

Iná čínska izba **(o rozumení)**

Ján Šefránek

Úvod

V druhej polovici 20. a začiatkom 21. storočia sa do dejín slovenskej filozofie výrazným spôsobom zapísal muž, ktorý sa úporne usiloval o novú metódu filozofovania. Charakteristickou črtou tejto metódy je dôraz na zrozumiteľnosť. Mám na mysli rozprávanie a písanie, pri ktorom autor vie, o čom hovorí, a čitateľovi, poslucháčovi to prezentuje – často so značným vypätím – jasne a dôsledne. Meno toho muža je Paľo Cmorej.

Metóda, o ktorej je reč, preferuje dôkladnú pojmovú analýzu a precíznú konštrukciu pojmového aparátu. Treba zdôrazniť, že ide o analýzu a konštrukciu využívajúcu prostriedky logiky. Ak použijeme negatívne vymedzenie, ide o vyhýbanie sa špekuláciám porušujúcim elementárne princípy logiky a ignorujúcim empiriu, ide o nedôveru voči textom, v ktorých honosná škrupina dobre znejúcich slov neobaľuje zreteľne rozpoznateľný význam. Dodávam: uvedomujem si, že táto metóda nie je ani postačujúca, ani vhodná pre všetky ambície filozofov a filozofie.

Určite išlo o úsilie, ktoré v kontexte súdobej slovenskej filozofie pôsobilo sviežo a kacírsky. Spracúvalo podnety logického empirizmu (a príbuzných smerov či jednotlivcov, azda zvlášť treba vyzdvihnúť poľskú školu) – viac ako metódy než ako doktríny.

Trochu čisto osobného: Paľo mal rozhodujúci vplyv na to, čo som sa pokúšal v živote robiť (vo sfére, ktorá by sa dala nazvať odbornou). Napriek tomu, že väčšinu profesionálneho života som strávil pomerne ďaleko (niekoľko ulíc) od neho. Samozrejme, ani týmto amatérskym textom, ani týmto osobným vyznaním nechcem vrhnúť negatívny tieň na Paľovu dominantnú rolu v slovenskej filozofii. Verím, že ju ešte významne podčiarkne dielom sedemdesiatnika.

V texte sa venujem rozumeniu, zrozumiteľnosti. Poznamenajme, že zrozumiteľnosť sa v tradícii myšlienkových prúdov, ovplyvnených modernou logikou, môže bližšie zrozumiteľnosti pre stroj. (Tento text však nechce byť zrozumiteľný pre stroj.)

Téma čínskej izby sa teda vynára celkom prirodzene. I tak sa však za ňu ospravedlňujem. Pomocou čínskej izby chcem motivovať to, ako neprístupovať k analýze rozumenia. Potom prídem s metaforou inej čínskej izby, ktorou chcem motivovať to, ako k analýze rozumenia pristupovať. Mojm cieľom je postaviť problém tak široko, aby rozumenie nebolo nevyhnutne viazané na ľudskú myseľ. Načrtnem sémantický rámec, v ktorom možno hovoriť o rozumení opíc alebo psov. Ukáže sa, že rozumenie a usudzovanie sú dve strany tej istej mince. Potiaľ bude text bežať skôr po príkladoch. Na záver príde pokus o všeobecnejšie uchopenie témy. Uvedomujem si, že text nie je cmorejovsky zrozumiteľný. Viem si však predstaviť, že sa dá dotiahnuť (v priebehu jedného desaťročia :-)) do takého tvaru. Dúfam, že to bude vidieť aj čitateľ.

Searlova čínska izba

Searlovu metaforu čínskej izby (Searle 1980, 1984, 1990, 1999) zrejme netreba čitateľovi približovať, pre úplnosť to však urobím (vychádzam 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 škatule 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 poselať 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 tá *nerozumie* ani slovo.

