? Sú vrodené (predprogramované), alebo naučené? Takéto otázky nás budú zaujímať.

Niektorí autori odmietajú používať pojmy ako „rozumienie“ či „význam“ v súvislosti s inými ako ľudskými agentami. Často považujú za predpoklad rozumenia intencionalitu alebo vedomie [67]. Takýto prístup z definície vylučuje niektoré typy agentov, alebo aspoň celú vec ešte viac zahmlieva redukciami problému významu na ešte ťažšie (keďže detekcia alebo dôkaz vedomia či intencionality u nie ľudských agentov sú problematické). My sa preto pokúsime o odlišný prístup. Budeme hľadať čo najneutrálnejšiu a najmenej antropocentrickú charakterizáciu rozumenia, aby bola použiteľná pre štúdium významov aj u predverbálnych živých organizmov a umelých systémov. Tento prístup je v súlade so snahami definovať život a vedomie všeobecnejším spôsobom tak, že ľudský život a ľudské vedomie sú len jednou z jeho možných inštancií [3, 38, 51].

Po krátkom úvode do formálnych teórií významu prejdeme lekciami z predverbálnych štádií fylogenézy a ontogenézy k štúdiu senzomotorickej inteligencie zvierat a malých detí. Počiatky významu totiž treba hľadať práve tu, omnoho skôr ako sa na evolučnej scéne objavil jazyk. Potom opíšeme problémy s významom u umelých systémov a budeme analyzovať niektoré príklady. Hlavne nás však bude zaujímať otázka pôvodu významov, preto preskúame možnosti a limity konštruktivistického prístupu metodológiou výpočtového modelovania.

Prínos takéhoto prístupu je mnohoraký. Po prvé, žijeme v časoch, keď komunikácia a interakcia medzi človekom a strojom, či medzi strojmi navzájom nie je sci-fi, ale praktický inžiniersky problém. Potrebujeme navrhovať také reprezentačné formalizmy, ktoré nám umožnia vybaviť stroje/počítače ontológiami nutnými pre úspešné riešenie daných úloh a pre vzájomnú koordináciu. Reprezentácia musí byť dostatočne komplexná, aby postihla špecifiká materiálnych a sociálnych prostredí, vrátane ich dynamického charakteru. V otvorených prostrediach je schopnosť učiť sa autonómne konštruovať reprezentáciu relevantných znalostí kľúčovou. Po druhé, operacionalizácia sémantických teórií a budovanie relevantných výpočtových modelov pomáha vyjasniť pojem „rozumienia“, ktorý je zdrojom kontroverzií v umelej inteligencii od jej počiatku, a poskytuje mechanizmy ukotvenia symbolov. A nakoniec, konštruovanie výpočtových modelov nám môže pomôcť lepšie porozumieť sebe samým. Úspešné modely môžu mať spätný vplyv na teórie učenia a jazykového vývinu aj na kognitívnu vedu ako takú.

2 Teórie významu

Filozofia a lingvisti sa otázkou „čo to znamená *znamenat'* niečo“ zaoberajú už mnoho storočí. Dnes sa štúdiu významu venuje najmä sémantika a semiotika. V tejto časti predstavíme niekoľko vplyvných teórií, ktoré dávajú na otázku povahy významov a ich vzťahu k jazyku rôzne odpovede.

2.1 Funkcionalistická sémantika

Základnou otázkou je, či existujú nejaké objekty – materiálne, mentálne, ideálne, atď., ktoré konštituujú význam jazykových výrazov. Sémantické teórie, ktoré na túto otázku odpovedajú kladne, sú *denotačné*; tie, čo odpovedajú záporne, nazveme *nedenotačné*. Príkladom nendenotačného prístupu je funkcionalistická sémantika, ktorej základy položil Wittgenstein. Podľa (neskorého) Wittgensteina, významom slov nie je referencia na objekty alebo veci označené vo vonkajšom svete, ani žiadne idey či mentálne reprezentácie s nimi asociované, ale spôsob ich používania v komunikácii [101]. Jazyková výpoveď je v prvom rade aktom v reálnom svete, význam je preto viac vecou pragmatiky ako sémantiky. Wittgenstein prirovnáva použitie jazykového výrazu k ťahu v hre podľa nejakých pravidiel.

2.2 Realistická sémantika

V denotačných sémantikách sú jazykovými významami nejaké objekty. Podľa názoru na povahu týchto objektov môžeme rozlíšiť *realistické* a *kognitivistické* (alebo *konceptualistické*) prístupy. Podľa realistického prístupu sú významami entity „vo vonkajšom svete“ (porovnaj obr. 2a). Podľa kognitivistického prístupu sú významy mentálne entity „v hlave“ (obr. 2b).

V *extenzionálnych* realistických sémantikách sa význam definuje vzťahom k objektom vo svete, alebo formálnejšie k modelovej štruktúre M . Mená sa zobrazia na konkrétne objekty – prvky M , predikáty sa zobrazia na množiny objektov resp. relácie v M . Pomocou kompozície sa vety zobrazia na pravdivostné hodnoty. Dôsledkom tohto prístupu je, že významy sú objektívne, spoločné a nezávislé od pochopenia jednotlivých používateľov. Základy extenzionálnej sémantiky položil Tarski [87].

Čoskoro však boli objavené medzery množinovo-teoretického prístupu, pretože niektoré javy (kompozície predikátov, čas, modality) sa nedali napasovať na extenzionálnu definíciu významu. Pokusom o nápravu problémov extenzionálnej sémantiky bol návrh tzv. *intenzionálnej sémantiky* [15, 46, 55], v ktorej sa prvky jazyka zobrazujú na množiny možných svetov namiesto jediného. Tvrdenie – propozícia sa definuje zobrazením z možných svetov na pravdivostné hodnoty, určujúcim množinu svetov, v ktorých je dané tvrdenie pravdivé.

2.3 Kognitívna sémantika

Gärdenfors charakterizuje kognitívnu sémantiku nasledovne [28]:

1. Význam je konceptuálna štruktúra kognitívneho systému (nie podmienky pravdivosti v možných svetoch).
2. Konceptuálne štruktúry sú stelesnené (význam nie je nezávislý od percepcie a telesnej skúsenosti).
3. Sémantické elementy sa konštruujú z geometrických alebo topologických štruktúr (nie zo symbolov skladaných na základe nejakého systému pravidiel).

Martin Takáč

4. Kognitívne modely sú predovšetkým obrazovo-schematické (nie propozičné). Obrazové schémy sú transformované metaforickými a metonymickými operáciami [48].
5. Sémantika predchádza syntax a čiastočne ju determinuje (syntax nemožno opísať úplne nezávisle od sémantiky).
6. V kontraste s Aristotelovským poňatím kategórií založenom na nutných a postačujúcich podmienkach, koncepty vykazujú prototypové efekty [65].

Prvé dva body vymedzujú, že porozumenie jazyka je integrálnou súčasťou konceptuálneho systému, ktorý slúži aj na usudzovanie, orientáciu a konanie vo svete [47], teda nemôže byť zabezpečené nejakým izolovaným jazykovým modulom (v zmysle Fodora [25]). Na neurálnej úrovni koreluje s aktiváciou štruktúr zúčastňujúcich sa na vnímaní, konaní, predstavách, či rozprávaní o reprezentovanom obsahu [63, 64].

Geometrické štruktúry reprezentujúce sémantické elementy (bod 3) Gärdenfors umiestňuje do tzv. *konceptuálneho priestoru*. Konceptuálny priestor pozostáva z množstva dimenzií ako farba, výška tónu, teplota, váha, tri priestorové dimenzie, ktoré zodpovedajú vlastnostiam reprezentovaných objektov. Dimenzie dôležité pre naše prežitie sú pravdepodobne vrodene (vyvinuté v priebehu evolúcie) a vbudované do nášho nervového systému. Iné dimenzie sú naučené a niektoré z nich môžu byť kultúrne podmienené. Dimenzie sú organizované do domén (pre reprezentované objekty môžu byť relevantné dimenzie iba z niektorých domén). Základnou vlastnosťou konceptuálnych priestorov je to, že podobné objekty sú umiestnené blízko pri sebe. Konkrétny objekt je reprezentovaný ako bod v podpriestore zloženom z jednej alebo viacerých domén. Vektor súradníc bodu určuje hodnoty vlastností na jednotlivých dimenziách. Podobnosť dvoch objektov je nepriamo úmerná vzdialenosti ich bodových reprezentácií v konceptuálnom priestore. (Prírodné) kategórie sú reprezentované (konvexnými) oblasťami v priestore. Geometrické centrá oblasti zodpovedajú najlepším resp. typickým reprezentantom kategórií – prototypom.

Gärdenfors vybudoval pomocou konceptuálnych priestorov vlastnú teóriu kognitívnej sémantiky [28]. Medzi iné vplyvné teórie patrí silová dynamika [86], rámcová sémantika [23], mentálne priestory [21] a kognitívna gramatika [49, 50].

Teraz sa bližšie pozrieme na Šefránkovu biologicky motivovanú kognitívnu teóriu reprezentácie [79]. Deklarovaným cieľom tejto teórie je vybudovať netriviálnu falzifikovateľnú úroveň analýzy kognície a rozumenia abstrahujúcu od mozgu a neurálnych procesov. Teória namiesto neurálnych procesov pracuje s obsahmi kognície – význammi. Jej kľúčovým predpokladom je, že významy sú oddeliteľné od jazyka, teda existujú aj u zvierat a detí v predverbálnom vývinovom štádiu. Všeobecne povedané, skúma významy u organizmov/agentov situovaných v nejakom prostredí. Organizmy majú potreby a ciele, ktoré naplňajú svojim správaním (vykonávaním akcií). Organizmy disponujú reprezentáciami zloženými z významov – *rozlišovacích kritérií*. Rozlišovacie kritériá sú abstrakciou schopnosti organizmu rozlišovať (rozpoznať, identifikovať) isté aspekty svojho (vnútorného či vonkajšieho) prostredia. Elementárne kritériá rozlišujú objekty (individua), prirodzené triedy a prirodzené vlastnosti objektov, a prirodzené vzťahy medzi objektami. Kompozíciou a transformáciami sa dajú z elementárnych kritérií konštruovať komplexnejšie kritériá

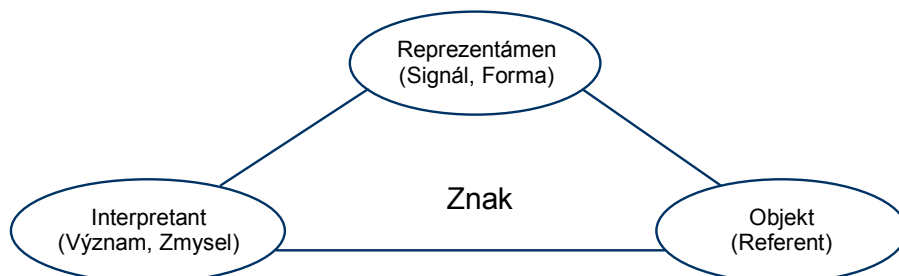
situácií, pravidiel (typov situácií), cieľov (želaných situácií), zmien v prostredí, plánov (projektovaných zmien), metód (osvedčených plánov), udalostí a typov udalostí.

Šefránek ďalej načrtáva cestu ako sa dostať od protosémantiky, protoinferencie a protokomunikácie jednoduchých organizmov cez dvojslovný protojazyk až k plnému ľudskému jazyku s propozičnou reprezentáciou a rekurzívnou syntaxou. Schopnosť rozumieť komplexnému jazyku je úzko prepojená s usudzovaním, presnejšie s hypotetickým (nemonotónnym) usudzovaním [34].

2.4 Semiotika

Podľa semiotických teórií vytvárame význam tvorbou a interpretáciou *znakov*. Znak môže mať podobu slov, obrázkov, zvukov, aktov, či objektov, ale stávajú sa znakmi len ak im pripíšeme nejaký význam. „Nič nie je znakom, pokiaľ to nie je interpretované ako znak“ [59]. A tiež, čokoľvek môže byť znakom, ak to niekto interpretuje ako *označujúce* niečo, teda odkazujúce na niečo alebo zastupujúce niečo iné ako samo seba [16]. Peirce definuje znak ako triádu – tri časti poprepájané reláciami:

- *reprezentámen* je forma, ktorú znak nadobúda,
- *interpretant* je zmysel, ktorý má znak pre nejakého interpretátora,
- *objekt* je to, na čo znak odkazuje.



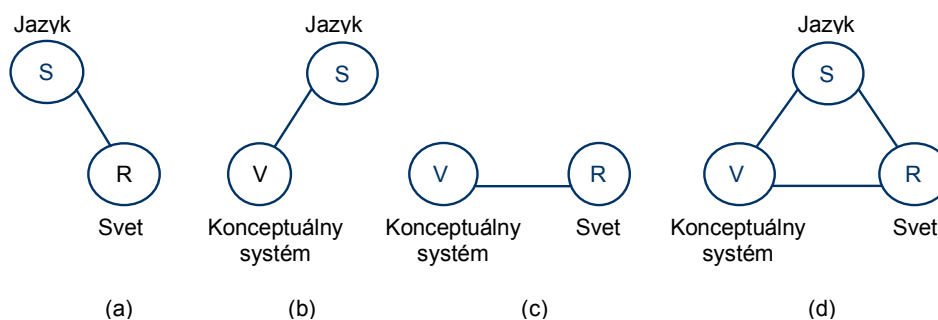
Obr. 1. Semiotický trojuholník

Triadické poňatie znaku sa niekedy označuje ako *semiotický trojuholník* (obr. 1). (Odteraz budeme namiesto pôvodnej terminológie *reprezentámen*, *interpretant*, *objekt* používať zaužívanejšie označenia *signál*, *význam*, *referent*.)

Je dôležité zdôrazniť, že znak v Peirceovom poňatí nie je absolútna alebo ontologická vlastnosť veci, ale je to relačná situovaná interpretatívna rola, ktorú vec nadobúda v konkrétnom kontexte. Čo je pre jedného pozorovateľa znakom, môže byť pre iného len zbytočným alebo nevnímateľným šumom, v závislosti od pozorovateľovho stelesnenia (embodiment), spoločnosti a interakčnej histórie. Takýto pohľad má blízko k *enaktívnemu prístupu* [53]. Enaktívny prístup je asi najsilnejšie vyjadrený v dynamickej kognitívnej paradigme [89], ktorá nahliada na kogníciu ako na komplexný dynamický systém (resp. viacero spriahnutých dynamických systémov)

Martin Takáč

a „zdôrazňuje časový rozmer kognície a emergenciu samoorganizovaných vzorcov správania jedinca ako stabilných stavov v interakciách mozgu, tela a prostredia” [31]. Takáto perspektíva ide naprieč hranicami medzi mozgom, telom a svetom. Niektoré dynamické modely celkom odmietajú reprezentácie, iné ich považujú za dynamické entity (napr. momentálne stavy systému, trajektórie či atraktory a ich spádové oblasti). „Reprezentácie sú dočasné, kontextovo závislé stability uprostred zmeny a nie statické, nemenné a na kontexte nezávislé jednotky” [30].



Obr. 2. Interpretácia neúplných verzií semiotického trojuholníka „význam (V) – signál (S) – referent (R)”. a) Realistická sémantika umiestňuje významy do sveta. b) V „čisto” kognitívnej sémantike sú lexikálnymi významami interné reprezentácie. Bez vzťahu k svetu by však takáto sémantika trpela *problémom ukotvenia symbolov* [37]. c) Predverbálne konceptualizácie sveta môžu predchádzať jazyk. d) V ukotvenej kognitívnej sémantike sú vnútorné významy formované v interakciách so svetom.

3 Od predverbálnych významov k jazyku: evolučný pohľad

V rámci kognitívnej sémantiky sa významy považujú za súčasť konceptuálneho systému, ktorý bol tvarovaný skúsenosťou s okolitým svetom. Z toho vyplýva, že rozumenie nezačína až na úrovni jazyka: môžeme hovoriť o rozumení a významoch aj na predverbálnej úrovni (obr. 2c).

Pozrime sa teraz z pohľadu evolúcie na vznik rozumenia v živých organizmoch. Niektorí vedci postulujú elementárne formy kognície na veľmi hlbokých úrovniach fylogenetického stromu, na bakteriálnej, bunečnej, či dokonca molekulárnej úrovni [35, 44]. Z organizmov, ktoré sa objavili v priebehu evolúcie, pretrvali tie, ktorých štruktúra zodpovedala relevantným vlastnostiam ich prostredia. Biologická evolúcia spočíva v generovaní hypotéz o povahe prostredia a vo falzifikovaní týchto hypotéz [44]. V tomto zmysle organizmy, ktoré pretrvali, stelesňujú hypotézy, ktoré neboli zatiaľ falzifikované. Na evolučnú adaptáciu organizmov sa teda dá pozeráť ako na formu fylogenetického učenia, pričom znalosť sa uchováva v štruktúre (stelesnení) organizmov. Šance pretrvať sú vyššie u tých organizmov, ktoré aktívne explorujú svoje prostredie a prispôbujú sa podmienkam, alebo ich dokážu aktívne pretvárať na

Konstruktivistický prístup k štúdiu kognície

vhodnejšie svojim správaním. Najjednoduchšie kognitívne systémy pozostávajú z mechanizmov vnímania (senzie) a konania (akcie). U nich môžeme za najelementárnejší spôsob „pripisovania významu“ považovať to, že rozpoznávajú v prostredí pomocou svojich sensorov a aktuátorov informáciu relevantnú svojim cieľom resp. účelu [56]. Zložitejšie kognitívne mechanizmy – percepcia, afekcia a kognícia – sú postupne vkladané medzi senziu a akciu v procese interkalárnej evolúcie [45]. Myslenie sa objavuje na najvyšších stupňoch evolúcie ako abstraktná akcia – testovanie rôznych motorických akcií bez skutočnej realizácie svalmi. Deliberácia (zámer), krátkodobá anticipácia, proto-plánovanie, a konečne aj “čo-keby” myslenie (schopnosť mentálne simulovať rôzne scenáre a vyhodnotiť ich dôsledky bez nutnosti ich fyzickej realizácie) zvýšili šance organizmov na prežitie a poskytli im významnú evolučnú výhodu.

Vhodným teoretickým konštruktom na analýzu adaptívneho správania je pojem *repräsentácie*. Gärdenfors rozlišuje dva typy reprezentácie nutné pre evolúciu jazyka: *vyvolanú* (cued) a *oddelenú* (detached) [26]. Vyvolaná reprezentácia je vždy spustená nejakým externým stimulom prítomným v aktuálnej situácii organizmu. Organizmus, ktorý diferencovane reaguje na rôzne stavy svojho prostredia (napr. konzumuje objekty rozpoznané ako potravu a vyhýba sa objektom rozpoznávaným ako predátori) vykonáva (prínajmenšom implicitnú) kategorizáciu – hovoríme, že disponuje vyvolanými reprezentáciami príslušných kategórií. Vyvolanosť reprezentácií spočíva v tom, že sú vždy aktivované len v prítomnosti svojich referentov. Vyvolané reprezentácie sú deskriptívno-vysvetľujúcim konštruktom pre vrodené vôľou neovplyvniteľné diferencované behaviorálne reakcie, ktoré majú fylogenetický pôvod.

Dôležitým mechanizmom, ktorý rozširuje limitované pamäťové možnosti organizmov, je ukladanie externalizovaných pamäťových značiek do prostredia. Príkladom sú pachové stopy, ktoré pomáhajú zvieratám v orientácii, či u ľudí notoricky známy uzlík na vreckovke (a napokon vynález kníh a iných záznamových médií, ktoré výrazne akcelerovali ľudskú kultúru). Externalizovaná značka uložená do prostredia neskôr aktivuje príslušnú vyvolanú reprezentáciu.

Oddelená reprezentácia môže zastupovať objekty či udalosti, ktoré nie sú aktuálne prítomné ani nesúvisia so súčasnou situáciou v prostredí. Napríklad šimpanz, ktorý chce dočiahnuť banán a odíde preto hľadať dlhý konár, musí disponovať oddelenou reprezentáciou konára aj možnosti jeho použitia. Predpokladá sa, že oddelené reprezentácie sa vo fylogénéze objavujú približne s vývojom neokortexu [27]; v ontogenéze korešpondujú s vytvorením *stálosti predmetu* [62].

Disponovanie oddelenými reprezentáciami je nutnou podmienkou pre vyššie kognitívne funkcie ako plánovanie, klamanie, sebauvedomenie a jazyková komunikácia [26]. Plánovanie predpokladá schopnosť organizmu „mentálne“ (t.j. pomocou oddelených reprezentácií) vyhodnotiť dôsledky rôznych behaviorálnych scenárov a zvolíť postupnosť akcií, ktorá najviac vyhovuje aktuálnym cieľom. Dobré plánovanie by tiež malo zvažovať dôsledky akcií iných agentov. Predstieranie, resp. klamanie predpokladá schopnosť reprezentovať iné agenty aj s ich vlastnými reprezentáciami, zámermi, plánmi atď., teda akúsi „teóriu mysle“. Klamár musí taktiež disponovať reprezentáciou toho, ako ho budú vnímať klamané agenty. Týmto

Martin Takáč

vnoreným spôsobom vlastne disponuje (meta)reprezentáciou samého seba, čo je nutným prekursorom sebauvedomenia [26].

Za jednu z najvyšších kognitívnych schopností sa považuje jazyková kompetencia. Jazyk je symbolový znakový systém, ktorý umožňuje externalizáciu a komunikáciu oddelených reprezentácií. Práve oddelenosť reprezentácií umožňuje vyjadriť významy prekračujúce aktuálne „tu a teraz“, dokonca aj významy, ktorých referenty v reálnom svete neexistujú (napr. ježibaba, či trojuholník s tromi pravými uhlami). Podľa Gärdenforsa bola najsilnejším evolučným impulzom pre vznik jazyka potreba spoločenstva koordinovať sa pri spoločnom dosahovaní budúcich cieľov [29].

Koncepty, v ktorých je ukotvená sémantika jazyka, môžu mať rôzny pôvod. Tie, ktoré boli v priebehu evolúcie podstatné pre prežitie, môžu byť vrodené [24]. Iné sú konštruované v priebehu ontogenézy senzomotorickou interakciou s prostredím a pozorovaním dôsledkov vlastných akcií [2, 62]. Ešte iné koncepty sú prenášané kultúrne a ich osvojenie je stimulované samotným jazykom [99]. Vplyv jazyka na proces formovania významov v rámci detského lexikálneho vývinu má aj empirickú podporu [97].

4 Reprezentácia významov v umelých systémoch

4.1 Základné problémy

Vytvorenie prvých umelointeligentných systémov, ktoré do istej miery používali prirodzený jazyk (napr. ELIZA, SHRDLU) prinieslo so sebou otázky o povahe rozumenia u týchto systémov. Môžu umelo naprogramované systémy naozaj porozumieť jazyku, alebo iba manipulujú s (pre ne) nezmyselnými symbolmi? Tieto otázky vyvolali mnoho kontroverzií a emočne nabitých debát. Príkladmi pokusov o zodpovedanie, alebo aspoň reformuláciu, týchto otázok sú *Metafora čínskej izby* [67] a *Problém ukotvenia symbolov* [37].