Podrobnú analýzu Searlovej metafory pozri napríklad v Cole (2004). Trochu sme si ju všimli aj v Šefránek – Farkaš – Takáč (2007) a Šefránek (2002).

Tu sa chcem venovať rozumeniu. Searle si nerobí žiadny problém s významom slova rozumieť.¹ Ten je však pre pochopenie (rozumenie) príbehu o čínskej izbe kľúčový. Na základe niektorých Searlových 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?

Ďalej sa Searle opiera o zdôrazňovanie úlohy ľudského biologického „hardvéru“ ako „kauzálnej sily“, ktorá generuje myslenie. Dostávame tautologický argument typu „počítač nemôže rozumieť, myslieť, lebo počítač nemôže generovať myslenie“. Aký hardvér môže ešte 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.

Navyše, aj keby sme introspekciu (alebo ľudské mentálne procesy) považovali za nevyhnutnú podmienku rozumenia, určite nie je postačujúcou podmienkou. *Introspekciou (mentálnymi procesmi) môžeme nadobudnúť presvedčenie, že rozumieme i vtedy, keď nerozumieme.*

Pointa:

Táto čínska izba je založená na predpoklade, že je jasné, čo znamená (ne)rozumieť. Tento predpoklad je minimálne neproduktívny. Iba na okraj – obávam sa, že disponujúc mašineriou čínskej izby a používajúc ju, by som rozumel čínsky postupne stále viac. Kvôli zjednodušeniu situácie si miesto čínštiny predstavme portugalčinu. Treba dodať, že opisovaná mašineria čínskej izby je kdesi za dosiahnuteľnou latkou súčasného poznania a jej implementácia by vyžadovala extrémne detailné pochopenie čínštiny.

¹ Prípomínam, že v tomto kontexte nie je zaujímavé poslanstvo jeho metafory pre poňatie mysle alebo pre možnosti tzv. silnej umelej inteligencie.

Iná čínska izba

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 skrupina Číňanov. Spolu s nimi je v miestnosti Armén. Armén s nimi nekonverzuje, vyzerá to tak, že nevie a nerozumie čínsky. Číňania zamýšľajú napláňovať 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 dovoľí 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, ...).²

Teda: o tom, či niekto rozumie, sa môžeme presvedčiť iba tak, že pozorujeme *správanie* (verbálne či neverbálne), ktoré zodpovedá hypotéze o pochopení (na základe nejakej dobre zdôvodnenej „teórie“). Zdôrazňujem, že nejde o *čisto* behaviorálne kritérium: pozorovať správanie samo osebe nestačí, pozorovanie sa musí opierať o solídne poznanie relevantného kontextu a táto korešpondencia sa potvrdzuje/vyvracia nejakými usudzovacími operáciami.

Pointa:

To, že X rozumie niečomu, možno a treba overovať pozorovaním jeho správania opierajúcim sa o „teóriu“ (toho niečoho).

Rozumenie ako usudzovanie

Uvedme 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 tomuto poj-

² Keď hovorím o rozumení, nemám na mysli iba rozumenie prirodzenému jazyku. Uvediem niekoľko kontextov, ktoré naznačujú, že je primerané chápať predikát *rozumieť* veľmi všeobecne: „rozumie, prečo to Robert urobil“, „nerozumie teórii relativity“, „títo dvaja futbalisti si na ihrisku rozumejú“.

mu („teória“, o ktorú sa pritom dá oprieť, umožňuje odfiltrovať napríklad numerické chyby). Zdôraznime úlohu usudzovacích operácií – tie sú nevyhnutné pri zdôvodnení korektnosti výpočtu; napokon aj pri výpočte. Tento príklad a táto analýza ešte 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. Podľa Ladislava Kováča, zakladateľa kognitívnej biológie (1986), evolúcia je vlastne evolúciou poznania.