John Searle tvrdí, že čisto na základe pozorovania správania (v tomto prípade komunikácie) nie je možné posúdiť, či systém naozaj rozumie tomu, o čom komunikuje [67]. Vymyslel myšlienkový experiment známy ako Čínska izba: predstavme si, že v izbe je zavretý človek, ktorý nevie po čínsky, a komunikuje s okolitým svetom pomocou lístkov. Ľudia zvonku mu posielajú otázky v čínštine. V izbe je škatuľa s čínskymi znakmi a kniha s pravidlami ako kombinovať znaky, ktorá umožňuje formulovať v čínštine odpovede na otázky. Človek v miestnosti skladá odpovede z čínskych znakov iba na základe formálnych pravidiel a porovnávaním tvaru znakov v knižke, na lístkoch a v škatuli. Predpokladajme, že sa v manipulácii so znakmi tak zdokonalí, že dokáže dávať na otázky korektné odpovede. Nikto mimo miestnosti sa nedozvie, že človek vo vnútri nerozumie ani slovo po čínsky.

Steven Harnad sformuloval svoju vlastnú verziu tohto problému, známu ako *problém ukotvenia symbolov*: „Ako možno urobiť sémantickú interpretáciu formálneho symbolového systému tak, aby bola systémom vlastná, a nie len parazitovala na významoch v našich hlavách? Ako môžu byť významy nezmyselných symbolových

Konstruktivistický prístup k štúdiu kognície

jednotiek, s ktorými sa manipuluje iba na základe ich (ľubovoľných) tvarov, ukotvené v niečom inom ako len v ďalších nezmyselných symboloch?” [37] (pozri obr. 2b).

Harnad navrhol aj vlastné riešenie tohto problému – hybridný subsymbolovo-symbolový model mysle, v ktorom symbolická funkcia emerguje zdola nahor ako dôsledok ukotvenia názvov kategórií v ich senzorických (subsymbolorých) reprezentáciách dvoch typov:

1. *ikonické* reprezentácie, ktoré sú analogické lokálnym priemetom (projekciám) vzdialených objektov a udalostí,
2. *kategoriálne* reprezentácie – naučené alebo vrodené detektory vlastností, ktoré zo senzorických projekcií vyberajú invariantné črty kategórií objektov a udalostí.

Vnútorne reprezentácie nemajú žiaden vlastný význam samy o sebe, ale získavajú ho štruktúrnym spriahnutím s prostredím [53]. Toto spriahnutie má dve zložky: individuálnu a sociálnu. Individuálna, ktorá sa nazýva aj *fyzickým ukotvením symbolov* [92], označuje schopnosť každého jedinca vytvoriť si vlastné prepojenie medzi entitami sveta a vnútornými reprezentáciami. Sociálna, nazývaná aj *sociálne* (alebo *externé*) *ukotvenie symbolov*, označuje kolektívne zjednávanie výberu zdieľaných symbolov a ich významov [12].

Významy sú teda individuálne vytvorené subjektívne konštrukty, ktoré sú kolektívne navzájom zladené. Fyzické ukotvenie symbolov po filozofickej stránke korešponduje s kognitívnym konštruktivizmom založeným na prácach Jeana Piageta [61] tým, že jedinci aktívne konštruujú svoje vlastné významy pomocou kognitívnych procesov na základe predchádzajúcej skúsenosti a interakcií s okolím. Sociálne ukotvenie symbolov je blízke sociálnemu konštruktivizmu [96] tým, že individuálne vytvorené významy sú motivované a formované sociálnym kontextom.

Vo zvyšku tejto časti urobíme stručný prehľad reprezentácie významov v niekoľkých umelých systémoch a výpočtových modeloch. Budeme analyzovať ich vyjadrovaciu silu a schopnosť vysporiadať sa s problémom ukotvenia symbolov.

4.2 Procedúry, pravidlá a predikátová logika

Začneme s niekoľkými systémami, ktoré sú považované za míľniky v ranej histórii umelej inteligencie. U prvého presláveného konverzačného programu ELIZA [98], ktorý simuloval nedirektívneho psychoterapeuta, je ťažko vôbec hovoriť o nejakých významoch. ELIZA nemala žiaden model vonkajšieho sveta, ani žiadnu sémantickú reprezentáciu prebiehajúcej konverzácie. Fungovala čisto reaktívnym spôsobom podľa niekoľkých šikovne naprogramovaných transformačných a dekompozičných pravidiel a nemala ani schopnosť učiť sa.

SHRDLU – systém pre porozumenie prirodzeného jazyka naprogramovaný Terryom Winogradom [100] umožňoval používateľovi komunikovať o jednoduchom simulovanom (vizualizovanom) „svete kociek“, pozostávajúcom z trojrozmerných útvarov rôznych farieb a veľkostí, v ktorom SHRDLU konal ako robot vybavený okom a magnetickou rukou. Užívateľ mohol dávať (textové) príkazy v prirodzenom jazyku, na základe ktorých SHRDLU manipuloval s objektami v simulovanom svete, alebo sa

Martin Takáč

mohol pýtať otázky o svete a o vykonaných akciách. Dôsledky akcií boli vizualizované na obrazovke. SHRDLU-ova znalosť o svete bola reprezentovaná z časti ako formuly predikátovej logiky a z časti ako procedúry operujúce na formulách. Atomické prvky reprezentácie boli symbolové a dané vopred. Obmedzená schopnosť SHRDLU učiť sa spočívala v pridávaní nových tvrdení do modelu sveta, na základe vstupu od užívateľa a predchádzajúcej histórie akcií v simulovanom svete. Keďže SHRDLU dokázal v simulovanom svete vykonávať akcie podľa pokynov v prirodzenom jazyku a dokázal správne odpovedať na otázky o stave sveta kociek, možno povedať, že jeho limitované významy boli ukotvené vo virtuálnom prostredí.

Shakey – prvý fyzicky realizovaný mobilný robot schopný usudzovať o svojom konaní bol vyvinutý v Stanfordskom výskumnom inštitúte (SRI) v rokoch 1966-72. Disponoval televíznou kamerou a triangulačným zariadením na určovanie vzdialeností, umiestnenými na pohyblivej hlave, nárazovými senzormi, dvomi krokovými motormi na ovládanie kolies a ďalšími motorčekmi na riadenie zaostrenia kamery a naklonenia hlavy [57]. Shakey sa dokázal pohybovať po rovných povrchoch medzi kancelárskymi a vykonával jednoduché úlohy, ktoré si vyžadovali plánovanie, určovanie trasy a prekladanie jednoduchých objektov. Udržoval si axiomatický model sveta a využíval hierarchiu programov na percepciu (vrátane jednoduchého videnia), modelovanie sveta a konanie. Využíval plánovací systém STRIPS [22], ktorý dokázal zreťaziť akcie do plánov a vykonať ich na dosiahnutie cieľa zadaného používateľom. Ciele boli reprezentované ako formuly prvorádovej predikátovej logiky. STRIPS mal aj istú schopnosť učiť sa – dokázal zovšeobecňovať plány a uchovávať ich pre budúce použitie. Môžeme povedať, že Shakey „rozumel“ svojmu svetu do tej miery, že disponoval reprezentáciami dôsledkov svojich akcií a využíval ich na plánovanie akcií za účelom dosiahnutia cieľov pomocou modifikovania svojho prostredia. Jeho znalosť, z väčšej časti predprogramovaná a reprezentovaná symbolmi prvorádových logických formúl, bola ukotvená v reálnom svete.

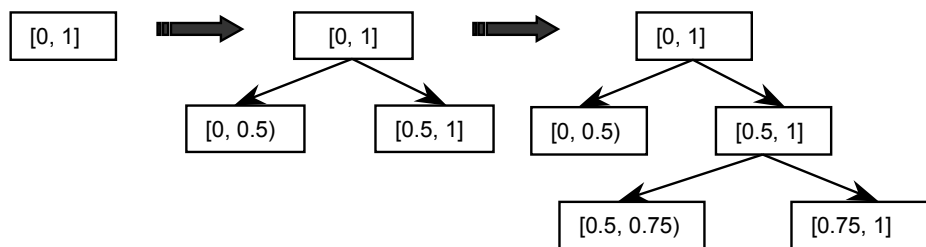
4.3 Diskriminačné stromy a prototypy

Teraz opíšeme niekoľko modelov, v ktorých sú významy reprezentované ako oblasti v nejakom geometrickom priestore obvykle definovanom možnými hodnotami senzorov, pomocou ktorých agenty vnímajú zdieľané prostredie: buď v jednorozmernom priestore pre každý senzor zvlášť, alebo vo viacrozmernom karteziánskom súčine oborov sensorických hodnôt. Významy v týchto modeloch nie sú dané vopred, ale každý agent si ich vytvára individuálne.

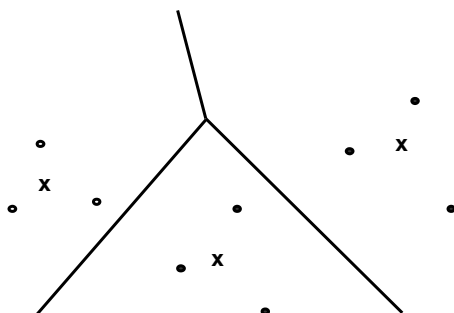
Senzorické priestory je možné rozdeliť na oblasti pomocou tzv. diskriminačných stromov [71], ktoré boli použité v mnohých modeloch emergencie jazyka, či už vo virtuálnych simuláciách [4, 69, 70] alebo v experimentoch s reálnymi robotmi [40, 72, 75, 76, 77]. Agenty konštruovali separátny diskriminačný strom pre každý sensorický kanál tak, že vrcholy diskriminačného stromu zodpovedali podintervalom oboru hodnôt príslušného senzora. Tým zároveň determinovali granularitu reprezentácie: všetky hodnoty senzora, ktoré padli do nejakého intervalu, boli reprezentované príslušným vrcholom (teda boli navzájom nerozlíšiteľné). Na začiatku každý strom pozostával iba z koreňa reprezentujúceho celý obor hodnôt daného senzora. Stromy sa

Konstruktivistický prístup k štúdiu kognície

postupne adaptívne zjemňovali rozdeľovaním intervalov vrcholov na dve polovice, t.j. vytvorením dvoch synov daného vrchola (obr. 3), v tzv. diskriminačných hrách [75] zameraných na nájdenie významu (vrchola), ktorý jednoznačne odlišuje zvolený objekt (tému komunikácie) od všetkých ostatných súčasne prítomných objektov (kontextu). Užitočnosť každého rozdelenia sa monitorovala zaznamenávaním úspešnosti použitia v nasledujúcich diskriminačných hrách; irelevantné (nepoužívané alebo neúspešné) rozlišovania sa zahadzovali. Týmto spôsobom diskriminačné úlohy poskytli konštruovaným významom ukotvenie.



Obr. 3. Rozdelenie diskriminačného stromu. Obor hodnôt senzora sa opakovane delí na polovice (situácie, teda výsledný strom nemusí byť vyvážený).



Obr. 4. Voronoiho mozaika priestoru kategórií generovaná prototypmi. Okrúhle body reprezentujú exempláre kategórií, krížiky (x) sú prototypy vypočítané ako ťažiská (stredy) príkladov.

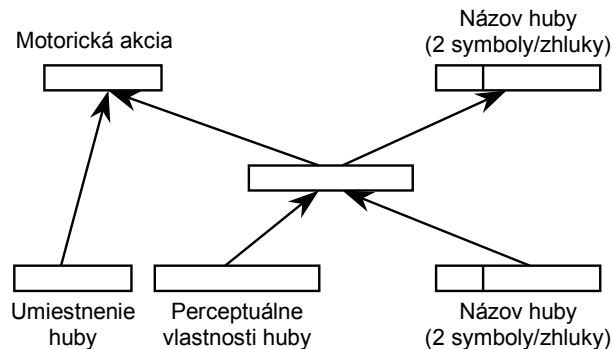
Alternatívou diskriminačným stromom je prototypová reprezentácia významov, ktorá má svoj pôvod v empirickom zistení, že niektoré exempláre kategórií sú reprezentatívnejšie ako iné [65]. Takáto reprezentácia kategórií je efektívna, pretože si stačí pamätať najlepšie príklady (prototypy). Každý bod sensorického priestoru sa potom považuje za člena kategórie reprezentovanej priestorovo najbližším prototypom (obr. 4). Prototypy boli použité na reprezentáciu významov v experimentoch so

Martin Takáč

softvérovými agentami v simulovanom prostredí [93, 95] ako aj u reálnych robotov [91, 92]. My si použitie diskriminačných stromov a prototypov predvedieme vo výpočtových modeloch horizontálnej koordinácie lexikónu v časti 6.2.

4.4 Dvojcestné neurónové siete

V konekcionistických modeloch ukotvenia symbolov sa často používa dvojcestná architektúra, ktorá obvykle obsahuje vizuálny vstup (napr. projekciu umelej sietnice) a jazykový vstup (napr. lokalistické alebo grafémové/fonémové kódovanie symbolov). Výstupná vrstva obsahuje symbolové jednotky na reprezentáciu slov a buď kategoriálnu reprezentáciu vstupných stimulov (napr. lokalistické výstupné jednotky pre každú kategóriu alebo vizuálnu reprezentáciu prototypov kategórií) alebo reprezentáciu želannej akcie (napr. hodnoty uhlov v kĺboch ramena). Všetky vstupné a výstupné vrstvy sú prepojené pomocou zdieľanej skrytej vrstvy. Cesta od vizuálneho vstupu k jazykovému výstupu sa používa na produkciu jazyka, napr. na pomenovanie (kategórie) objektu na vizuálnej scéne. Cesta od jazykového vstupu k vizuálnemu/kategoriálnemu/motorickému výstupu sa používa na porozumenie jazyka (s prípadnou akciou na základe jazykovej inštrukcie). Dve ďalšie cesty sa používajú na kategorizáciu resp. akciu na základe vnemu (cesta od vizuálneho vstupu ku kategoriálnym/motorickým jednotkám) a na jazykovú imitáciu (cesta od jazykového vstupu k jazykovému výstupu) [11]. Dvojcestné neurónové siete boli použité napr. v modeloch [10, 13, 54].



Obr. 5. Príklad dvojcestnej neurónovej siete použitej v experimente Cangelosiho [10].

V modeli [10] žije populácia 80 organizmov vo virtuálnom prostredí obsahujúcom tri druhy jedlých a tri druhy jedovatých húb. Model využíva genetickú evolúciu organizmov. Organizmy získavajú fitness vtedy, ak sa vyhýbajú jedovatým hubám a vedú identifikovať druh jedlej huby a zjedia ju. Správanie každého organizmu je riadené dvojcestnou neurónovou sieťou (obr. 5) ktorá umožňuje vykonávať akcie na základe miesta výskytu a perceptuálnych vlastností huby a/alebo jazykovej inštrukcie od iných agentov, a tiež pomenovávať huby (buď na základe

miesta výskytu a perceptuálnych vlastností alebo ako imitácia jazykového vstupu iného agenta). Na začiatku je lexikón úplne náhodný a nezmyselný. Ku koncu evolúcie sú však agenty schopné vyvinúť si zdieľaný kompozičný jazyk. Hoci prezentovaný model je zámerne zjednodušený, predstavuje korektný konekcionistický prístup k ukotveniu symbolov. Významy, implicitne reprezentované vo váhach spojení siete, sú súkromné a vytvorené každým agentom samostatne. Osvojovanie si jazyka je spriahnuté so senzomotorickými aktivitami cez spoločnú skrytú vrstvu dvojcestnej architektúry.

4.5 Korpusový prístup k významom

Hoci distribučné a kontextuálne charakteristiky slov nie sú postačujúce pre ukotvenie jazyka v reálnom svete (pozri Argument čínskej izby), predsa predstavujú dôležitú pomoc pre určenie významu slov [9, 52], čoho impresívnou demonštráciou sú konverzačné agenty (angl. chatter bots, chatbots) založené na kontextovom prehľadávaní korpusu. Chatbot je počítačový program, ktorý má simulovať inteligentnú konverzáciu s ľudskými užívateľmi. Mnoho chatbotov (ako ELIZA) je založených na rozpoznaní kľúčových slov v textovom vstupe a odpovedaní podľa predprogramovaných pravidiel. Chatbot Jabberwacky, ktorý naprogramoval Rolo Carpenter, je založený na odlišných princípoch: nie sú v ňom žiadne pevné predprogramované pravidlá a pracuje výhradne na základe interakcie s užívateľom [14]. Systém si uchováva obrovskú databázu všetkých predchádzajúcich rozhovorov a pokúša sa ju využiť na nájdenie najpriliehavejšej odpovede v aktuálnom kontexte. Prehľadávanie je založené na komplexnej viacvrstvovej sade heuristik, ktoré analyzujú konverzačný kontext a využívajú pozitívnu spätnú väzbu [39].

Úspešné budenie dojmu, že Jabberwacky naozaj rozmýšľa, spočíva vo veľkosti databázy. Jabberwacky bol on-line na webe od roku 1997 a zaznamenal viac ako 13 miliónov konverzácií. V komunikácii používa to, čo sa naučil od svojich konverzačných partnerov. Takto používa vtipy, idiómy, slovné hračky, slang a dokonca hovorí cudzími jazykmi. Dve nedávne inštancie Jabberwackyho – George a Joan vyhrali Loebnerovu cenu za rok 2005 a 2006. Loebnerova cena je každoročná súťaž, v ktorej cenu dostane ten chatbot, ktorého rozhodcovia považujú za všetkých zúčastnených za najbližšieho človeku. Formátom súťaže je štandardný Turingov test [90].

Hoci Jabberwackyho konverzácia môže byť ľuďmi interpretovaná ako zmysluplná – majúca nejaký sémantický obsah, tento obsah nie je vlastný a Jabberwacky o ňom nič nevie. Jazyková znalosť nie je ukotvená vo svete. Avšak z odlišného uhla pohľadu je Jabberwacky situovaný vo svete konverzačných postupností a v tomto svete sa naučí adekvátne reagovať v rôznych kontextoch. Znalosť o tom, čo je adekvátne je zakódovaná v zaznamenaných odpovediach Jabberwackyho ľudských partnerov. V súlade s Brooksom [6, 7, 8] by sme mohli povedať, že Jabberwackyho inteligentné správanie je emergentným efektom jeho interakcií, pričom znalosť je distribuovaná z časti v jeho architektúre a z časti v prostredí.

Martin Takáč

4.6 Ďalšie modely

Z priestorových dôvodov nemôže byť náš prehľad modelov vyčerpávajúci. Medzi ďalšie zaujímavé prístupy k reprezentácii významov patria semiotické schémy [66], dynamické mapy [18], prepisy súčasných udalostí [19] a stelesnená konštrukčná gramatika ECG [1]. Čitateľa odkazujeme na pôvodné zdroje.

5 Konštrukcia významov

5.1 Individuálne významy

Predstavili sme základné teoretické prístupy, evolučný kontext rozumenia a tiež niektoré problémy s rozumením v umelých systémoch. Z toho všetkého možno vyvodiť ponaučenie, že na prítomnosť rozumenia či významov v živočíšnej ríši a v umelých systémoch treba nazerať skôr ako na kontinuum od veľmi primitívnych foriem až po zložité a nie ako na niečo, čo buď je v plnej forme alebo nie je vôbec. Pripravili sme si tak pôdu pre vyslovenie našej vlastnej predstavy o rozumení v podľa možnosti čo najneutrálnejších, nie antropocentrických (a dokonca nie biologických) pojmoch.

O rozumení budeme hovoriť iba u agentov (živých či umelých) s ohľadom na (reálne či virtuálne) prostredie, v ktorom sú situované. *Situovaný agent* je autonómna entita, ktorá v prostredí dosahuje svoje ciele pomocou vnímania a konania [42]. Všimnime si, že táto definícia neimplikuje ani nevyžaduje vedomie či intencionalitu: agent môže dosahovať nejaké ciele (alebo plniť nejaký účel) bez toho, aby o tom „vedel“.

Adaptívne agenty modifikujú samy seba, aby lepšie dosahovali svoje ciele. Budeme rozlišovať medzi fylogenetickými (dizajn-time, pri návrhu) a ontogenetickými adaptáciami (run-time, počas behu). Učenie je ontogenetická adaptácia založená na spätnoväzobnej slučke medzi percepciou a akciou (alebo všeobecnejšie, medzi vstupom a výstupom). Vplyv adaptácií na správanie agenta možno pohodlne opísať pomocou teoretických konštruktov „reprezentácia“ a „semiotický znak“ bez ohľadu na ich realnosť, pôvod či implementáciu. Ak agent kategorizuje svet tým, že diferencovane reaguje na rôzne skupiny podnetov, hovoríme, že disponuje vyvolanými reprezentáciami, ktoré sú najelementárnejšou formou významov. Zložitejšie kognitívne schopnosti ako rozhodovanie a plánovanie si vyžadujú oddelené reprezentácie, ktoré môžu byť uchovávané, vyvolávané a spracovávané nezávisle od vonkajších environmentálnych spúšťačov.

5.2 Sociálne zdieľané významy

Doteraz sme hovorili o subjektívnych a súkromných významoch, ktoré záviseli od fylogenetickej a ontogenetickej histórie každého agenta. Kooperatívne plánovanie medzi agentami si vyžaduje externalizáciu a komunikáciu významov pomocou verejne vnímateľných signálov. Povedané v semiotických termínoch, každý agent interpretuje

signály ako znaky tým, že doplní dva zvyšné vrcholy a príslušné hrany semiotického trojuholníka (obr. 2d). Hovoríme, že agent pozorujúci signál ho chápe akoby mal sémantický obsah, ak sa správa v súlade s týmto obsahom [36]. Tento pohľad na rozumenie je interakcionistický: súkromné významy nemožno prenášať priamo; možno ich iba odvodiť z behaviorálnych interakcií. Komunikácia však môže byť úspešná, iba ak si budú súkromné významy dostatočne podobné. Na to sa dá spoľahnúť, ak sú asociácie medzi vrcholmi semiotického trojuholníka nearbitrárne a boli formované rovnakým mechanizmom, ako je to napr. v prípade vrodenných zvieracích signálnych systémov alebo u neadaptívnych umelých agentov riadených rovnakým predprogramovaným kódom. Avšak v *symbolických* komunikačných systémoch sú vzťahy medzi významami a signálmi arbitrárne z definície [16]. Ak boli významy formované alebo ovplyvnené individuálnou adaptáciou, vzájomná podobnosť významov u rôznych agentov nie je zaručená. Významy jednotlivých agentov je potrebné skoordinať a vzájomne zladit' adaptáciami v sociálnych interakciách. Tieto interakcie môžu byť jednosmerné (vertikálne), ak agent prispôsobí svoje vlastné reprezentácie existujúcemu komunikačnému systému bez toho, aby ho nejako ovplyvnil (napr. ak sa malé dieťa učí jazyk, ktorému je vystavené). Ak sú interakcie obojsmerné (horizontálne), t.j. vedú k adaptáciám na oboch stranách, celý jazyk sa správa ako dynamický systém, v ktorom „verejnú“ (zjednanú) významy emergujú ako (pohyblivé) ekvilibriá [28].