Pointa:

Môžeme povedať viac, než to, že rozumenie možno a treba preverovať pozorovaním správania. Test, že X rozumie čomusi, je presvedčivý vtedy, keď pomocou tohto testu overujeme dôsledky predpokladu, že X tomu rozumie (alebo nerozumie) s využitím poznatkov o príslušnej doméne. Neočakávajme však, že X je deduktívny stroj. Pod dôsledkami si predstavujem aj výsledky zvieracieho usudzovania (opretého o zvieracie poznanie).

Rozumenie ako biologický fenomén

Pointa:

Mnohé experimenty a pozorovania (minimálne) naznačujú, že chápanie/rozumenie a usudzovanie možno pripísať aj vyšším živočíchom. Ako sa k tomu postaviť? Považovať to za čiru metaforu a skutočné rozumenie „dovoliť“ iba u „koruny tvorstva“, človeka? Myslí si, že plodnejšie je nájsť také sémantické pojmy, ktoré umožnia zmysluplne hovoriť o rozumení a usudzovaní vlkov, gepardov alebo kačíc. Mali by sme tak väčšiu šancu lepšie pochopiť ľudské chápanie a usudzovanie, ak veríme, že sa vyvinulo z elementárnejšieho biologického podhubia. Mimochodom, človek úspešne usudzuje tak, že v okamihu vidí dôsledok (riešenie) bez toho, že by musel v mysli konštruovať reťaz odvodení. Domnienka, že táto schopnosť má evolučné korene a že jej opísanie a pochopenie by sa mohlo opierať o opísanie a pochopenie zvieracieho usudzovania, sa zdá byť sľubná, hodna preverenia.

Ak nájdeme takýto sémantický opis, mohli by sme ho aplikovať na jednoduchých (abstraktných) agentov bez toho, že by sme u nich predpokladali mentálne procesy. (Filozofická) konzekvencia takéhoto pokusu

je aspoň dvojaká. Jednak ide o prihlásenie sa k ne-psychologizmu, k tradícii, ktorú veľmi jasne artikuloval pre štúdium významov Frege. Jednak ide o akceptovanie podnetov psychológie a kognitívnej vedy s cieľom rozširovať horizont sémantických a logických (hovoríme o usudzovaní) konštrukcií; dokonca až k opisu predjazykových správání.

Autori experimentu opísaného v (Bräuer 2006) testovali schopnosť zvieratá (šimpanza a domestikovaného psa) rozpoznať, kde je ukrytá potrava. Toto rozpoznanie mohlo využívať tri typy podnetov: komunikatívne, kauzálne a behaviorálne. Zameriame sa na prvé dva typy, ktoré sú pre tento experiment podstatné.

Úlohou pokusného zvieratá bolo vybrať si z dvoch nádob obrátených hore dnom tú, pod ktorou sa nachádza potrava. Ako komunikatívne podnety sa použili ukazovanie na nádobu s potravou a pohľad smerom na nádobu. Kauzálne podnety boli realizované fyzickou manipuláciou (napríklad hrkaním nádobou). Ak zviera rozumelo kauzálnym vzťahom medzi lokalizáciou zvuku a umiestnením potravy mohlo na tom základe usudzovať.

Šimpanzy podľa očakávania veľmi dobre reagovali na kauzálne podnety, no nevenovali veľkú pozornosť ukazovaniu. Psy naopak spoľahlivo nasledovali komunikačné podnety, ale umiestnenie potravy neboli schopné odvodiť na základe kauzality.

Podľa citovanej práce (a mnohých ďalších) má zmysel hovoriť o tom, že zvieratá *rozumejú a usudzujú, odvodzujú dôsledky z toho, čo pochopili*. Opice rozumejú kauzálnym podnetom, psy komunikačným podnetom.

Vedy, ktoré sa tradične venujú skúmaniu usudzovania, chápania (a súvisiacich fenoménov, ako je *poznanie a význam*) sa sústreďujú na poznatky a úsudky sformulované v *jazyku*. Veľkou výzvou (aj pre tieto vedy) je skúmať, čo je usudzovanie a pochopenie ako *biologický fenomén*, ktorý sa vyskytuje v prírode v *predjazykovom štádiu*.