5.3 Princípy dizajnu rozumejúcich agentov

Z prezentovaného interakcionistického pohľadu na významy možno vyvodit' kritériá rozumenia u umelých agentov a princípy návrhu takýchto agentov:

- 1.) Významy musia pochádzať z interakcií agentov s prostredím, v ktorom sú situované. Prostredie môže byť reálne ale aj virtuálne – napr. softvérové agenty môžu vnímať prehľadávaním databáz a vykonávať akcie napr. uzatváraním elektronických obchodných transakcií.
- 2.) Netriviálne prostredia sú dynamické, menia sa v čase. Preto musia byť agenty schopné okrem statických objektov a vzťahov zachytiť a reprezentovať dynamické vlastnosti sveta ako sú zmeny, akcie a ich dôsledky, udalosti.
- 3.) Keďže prostredie je otvorené, nie je možné všetky významy anticipovať a predprogramovať vopred (v čase návrhu), preto je potrebná ontogenetická adaptácia (učenie). Učenie – konštrukcia významov – by malo byť inkrementálne a permanentné.
- 4.) Aby sa zabezpečilo vzájomné porozumenie v komunikácii, agenty musia byť vybavené mechanizmami na sociálnu koordináciu individuálne konštruovaných významov.

6 Výpočtové modely konštrukcie významov

V tejto časti predstavíme niekoľko výpočtových modelov ilustrujúcich horeuvedené princípy. Z didaktických dôvodov sme vybrali modely zamerané na jednotlivé princípy či konkrétny spôsob ich implementácie, aj keď nespĺňajú všetky princípy zároveň.

6.1 Individuálna senzomotorická explorácia

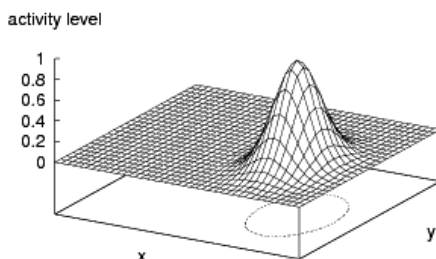
Ako prvý predstavíme výpočtový model individuálneho konštruovania významov ukotvených v senzomotorických interakciách [81]. Model pozostáva z jediného agenta situovaného v simulovanom prostredí – na dvojrozmernej mriežke s náhodne rozmiestnenými objektami štyroch typov: ovocie, hračky, nábytok a sám agent. Na začiatku agent nerozlišuje žiadne typy, len dostáva informáciu o vnímateľných vlastnostiach objektov okolo seba vo forme rámcov – množín dvojíc *⟨atribút: numerická hodnota⟩*. Atribúty kódujú vlastnosti ako pozícia na mriežke, veľkosť, hmotnosť, farba, tvar, atď. Čas je v modeli diskretný. V každom časovom kroku agent dostáva na vstupe jeden rámec pre každý vnímaný objekt. Agent si náhodne vyberá jeden z objektov a pokúša sa ním manipulovať – zdvihnúť ho alebo položiť. Výsledkom úspešných manipulácií sú zmeny hodnôt atribútov manipulovaného objektu (v prípade dvíhania a kladenia vertikálna poloha). Akcie sú parametrizované – agent si volí silu vynaloženú na dvíhanie a vertikálnu polohu vlastného ramena (tieto parametre sú zjednodušenou abstrakciou simulovanej motorickej akcie – v skutočnosti agent nemá žiadne rameno, ani motory, atď.). Maximálne hodnoty sily a polohy sú obmedzené „konštrukciou“ agenta.

Agent pozoruje dôsledky svojich akcií porovnaním hodnôt atribútov objektu v dvoch po sebe nasledujúcich časových krokoch. Skutočné dôsledky akcií sú vyhodnocované simulátorom prostredia podľa zjednodušených „fyzikálnych zákonov“, napr. ak je sila vynaložená na dvíhanie príliš malá v porovnaní s hmotnosťou dvíhaného objektu, objekt sa nepohne, v opačnom prípade sa dvihne priamo úmerne zvolenej vertikálnej polohe agentovho ramena.

Agent aktívne exploruje svoje prostredie vykonávaním náhodných akcií a postupne sa učí rozlišovať relevantné vlastnosti prostredia a buduje si kategórie objektov, akcií a zmien. Kategórie sú reprezentované tzv. *rozlišovacími kritériami* inšpirovanými už spomínanou kognitívnou sémantikou Šefránka [79]. Každé rozlišovacie kritérium je aktivačnou funkciou, ktorá pre svoj vstup vracia mieru príslušnosti vstupu v reprezentovanej kategórii (obr. 6). Medzi možné vstupy patrí perceptuálny rámec jedného objektu (u kritérií objektov a ich vlastností), „proprioceptívny“ rámec parametrov a typu akcie (u kritérií akcií), perceptuálne rámce viacerých objektov (u kritérií vzťahov), rámce toho istého objektu v rôznych okamihoch (u kritérií zmien) a výstupné aktivity iných kritérií (u kompozičných kritérií situácií a udalostí). Na začiatku agent nemá žiadne kritériá, tie si postupne konštruuje extrakciou spoločných štatistických vlastností exemplárov kategórií. Technické detaily čitateľ nájde v [82, 84]. Ak nejaká akcia vykonaná na rôznych objektoch vedie k dostatočne podobným dôsledkom (zmenám atribútov), tieto objekty budú považované za členov tej istej kategórie (vzhľadom na danú akciu a dôsledok).

Konstruktivistický prístup k štúdiu kognície

V prípade príliš odlišných dôsledkov sa vytvoria nové kategórie. Rovnaký princíp sa používa aj na združovanie rôznych akcií vykonaných na tom istom objekte.



Obr. 6. Rozlišovacie kritérium s Gaussovskou krivkou aktivácie nad dvojrozmerným vstupným priestorom.

Propozíčné znalosti o kauzálnych vzťahoch medzi akciami, objektami a zmenami sú reprezentované pomocou asociácií medzi príslušnými kategóriami (ako analógia asociačných oblastí v mozgovej kôre). Takáto reprezentácia je v súlade s poznatkami o ľudskej organizácii znalostí. Podľa Gibsona dokážeme z prostredia priamo vnímať možnosti konania – tzv. *afordancie* [32]. Afordancie zahŕňajú spôsoby manipulovania s vecami, napr. dvere, ktoré sa dajú otvoriť, potrava, ktorá sa dá zjesť, povrchy, po ktorých sa dá kráčať, atď. V našom modeli sú afordancie objektu reprezentované ako všetky s ním asociované akčné kategórie. Objekty a zmeny asociované s akčnými kategóriami zase tvoria zárodok sémantickej reprezentácie centrovanej okolo slovíec, tzv. *slovesné ostrovy*, ktoré v organizácii detského lexikónu opísal Tomasello [88].

V počítačovej simulácii nášho modelu sme nechali agenta interagovať s prostredím počas 5000 krokov a merali sme užitočnosť vytvorenej reprezentácie ako schopnosť predikovať dôsledky vlastných akcií. Experimentálne výsledky potvrdili, že navrhnutý mechanizmus vytvárania kategórií na základe akcií vedie k ekologicky relevantným kategóriám. Technické detaily algoritmu a výsledky možno nájsť v [81].

Kvalitatívna analýza skonštruovaných významov vo vybranom simulačnom behu ukázala, že agent si vytvoril štyri objektové kategórie. Dve z nich by bolo možné vzhľadom na asociovanú kauzálnu znalosť interpretovať ako „veci príliš ťažké na to, aby sa dali zdvihnúť“ a „veci, ktoré sa nedajú položiť, pretože už sú na zemi“. Do tretej kategórie padla väčšina objektov typu ovocie a do štvrtej väčšina hračiek. V kritériách ostali prítomné aj iné atribúty ako hmotnosť a vertikálna poloha, čo znamená, že nejakým spôsobom pomáhali agentovi v klasifikácii (napr. ak by sa všetky ťažké objekty nachádzali v rovnakej časti mriežky alebo by mali rovnakú farbu). Konštruovaná reprezentácia bola teda situovaná a obsahovala aj subjektívny interakčný kontext agenta.

Tento experiment modeluje aktivity dieťaťa v senzomotorickom štádiu detského vývinu [62], v ktorom dieťa pomocou náhodných a spočiatku nekoordinovaných pohybov spoznáva objekty okolo seba, vlastné telo, a možnosti ich vzájomného interagovania. Opakujúce sa kauzálne vzorce sú zaznamenané, čo neskôr umožní

Martin Takáč

zámerné plánovanie akcií na splnenie vytýčených cieľov. Významným aspektom prezentovaného modelu je to, že agent je v čase návrhu vybavený iba učiacimi mechanizmami, a všetky znalosti o konkrétnom prostredí získa až počas behu. To umožňuje agentovi byť flexibilný a adaptovať sa na rôzne prostredia, ktoré nemuseli byť v čase návrhu známe, alebo sa mohli zmeniť počas agentovho „života“ (predstavme si analogickú úlohu naprogramovať robota, ktorý bude vyslaný na vzdialenú planétu s neznámymi podmienkami).

Z vytýčených princípov dizajnu rozumejúcich agentov tento model spĺňa princípy 1 až 3. Percepčno-akčné rozhranie modelu bolo (zámerne) značne abstraktné a zjednodušené. V stelesnených systémoch situovaných v reálnych prostrediach patria úlohy spojené s percepciou a spracovaním zašumeného analógového senzorického signálu a riadením efektorov (napr. servomotorov mechanického ramena) medzi tie ťažšie. Keďže model pozostával iba z jediného agenta, nebola v ňom žiadna sociálna koordinácia významov ani jazyk (princíp 4). Týmto aspektom sú venované nasledujúce časti.

6.2 Horizontálna koordinácia významov

V tejto časti predstavíme modelársky prístup skupiny Luca Steelsa ilustrujúci emergenciu koordinovaného komunikačného systému na základe horizontálnych kultúrnych interakcií [73]. Horizontálne interakcie sú obojsmerné, čo znamená, že každá z komunikujúcich strán sa snaží prispôsobiť tej druhej. Ak by sme si opäť predstavili semiotický trojuholník (obr. 1 a 2d), adaptácia prebieha na úrovni hrán aj vrcholov: agent mení svoje súkromné asociácie medzi externe pozorovateľnými signálmi, internými významami a ich príslušnými referentami vo svete, a tiež mení aj samotné interné významy.

Adaptácia je riadená spätnou väzbou o úspešnosti resp. zlyhaní komunikácie. Úspech v komunikácii nemusí nutne znamenať, že agenti majú s rovnakým signálom asociované identické významy: keďže interné významy nie sú viditeľné a nie je ich možné priamo porovnávať, kritériá úspešnosti sú pragmatické, teda znaky sa musia zhodnúť na úrovni referentov, čo je manifestované správaním agentov. Behaviorálne kritériá sú plauzibilné: ak napr. požiadam niekoho, aby mi podal pero, a on mi namiesto toho podá jablko, alebo otvorí dvere, môžem odôvodnene predpokladať, že ma neporozumel (príp. žartuje, je zlomyseľný, atď.).

V modeloch založených na tomto prístupe agenti (simulované alebo stelesnené v reálnych robotoch) hrajú rôzne typy tzv. *jazykových hier* s rôznymi pragmatickými cieľmi [75]. Hry sa hrajú v kolách: v každom kole sa z populácie náhodne vyberú dva agenti – jeden v role hovorcu a druhý v role poslucháča. Najprv zohrá hovorca tzv. *diskriminačnú hru*, v ktorej sa snaží jednoznačne identifikovať zvolený objekt na scéne (tému komunikácie) pomocou významov a tie následne lexikalizovať do signálov. Na základe týchto signálov sa potom poslucháč pokúša uhádnuť tému komunikácie v tzv. *hádacej hre*.

Agenty si ukladajú konceptuálne znalosti (významy) v podobe diskriminačných stromov, a jazykovú znalosť v podobe viacnásobných asociácií medzi vrcholmi stromov a signálmi. Stromy aj asociácie sú konštruované a adaptované individuálne na

Konstruktivistický prístup k štúdiu kognície

základe úspešnosti v jazykových hrách (asociácie sa adaptujú tak, že sa mení ich číselne vyjadrená sila).

V prípade úspešnej diskriminačnej hry sa hovorca snaží lexikalizovať vybrané významy použitím tzv. stratégie *introspektívneho obvertera* [69]: zo všetkých signálov asociovaných s daným významom vyberie ten, ktorý by on sám s najväčšou pravdepodobnosťou interpretoval ako zvolený význam. Poslucháč sa potom pokúša interpretovať signál pomocou vlastných interne asociovaných významov a hľadá medzi objektami na scéne zodpovedajúci referent. Na konci poslucháč dostane (neverbálnu) spätnú väzbu o tom, či jeho tip na referenta bol správny a obaja si príslušným spôsobom adaptujú silu zúčastnených asociácií. V inom variante nazvanom *pozorovacia hra* [58] poslucháč nedostáva žiadnu spätnú väzbu. Namiesto toho sa množina možných referentov obmedzí sledovaním pohľadu (alebo ukazovania) hovorca a poslucháč si upravuje asociácie vo svojom lexikóne Hebbovským učením (ak je viac možných kandidátov, vytvorí/posilnia sa asociácie s každým z nich). Nejednoznačnosti sa časom redukujú učením z viacerých situácií [68, 70].

Známou materiálnou realizáciou hádacích hier bol experiment *Hovoriace hlavy* – *Talking Heads* [72]. Na scéne boli inštalované dve kamery, do ktorých sa dali uploadovať viaceré agenty. Agenty pomocou kamier vnímali zdieľané prostredie – magnetickú tabuľu s geometrickými tvarmi rôznych farieb a veľkostí a komunikovali o nich.

Ešte sa zmienime o modeli hádacej hry, ktorý realizoval Vogt [92]. V tomto modeli boli agenty stelesnené v dvoch mobilných LOGO robotoch. Agenty používali prototypovú reprezentáciu významov. Cieľom jazykovej hry bolo komunikovať názov jedného zo svetelných zdrojov, ktoré roboty dokázali detekovať v prostredí.

V oboch zmienených modeloch sa agentom podarilo vytvoriť koordinovaný lexikón ukotvený v reálnych podmienkach. Význam týchto modelov je v tom, že demonštrujú, ako môže globálne koherentný komunikačný systém vzniknúť „z ničoho“ ako výsledok lokálnych interakcií jeho používateľov. Agenty začínajú bez významov (len s nediferencovanými senzorickejšími kanálmi) a s prázdny lexikónom – celý systém sa naštartuje vďaka mechanizmu invencie: v prípade, že agent nemá v lexikóne slovo na vyjadrenie nejakého významu, vytvorí si a asociuje nové náhodné slovo. Samozrejme, takémuto signálu spočiatku poslucháč neporozumie; pri adaptácii si však zapamätá asociáciu medzi signálom a významom a v budúcnosti ho už môže úspešne použiť. Týmto spôsobom sa signály šíria v populácii. Tendencia agentov používať tie signály a významy, ktoré sa v minulosti osvedčili, vytvára pozitívnu spätnoväzobnú slučku. Viaceré alternatívne signály a významy medzi sebou súťažia a víťazné sa posilňujú. To vedie k samoorganizácii a k emergencii koherentného komunikačného systému. Spontánna samoorganizácia jazyka v rámci jednej generácie a na základe čisto kultúrnych horizontálnych interakcií bola pozorovaná aj medzi ľuďmi – hluchonemými deťmi v prípade Nikaragujskej posunkovej reči [41].

Ak sa poohliadneme za vytýčenými princípmi dizajnu rozumejúci agentov, tu prezentované modely spĺňajú kritériá 1, 3 a 4. Každý agent si individuálne vytvára a permanentne adaptuje významy na základe interakcií s prostredím. Významy sú interné – nedostupné iným agentom a potenciálne rôzne u rôznych agentov. Sú však navzájom zladené, čo umožňuje agentom dosiahnuť ciele v jazykových hrách.

Martin Takáč

Kritérium 2 je však porušené: agenti dokážu komunikovať iba o statických objektoch aktuálne prítomných v ich dohľade. Nemajú žiadne prostriedky ani na reprezentáciu, ani na komunikáciu dynamických aspektov prostredia. Komunikácia je založená iba na vyvolaných reprezentáciách.

6.3 Učenie na základe (vertikálnej) sociálnej inštrukcie

V tejto časti sa zameriame na jednosmerné interakcie medzi agentami a ich vplyv na proces adaptácie významov. Jednosmerný adaptačný proces je učením, preto budeme agenta, ktorý sa neadaptuje, nazývať *učiteľ* a agenta, ktorý sa adaptuje, *učiaci sa*. Tento typ učenia hrá významnú úlohu v procese osvojovania si jazyka u detí. Predstavíme výpočtový model inšpirovaný empirickým výskumom Sandry Waxman, ktorá skúmala vplyv verbálnej inštrukcie (pomenovania) na proces formovania kategórií [97]. Výsledky jej výskumu napovedajú, že konzistentné používanie rovnakého názvu pre rôzne objekty motivuje deti k hľadaniu podobností medzi nimi a k formovaniu kategórií.

Výpočtový model [85] pozostáva z dvoch agentov situovaných v simulovanom prostredí: učiteľ opisuje rôzne aspekty meniacej sa scény a učiaci sa agent si vytvára významy indukciou podobností medzi referentami vo viacerých situáciách. Simulované prostredie obsahuje plošné geometrické tvary charakterizované náhodne generovanými numerickými atribútmi ako počet vrcholov, poloha (súradnice stredu objektu) a veľkosť (opísaného obdĺžnika). Prostredie je dynamické – náhodne zvolené objekty môžu zmeniť veľkosť, polohu, alebo môžu byť odobraté z prostredia, tiež môžu byť pridané nové objekty (v jednom časovom kroku sa môže udiť aj viac zmien naraz).

Učiaci sa agent si vytvára reprezentáciu významov v podobe už zmienených rozlišovacích kritérií. Na začiatku nemá žiadne – tie si vytvára indukciou zo spoločných štatistických vlastností prezentovaných príkladov. Učiteľ v priebehu učenia pomenúva rôzne aspekty objektov, ich vlastností, vzťahov a dynamických zmien a učiaci sa agent musí nejako rozhodnúť, ku ktorým významom tieto pomenovania priradiť, kedy aktualizovať existujúce kritériá a kedy vytvárať nové. Indukcia významov sa riadi dvoma princípmi: predpoklad *žiadnej pravej synonymie* a *žiadnej pravej homonymie*. Aj keď prirodzené jazyky obsahujú slová s viacerými významami (homonymá) a aj viaceré výrazy pre ten istý význam (synonymá), v prípade tvorenia jazyka a významov z ničoho je užitočné začať s takýmito predpokladmi.

- *Žiadna pravá synonymia*: Rôzne slová majú rôzny význam, aj keď zdieľajú referent (v tom prípade označujú rozličné aspekty referenta). Tento predpoklad zodpovedá *Princípu kontrastu* [17], ktorý je známy z akvizície jazyka deťmi.
- *Žiadna pravá homonymia*: Jedno slovo má len jeden význam, aj keď označuje viacero referentov. Tento predpoklad je kľúčový pre riešenie nejednoznačnosti označenia učením sa z viacerých situácií: referenty toho istého slova tvoria vzorku – množinu inštancií kategórie označenej daným slovom.

Príklad: Majme agenta, ktorý je situovaný vo svete geometrických tvarov na mriežke 50×50 pričom bod so súradnicami (1, 1) je najviac vľavo dolu a bod so súradnicami

(50, 50) najviac vpravo hore. Ak agent vníma objekt daný rámcom $f = \{ \text{počet_vrcholov: } 3; \text{ veľkosť: } 18; \text{ farba: } 3; \text{ pozícia_X: } 1; \text{ pozícia_Y: } 23 \}$ pomenovaný slovami „vľavo“, „veľký“, „trojuholník“, vytvorí si tri rozlišovacie kritériá, ktoré sú na začiatku úplne identické a reprezentujú „momentku“ objektu daný rámcom f . Ako sú kritériá postupne aktualizované viacerými inštanciami, začínajú sa diferencovať. Napríklad rámce rôznych objektov, ktoré majú najrôznejšie farby, polohy, veľkosti a ďalšie vlastnosti, ale všetky majú 3 vrcholy, budú aktualizovať rozlišovacie kritérium asociované so slovom „trojuholník“. Atribúty, ktoré sa nevyskytujú vo všetkých inštanciách, sa z kritéria odstránia, a tie, ktoré mali vo vzorke veľký rozptyl, budú mať malú dôležitosť. Tým sa vlastnosť mať tri vrcholy (ktorá je vo vzorke invariantná – má nulový rozptyl) stane rozhodujúcou. Podobne slovo „vľavo“ bude označovať najrôznejšie objekty, ktorých spoločnou vlastnosťou bude malá hodnota atribútu *pozícia_X*, atď. Čím viac kontextov použitia slova bude k dispozícii, tým je väčšia pravdepodobnosť, že irelevantné vlastnosti budú mať vo vzorke veľkú variabilitu. Avšak keby napríklad všetky trojuholníky, ktoré agent doteraz vnímal, boli veľké, vlastnosť byť veľký by sa stala súčasťou významu slova „trojuholník“. Takto vytvorená reprezentácia je preto vždy situovaná a kontextovo závislá.

Simulácia opísaného experimentu bežala 5000 krokov. V každom kroku učiteľ vybavený predprogramovanou ontológiou a lexikónom opísal aktuálnu scénu vrátane dynamických zmien. Učiteľov lexikón obsahoval 2 podstatné mená, 3 adjektíva a slovesá (štvorec, trojuholník, veľký, štíhly, malý, narásť, zmenšiť sa). Úlohou učiteľa bolo naštruktúrovať vzorku pre učiaceho sa agenta – rovnako dobre by mohol poslúžiť človek, ktorý by sledoval scénu a ručne dodával opisy toho, čo vníma. To by však bolo časovo neefektívne, preto bol pre urýchlenie experimentu použitý automatický učiteľ. Učiteľ zároveň ukázal na referent pomenovávanej kategórie, t.j. na objekt s danou vlastnosťou, zmenou alebo reláciou.

Toto zjednodušenie bolo použité zámerne, aby sme ukázali, že aj v neprítomnosti nejednoznačnosti *referenta* musí učiaci sa agent riešiť nejednoznačnosť *zmyslu*, pretože rôzne slová môžu opisovať rôzne aspekty toho istého (známeho) referenta. Povedané v termínoch semiotického trojuholníka, úlohou učiaceho sa agenta je na základe dvoch pozorovateľných vrcholov (signálu a referenta) skonštruovať alebo adaptovať tretí vrchol – svoj interný význam/zmysel a príslušné asociácie.

Za účelom merania kvality vytvorených významov učiaci sa agent tiež opisoval vnímanú scénu a opis sa porovnával s učiteľovým. Zaznamenával sa počet *chýbajúcich* slov (takých, ktoré na opis daného referenta použil učiteľ, ale nie učiaci sa) a *nesprávnych* slov (ktoré použil žiak, ale nie učiteľ). Okrem toho učiaci sa agent hral klasické hádacie hry a zaznamenávala sa úspešnosť v hádaní referenta opisu použitého učiteľom. Výsledky ukázali, že učenie sa z viacerých situácií je rýchle a spoľahlivé: učiaci sa agent si rýchlo vytvoril ontológiu a lexikón dostatočnej kvality, ktorý ostal stabilný po celý zvyšok simulácie. Detaily čitateľ nájde v [84, 85].

Významy skonštruované v tomto modeli neodrážajú priamo žiadne interakčné vlastnosti (afordancie) prostredia získané agentovou exploračiou. Namiesto toho sú determinované vopred existujúcou štruktúrou učiteľových kategórií sprostredkovanou pomocou jazyka. Experiment demonštruje ako možno kultúrne prenášať existujúci konceptuálny systém pomocou vertikálnych jazykových interakcií.