Verím, že takto postavený cieľ pomôže hlbšie, mnohotvárnejšie a v inom svetle pochopiť ľudské usudzovanie, poznanie, reprezentáciu poznania a významov.

V Retová et al. (2007) sme načrtli pojmovú aparatúru, v ktorej sa dá spresniť predstava o usudzovaní a chápaní agentov s predjazykovým správaním. Použili sme sémantiku rozlišovacích kritérií (Šefránek 2002). Jej spresnenie a rozpracovanie pre potreby výpočtového modelovania evolúcie a akvizície významov a jazyka možno nájsť v Takáč (2006a; b; c; 2007).

Významy modelujeme ako rozlišovacie kritériá: ak nejaký organizmus dokáže rozlíšiť, čo je jedlé (nebezpečné, v pohybe a pod.) a čo nie,

osvojil si akýsi význam. Malé deti skôr, než si osvoja nejaké slovo, napríklad pes, musia vedieť rozlíšiť psov od ne-psov (i keď sa môže stať, že za psov považujú aj holuby).

Pochopenie stotožňujeme s rozpoznávaním a tvorením významov (rozlišovacích kritérií). Rozlišovacie kritérium je abstrakcia schopnosti rozlišovať a možno ho definovať/implementovať ako nejakú funkciu. Ak organizmus (agent) dokáže používať nejaké rozlišovacie kritérium pri rozpoznávaní objektov, vlastností (atd.), alebo keď vytvára nové rozlišovacie kritériá, aby sa zorientoval v prostredí, môžeme hovoriť, že chápe (vlastnosť, vzťah, situáciu, udalosť, závislosť, pravidlo atd.).

Experimenty z Bräuer et al. (2006) sme teda analyzovali v sémantickom rámci rozlišovacích kritérií. Definovali sme rozlišovacie kritériá situácií, udalostí, typov situácií, typov udalostí, problémov a pravidiel.

Pravidlá sú špeciálne rozlišovacie kritériá, ktoré sa používajú pri usudzovaní. Nejakým rozlišovacím kritériám typov situácií alebo udalostí (významom predpokladov pravidla) priradia nejaké rozlišovacie kritériá (významy záveru).

Zaviedli sme dva druhy pravidiel – akčné a situačné. Dôsledkami prvého z nich sú rozlišovacie kritériá akcie, u druhých rozlišovacie kritériá typu situácie. Intuitívne, prvé vedú agenta ku konaniu, druhé k rozpoznaniu stavu prostredia.

Pri tejto sémantickej konštrukcii sa ukazuje, že usudzovanie je špeciálnym prípadom chápania: ak niekto alebo niečo dokáže produktívne usudzovať, v istom zmysle slova chápe vzťahy (napríklad medzi udalosťami).

Poznámka o tom, či a ako možno predpoklad o tom, že opice (alebo iné vyššie živočíchy) usudzujú, potvrdiť pozorovaním, t.j. ako možno operacionalizovať hypotézu o usudzovaní opíc (alebo iných živočíchov): Samozrejme, inú možnosť ako cez pozorovanie správania nemáme. Pozorujeme správanie opíc (alebo ľubovoľných agentov v predjazykovom štúdiu) v určitých (takmer) konštantných podmienkach počas udalosti toho istého typu. Predpokladajme, že za týchto podmienok má opica na výber z rôznych možných správání. Ak v štatisticky významnej miere si vyberá jedno z nich a jej rozhodovanie sa zmení pri určitej zmene daných podmienok, môžeme prijať, že daný výber správania sa deje na základe usudzovania. Dôležité je dodať, že presnejšia špecifikácia bude s najväčšou pravdepodobnosťou závislá od konkrétnej domény a konkrétneho rozhodovania (podobne sú aj defaultové pravidlá doménovo špecifické; pozri napríklad Šefránek 2000; mimochodom, v Retová et al.