Martin Takáč

Jazyky však nie sú prenášané ako zakonzervované systémy, ale sú živé a v priebehu generácií podliehajú zmenám a vyvíjajú sa [20], napr. sa objavujú novotvary a miznú archaizmy, zjednodušuje a spravidelňuje sa syntax, atď. Emergentná štruktúra evolvujúceho sa jazyka je tvarovaná ontogenetickým procesom akvizície jazyka, ktorý je zase determinovaný (poväčšine vrodennými resp. predprogramovanými) mechanizmami učenia [5]. Medzigeneračné významové posuny v rámci iterovaného vertikálneho prenosu sme skúmali *metódou iterovaného učenia* [43] v experimente [83].

7 Záver

V tejto kapitole sme sa snažili zdôrazniť dôležitosť konštruktivistického prístupu k významom pre štúdium kognície. Rozpracovali sme pojmový aparát pre (inherentné) významy, ukázali sme, ako môžu umelé agenty individuálne konštruovať významy na základe interakcií s prostredím a ako je možné tieto významy kolektívne koordinovať. To je veľmi dôležité najmä v neznámych, dynamických a otvorených prostrediach, kde nie je možné anticipovať všetky možné významy v čase návrhu systému.

Čo sa týka komunikácie, poukázali sme na to, že identickosť interných významov nie je pre vzájomné rozumenie nutná; postačuje koordinácia na úrovni správania, čo vytvára dobrý základ pre našu komunikáciu (a koexistenciu) s umelo vytvorenými agentami. Inherentné významy odlišne skonštruovaných agentov s odlišnými cieľmi a ekologickými nikami budú odlišné od našich ľudských. Napriek tomu môžeme navzájom viesť zmysluplnú komunikáciu, ak si zjednáme spoločný komunikačný systém.

7.1 Možnosti ďalšieho výskumu

Na viacerých výpočtových modeloch sme ilustrovali individuálnu konštrukciu a kolektívnu koordináciu významov. Tieto modely boli v mnohých aspektoch zjednodušené, na demonštráciu realizovateľnosti navrhnutých ideí však postačujú. Z metodologických dôvodov sme skúmali opísané fenomény každý zvlášť: v modeli individuálnej konštrukcie významov chýbala komunikácia, v modeli koordinácie významov neboli žiadne mimojazykové aktivity, atď. Prirodzeným (a realistickejším) pokračovaním by mal byť model, ktorý integruje všetky spomínané aspekty.

V modeli senzomotorickej explorácie agent nemal žiadne ciele a vykonával akcie náhodne. Keďže si úspešne dokázal osvojiť a reprezentovať propozičné znalosti o dôsledkoch svojich akcií v prostredí, ďalším logickým krokom je rozšírenie modelu o potreby, potrebami generované ciele a mechanizmus plánovania akcií.

V modeli integrujúcom plánovanie, komunikačnú úroveň a mechanizmy horizontálnej a vertikálnej koordinácie bude možné experimentálne testovať, či autonómne agenty profitujú (efektívnejšie dosahujú ciele) z vzájomnej výmeny znalostí pomocou zdieľaného komunikačného systému.

Ďalší výskum by sa mal tiež zaoberať analýzou vyjadrovacej sily rôznych typov reprezentácie významov. Reprezentácia by mala zachytávať hierarchické

a taxonomické vzťahy medzi významami a podporovať sémantiku slovies. V reprezentácii pomocou diskriminačných stromov dynamické aspekty prostredia neboli zachytené vôbec. V sémantike rozlišovacích kritérií boli významy slovies konštruované z kritérií jednokrokových zmien hodnôt atribútov. Pre niektoré slovesá môžu však byť potrebné dlhšie sekvencie, ešte iné možno vôbec nevystačia s diskrétnymi postupnosťami a budú potrebovať nejaký typ spojitej reprezentácie dynamiky (napr. fázové portréty).

Jazyk bol vo všetkých prezentovaných modeloch viazaný na aktuálnu situáciu (vyvolaný). Primárnou evolučnou výhodou ľudského jazyka je však jeho oddelené používanie, t.j. komunikácia o veciach, ktoré presahujú „tu a teraz“. Oddelená komunikácia by bola príliš nejednoznačná bez použitia nejakej formy gramatiky [27]. Zakomponovanie gramatiky do modelov emergencie a akvizície jazyka predstavuje dôležitý smer budúceho výskumu. Prvé kroky v tomto smere už boli vykonané [74, 78].

Demonštrovali sme, ako sa dajú skonštruovať kategoriálne významy extrakciou spoločných štatistických vlastností z príkladov. Pre konštrukciu komplexnejších významov by bolo vhodné obohatiť model o analytické a syntetické operácie a nemonotónne usudzovanie [79].

Doteraz nespomenutým problémom je škálovateľnosť modelov. Tu by sme chceli upozorniť na ambiciózny projekt New Ties (angl. skratka pre „nové a emergentné modely sveta pomocou individuálneho, evolučného a sociálneho učenia“, <http://www.new-ties.org>). Cieľom projektu je vyvinúť veľké umelé spoločenstvo (rádovo tisíce agentov) schopné explorácie a chápania svojho prostredia pomocou kooperácie a interakcie [33, 94]. Rovnaké mechanizmy koordinácie jazyka a významov tak otvárajú dvere pre skúmanie zmiešaných ľudsko-strojových spoločenstiev.

Podakovanie: Táto kapitola vznikla za podpory grantovej agentúry VEGA v rámci grantovej úlohy 1/0361/08.

Literatúra

- [1] Bergen, B., Chang, N.: Embodied construction grammar in simulation-based language understanding. In Ostman, J. O., Fried, M. (Eds.): *Construction grammar(s): Cognitive and cross-language dimensions*. Amsterdam, Netherlands: Johns Benjamins, 2003, pp. 147–190.
- [2] Bloom, P.: *How children learn the meanings of words*. Cambridge, MA: MIT Press 2000.
- [3] Blum, M., Williams, R., Juba, B., Humphrey, M.: *Toward a high-level definition of consciousness*. Invited talk presented at the Annual IEEE Computational Complexity Conference, San Jose, CA, June 2005.
- [4] Bodík, P., Takáč, M.: Formation of a common spatial lexicon and its change in a community of moving agents. In B. Tessem, P. Ala-Siuru, P. Doherty, B. Mayoh

Martin Takáč

- (Eds.): *Frontiers in AI: Proceedings of the Eighth Scandinavian Conference on Artificial Intelligence SCAI'03*. Amsterdam: IOS Press, 2003, pp. 37–46.
- [5] Briscoe, T. (Ed.): *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, U. K.: Cambridge University Press 2001.
- [6] Brooks, R. A.: Elephants don't play chess. *Robotics and Autonomous Systems* **6** (1–2) (1990) 3–15.
- [7] Brooks, R. A.: Intelligence without reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann, 1991, pp. 569–595.
- [8] Brooks, R. A.: Intelligence without representation. *Artificial Intelligence* **47** (1–3) (1991) 139–159.
- [9] Bullinaria, J. A., Levy, J. P.: Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods* **39** (2007) 510–526.
- [10] Cangelosi, A.: Modeling the evolution of communication: From stimulus associations to grounded symbolic associations. In D. Floreano, J. Nicoud, F. Mondada (Eds.): *Proceedings of the 5th European Conference on Advances in Artificial Life*. Berlin, Germany: Springer, 1999, pp. 654–663.
- [11] Cangelosi, A.: Approaches to grounding symbols in perceptual and sensorimotor categories. In H. Cohen, C. Lefebvre (Eds.): *Handbook of categorization in cognitive science*. Amsterdam, Netherlands: Elsevier, 2005, pp. 719–737.
- [12] Cangelosi, A.: The grounding and sharing of symbols. *Pragmatics and Cognition* **14** (2) (2006) 275–285.
- [13] Cangelosi, A., Parisi, D.: How nouns and verbs differentially affect the behavior of artificial organisms. In J. D. Moore, K. Stenning (Eds.): *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*. London, U. K.: Lawrence Erlbaum Associates, 2001, pp. 170–175.
- [14] Carpenter, R.: *Jabberwacky – live chatbot*. Retrieved December 8, 2007, from <http://www.jabberwacky.com>
- [15] Carnap, R.: *Meaning and necessity: a study in semantics and modal logic*. Chicago, IL: University of Chicago Press 1947/1956.
- [16] Chandler, D.: *Semiotics: the basics* (Second ed.). London, New York: Routledge.
- [17] Clark, E.: The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.): *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1987, pp. 1–33
- [18] Cohen, P.: Dynamic maps as representations of verbs. In *Proceedings of the 13th Biennial European Conference on Artificial Intelligence*. New York, NY: John Wiley Sons, 1998, pp. 145–149.
- [19] Cohen, P., Oates, T., Atkin, M., Beal, C.: Building a baby. In G. W. Cottrell (Ed.): *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, pp. 518–522.

Konstruktivistický prístup k štúdiu kognície

- [20] Deacon, T. W.: *The symbolic species: The co-evolution of language and the brain*. New York, NY: W.W. Norton Co 1997.
- [21] Fauconnier, G.: *Mental spaces: Aspects of meaning construction in natural language*. Cambridge, MA: MIT Press 1985.
- [22] Fikes, R., Nilsson, N.: *STRIPS: A new approach to the application of theorem proving to problem solving* (Tech. Rep. No. 43r). Menlo Park, CA: AI Center, SRI International 1971.
- [23] Fillmore, C. J.: Frame semantics. In *Linguistics in the morning calm* (pp. 111–137). Seoul: Hanshin Pub. Co. 1982.
- [24] Fodor, J. A.: *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press 1981.
- [25] Fodor, J. A.: *The modularity of mind*. Bradford Books. Cambridge, MA: MIT Press 1983.
- [26] Gärdenfors, P.: Cued and detached representations in animal cognition. *Behavioral Processes* **35** (1996) 263–273.
- [27] Gärdenfors, P.: Language and the evolution of cognition. In V. Rialle, D. Fisette (Eds.): *Penser l'esprit: Des sciences de la cognition a' une philosophie cognitive*. Grenoble: Presses Universitaires de Grenoble, 1996, pp. 151–172.
- [28] Gärdenfors, P.: *Conceptual spaces*. Cambridge, MA: MIT Press 2000.
- [29] Gärdenfors, P.: Cooperation and the evolution of symbolic communication. In K. Oller, U. Griebel (Eds.): *The evolution of communication systems*. Cambridge, MA: MIT Press, 2004, pp. 237–256.
- [30] Gelder, T. J.: Dynamic approaches to cognition. In R. A. Wilson, F. C. Keil (Eds.): *The MIT encyclopedia of the cognitive sciences*. Cambridge, MA: MIT Press, 1999, pp. 244–246.
- [31] Gibbs, R. W.: *Embodiment and cognitive science*. Cambridge, U. K.: Cambridge University Press 2006.
- [32] Gibson, J. J.: *The ecological approach to visual perception*. Boston: Houghton Mifflin 1979.
- [33] Gilbert, N., Besten, M. den, Bontovics, A., Craenen, B. G. W., Divina, F., Eiben, A. E., Griffioen, R., Hévízi, G., Lörincz, A., Paechter, B., Schuster, S., Schut, M. C., Tzolov, C., Vogt, P., Yang, L.: Emerging Artificial Societies Through Learning [Electronic version]. *Journal of Artificial Societies and Social Simulation* **9** (2) (2006) 9.
- [34] Ginsberg, M. (Ed.): *Readings in nonmonotonic reasoning*, San Mateo, CA: Morgan Kaufmann 1987.
- [35] Goodwin, B. C.: A cognitive view of biological process. *Journal of Social and Biological Structures* **1** (1978) 117–125.
- [36] Gulick, R. van: Consciousness, intrinsic intentionality and self-understanding machines. In A. J. Marcel, E. Bisiach (Eds.): *Consciousness in contemporary science*. Oxford, U. K.: Clarendon Press, 1988, pp. 78–100.

Martin Takáč

- [37] Harnad, S.: The symbol grounding problem. *Physica D* **42** (1990) 335–346.
- [38] Holland, O. (Ed.): Machine Consciousness [Special Issue]. *Journal of Consciousness Studies* **10** (4–5) (2003).
- [39] Icogno: *What AI techniques does Jabberwacky use?* Retrieved on December 8, 2007, from http://www.icogno.com/what_ai_techniques.html
- [40] Jong, E. de, Vogt, P.: How should a robot discriminate between objects? A comparison between two methods. In *Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior SAB'98*. Cambridge, MA: MIT Press, 1998, pp. 86–91.
- [41] Kegl, J., Senghas, A., Coppola, M.: Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. DeGraff (Ed.): *Language creation and language change: creolization, diachrony, and development*. Cambridge, MA: MIT Press, 1999.
- [42] Kelemen, J.: The agent paradigm. *Computing and Informatics* **22** (2003) 513–519.
- [43] Kirby, S., Hurford, J.: The emergence of linguistic structure: an overview of the iterated learning model. In D. Parisi, A. Cangelosi (Eds.): *Computational approaches to the evolution of language and communication*. Berlin, Germany: Springer-Verlag, 2001, pp. 121–148.
- [44] Kováč, L.: Fundamental principles of cognitive biology. *Evolution and Cognition* **6** (2000) 51–69.
- [45] Kováč, L.: Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu [Human consciousness is a product of evolutionary escalation of emotional selection]. In J. Kelemen (Ed.): *Kognície a umělý život III*. Opava, Czech Republic: Slezská univerzita, 2003, pp. 75–93.
- [46] Kripke, S. A.: A completeness theorem in modal logic. *Journal of Symbolic Logic* **24** (1959) 1–15.
- [47] Lakoff, G.: *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press 1987.
- [48] Lakoff, G., Johnson, M.: *Metaphors we live by*. Chicago, IL: University of Chicago Press 1980.
- [49] Langacker, R. W.: *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford, CA: Stanford University Press 1987.
- [50] Langacker, R. W.: *Foundations of cognitive grammar: Descriptive applications* (Vol. 2). Stanford, CA: Stanford University Press 1991.
- [51] Langton, C. G. (Ed.): *Artificial life*. Reading, MA: Addison-Wesley 1989.
- [52] Li, P., Farkaš, I., MacWhinney, B.: Early lexical acquisition in a self-organizing neural network. *Neural Networks* **17** (8–9) (2004) 1345–1362.
- [53] Maturana, H. R., Varela, F. J.: *The tree of knowledge: The biological roots of human understanding*. Boston, MA: Shambhala 1987.

- [54] Mirolli, M., Parisi, D.: Language as an aid to categorization: A neural network model of early language acquisition. In *Modelling language, cognition and action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific, 2005.
- [55] Montague, R.: *Formal philosophy: Selected papers of Richard Montague* (R. Thomason, Ed.). New Haven, CT: Yale University Press 1974.
- [56] Nehaniv, C.: The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In B. Edmonds, K. Dautenhahn (Eds.): *AISB Symposium: Starting from Society – the application of social analogies to computational systems*. Society for the Study of Artificial Intelligence and Adaptive Behaviour, 2000, pp. 73-84.
- [57] Nilsson, N. J.: *Shakey the robot* (Tech. Rep. No. 323). Menlo Park, CA: AI Center, SRI International 1984.
- [58] Oliphant, M.: *Formal approaches to innate and learned communication: Laying the foundation for language*. Unpublished doctoral dissertation, University of California, San Diego, CA 1997.
- [59] Peirce, C. S.: *Collected writings* (Hartshorne, C., Weiss, P., Burks, A. W., Eds.). Cambridge, MA: Harvard University Press 1931-58.
- [60] Pfeifer, R., Scheier, C.: *Understanding intelligence*. Cambridge, MA: MIT Press 1999.
- [61] Piaget, J.: *The child's construction of reality*. London: Routledge and Kegan Paul 1937/1955.
- [62] Piaget, J., Inhelder, B.: *La psychologie de l'enfant* [The psychology of the child]. Paris: PUF 1966.
- [63] Pulvermüller, F.: Words in the brain's language. *Behavioral and Brain Sciences* **22** (2) (1999) 253–279.
- [64] Rizolatti, G., et al.: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* **3** (1996) 131–141.
- [65] Rosch, E.: Principles of categorization. In E. Rosch, B. Lloyd (Eds.): *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978, pp. 27–48.
- [66] Roy, D.: Semiotic schemas: a framework for grounding language in action and perception. *Artificial Intelligence* **167** (1–2) (2005) 170–205.
- [67] Searle, J. R.: Minds, brains, and programs. *Behavioural and Brain Sciences* **3** (1980) 417–457.
- [68] Siskind, J. M.: A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* **61** (1–2) (1996) 1–38.
- [69] Smith, A. D. M.: *Evolving communication through the inference of meaning*. Unpublished doctoral dissertation, Theoretical and Applied Linguistics, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh 2003.

Martin Takáč

- [70] Smith, A. D. M.: The inferential transmission of language. *Adaptive Behavior* **13** (4) (2005) 311–324.
- [71] Steels, L.: Constructing and sharing perceptual distinctions. In M. van Someren, G. Widmer (Eds.): *Proceedings of the European Conference on Machine Learning*. Berlin, Germany: Springer, 1997, pp. 4–13.
- [72] Steels, L.: *The Talking Heads experiment. Words and meanings*. (Vol. 1). Antwerpen: Laboratorium 1999.
- [73] Steels, L.: Language as a complex adaptive system. In M. Schoenauer (Ed.): *Proceedings of PPSN-VI*. Berlin, Germany: Springer, 2000, pp. 17–26.
- [74] Steels, L.: Constructivist development of grounded construction grammars. In D. Scott, W. Daelemans, M. Walker (Eds.): *Proceedings Annual Meeting Association for Computational Linguistic Conference*. Barcelona, 2004, pp. 9–19.
- [75] Steels, L., Kaplan, F.: Situated grounded word semantics. In T. Dean (Ed.): *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*. San Francisco, CA: Morgan Kauffmann, 1999, pp. 862–867.
- [76] Steels, L., Kaplan, F., McIntyre, A., Looveren, J. V.: Crucial factors in the origins of word-meaning. In A. Wray (Ed.): *The transition to language*. Oxford, U. K.: Oxford University Press, 2002, pp. 252–271.
- [77] Steels, L., Vogt, P.: Grounding adaptive language games in robotic agents. In I. Harvey, P. Husbands (Eds.): *Advances in artificial life. Proceedings of the Fourth European Conference on Artificial Life*. Cambridge, MA: MIT Press, 1997, pp. 474–482.
- [78] Steels, L., Wellens, P.: How grammar emerges to dampen combinatorial search in parsing. In P. Vogt, Y. Sugita, E. Tuci, C. Nehaniv (Eds.): *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*. Berlin/Heidelberg, Germany: Springer, 2006, pp. 76–88.
- [79] Šefránek, J.: Kognícia bez mentálnych procesov [Cognition without mental processes]. In J. Rybár, L. Beňušková, V. Kvasnička (Eds.): *Kognitívne vedy*. Bratislava: Kalligram, 2002, pp. 200–256.
- [80] Šefránek, J., Takáč, M., Farkaš, I.: Vznik inteligencie v umelých systémoch. In D. Magdolen (Ed.): *Hmota, život, inteligencia: Vznik*. Bratislava: VEDA, 2008, pp. 245–270.
- [81] Takáč, M.: Categorization by sensory-motor interaction in artificial agents. In D. Fum, F. Del Missier, A. Stocco (Eds.): *Proceedings of the 7th International Conference on Cognitive Modeling*. Trieste, Italy: Edizioni Goliardiche, 2006, pp. 310–315.
- [82] Takáč, M.: Cognitive semantics for dynamic environments. In P. Hitzler, H. Schärfe, P. Øhrstrøm (Eds.): *Contributions to ICCS 2006 – 14th International Conference on Conceptual Structures*. Aalborg, Denmark: Aalborg University Press, 2006, pp. 202–215.

- [83] Takáč, M.: Konštrukcia významov a jej dynamika v procese iterovaného učenia [Construction of meanings and its dynamics in the iterated learning process]. In J. Kelemen, V. Kvasnička, J. Pospíchal (Eds.): *Kognície a umělý život VII*. Opava: Slezská univerzita, 2007, pp. 341–347.
- [84] Takáč, M.: *Construction of meanings in living and artificial agents*. Unpublished doctoral dissertation, Comenius University of Bratislava, Slovakia 2007.
- [85] Takáč, M.: Autonomous construction of ecologically and socially relevant semantics. *Cognitive Systems Research* **9** (4) (2008) 293–311.
- [86] Talmy, L.: *Toward a cognitive semantics*. Cambridge, MA: MIT Press 2000.
- [87] Tarski, A.: Pojecie prawdy w jezykach nauk dedukcyjnych [The concept of truth in the languages of the deductive sciences]. *Prace Towarzystwa Naukowego Warszawskiego, Wydział III Nauk Matematyczno-Fizycznych* **34** (1933) 13–172.
- [88] Tomasello, M.: *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press 1992.
- [89] Tschacher, W., Dauwalder, J.-P. (Eds.): *The dynamical systems approach to cognition: Concepts and empirical paradigms based on self-organization, embodiment, and coordination dynamics* (Vol. 10). Singapore: World Scientific, 1999.
- [90] Turing, A. M.: Computing machinery and intelligence. *Mind* **59** (1950) 433–460.
- [91] Vogt, P.: *Lexicon grounding on mobile robots*. Unpublished doctoral dissertation, Vrije Universiteit Brussel, Belgium 2000.
- [92] Vogt, P.: The physical symbol grounding problem. *Cognitive Systems Research* **3** (3) (2002) 429–457.
- [93] Vogt, P.: The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence* **167** (1-2) (2005) 206–242.
- [94] Vogt, P., Divina, F.: Language evolution in large populations of autonomous agents: issues in scaling. In *Proceedings of AISB 2005: Social Intelligence and Interaction in Animals, Robots and Agents*. 2005, pp. 80–87.
- [95] Vogt, P., Divina, F.: Social symbol grounding and language evolution. *Interaction Studies* **8** (1) (2007) 31–52.
- [96] Vygotsky, L. S.: *Mind in society: The development of higher mental processes* (Rice, E., Ed. Trans.). Cambridge, MA: Harvard University Press 1930-35/1978.
- [97] Waxman, S. R.: Everything had a name, and each name gave birth to a new thought: Links between early word-learning and conceptual organization. In D. G. Hall, S. R. Waxman (Eds.): *Weaving a lexicon*. Cambridge, MA: MIT Press, 2004, pp. 295–335.
- [98] Weizenbaum, J.: Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM* **9** (1) (1966) 36–45.
- [99] Whorf, B. L.: *Language, thought and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carrol, Ed.). Cambridge, MA: MIT Press 1956.

Martin Takáč

- [100] Winograd, T.: *Procedures as a representation for data in a computer program for understanding natural language*. Unpublished doctoral dissertation, MIT, Cambridge, MA 1971.
- [101] Wittgenstein, L.: *Philosophical investigations*. New York: Macmillan 1953.

Kľúčové slová

význam
agent
sémantika – denotačná, funkcionalistická, intenzionálna, extenzionálna,
kognitívna
konceptuálny priestor
rozlišovacie kritérium
semiotika
semiotický trojuholník
reprezentácia – vyvolaná, oddelená, ikonická, kategoriálna
metafora čínskej izby
problém ukotvenia symbolov
konštruktivizmus – kognitívny, sociálny
ELIZA
SHRDLU
Shakey
diskriminačný strom
prototyp
dvojcestná neurónová sieť
chatbot
afordancie
jazyková hra – diskriminačná, hádacia, pozorovacia