2007 sme použili aj trochu komplikovanejšie typy pravidiel, ktoré úplne zodpovedali defaultovým pravidlám).

Konceptuálny aparát rozlišovacích kritérií je flexibilný a možno ho použiť od opisu veľmi jednoduchých prípadov rozlišovania až po rozlišovanie opreté o používanie syntakticky bohato štruktúrovaného jazyka charakterizovateľného teoreticko-modelovou (alebo komplikovanejšou) sémantikou. Domnievame sa, že analýza, formalizácia a implementácia rýchleho (odlišného od logicky korektného, ale zdĺhavého) usudzovania v jazyku môže významne ťažiť z pokusov pochopiť zvieracie usudzovanie. Na rozlišovacie kritériá typov situácií, udalostí pravdepodobne možno naviazať rozmanité, heterogénne typy reprezentácií (od propozíčných po grafické vizualizácie). Od dôkladného štúdia zvieracieho usudzovania možno urobiť dôležité kroky pre pochopenie špecifickej ľudskej usudzovania (a predpokladáme, že tak možno dosiahnuť podstatne mnohotvárnejšie charakterizácie, než sú tie, ktoré dnes študuje logika).

Intermezzo

Napriek tomu, že rozumenie tu chápem veľmi široko (nie iba ako rozumenie jazyku), predsa len treba komentár, ktorým zameriam pozornosť na jadro problému (a tým na zamýšľanú interpretáciu rozumenia) tak, ako ho chcem vidieť.

V koncepciách blízkych logike je s rozumením (ak tomu dobre rozumiem) spätá nasledovná intuícia: vete rozumiem vtedy, keď viem, aký by bol svet, keby veta bola pravdivá. Táto myšlienka sa dá sformalizovať priamočiaro na báze rôznych logických sémantik.

Tu chcem byť o niečo náročnejší, už vieme, že rozumenie spájam s usudzovaním.

Všimnime si vetu: „Očakávala som, že po tom, čo si týždeň nebudete kvapkať do očí, váš vnútroočný tlak vyskočí na 45“.

(Každá) veta použitá v nejakom texte alebo rozhovore má nejaký doslovný obsah, ale okrem toho má aj nejaký zmysel v danom kontexte. Táto veta v tomto kontexte čosi hovorí o pokročilosti ochorenia, o tom, ako ho lekárka hodnotila pred týždenným experimentálnym vynechaním kvapkania a ako po ňom (reálne namerané hodnoty boli podstatne nižšie). Vete rozumiem iba potiaľ, že lekárka pôvodne čosi očakávala. Ani netuším, prečo to očakávala, aký bol jej názor na stav ochorenia a ako sa zmenil; vete dobre nerozumiem.

Tradične (a z rozumných dôvodov) sa poznanie významu a poznanie skutočnosti ostro oddeľuje. Vyššie uvedená veta je zrozumiteľne formulovaná. Ďalej budem (ešte podrobnejšie) motivovať to, prečo s rozumom chcem spájať prísnejšie požiadavky, i keď akceptujem aj tradičné chápanie a jeho motiváciu.

Mimochodom, aj v Searlovej (napokon, aj v mojej) čínskej izbe rozumenie čínštine nemôže byť dostatočne dobre implementované, ak popri významoch jazyka neberieme do úvahy aj znalosti o svete a zdravý ľudský úsudok. Výskumníci z oblasti výpočtovej lingvistiky vedia už vyše 30 rokov, že to platí pre každý prirodzený jazyk.

Uvedená veta však naznačuje viac. Naznačuje, že pri analýze rozumenia je (niekedy a pre niektoré ciele) dôležité brať do úvahy kontext, dôvody, vysvetlenia, fakty a hypotézy. V tejto súvislosti možno spomenúť Reason Maintenance Systems. Ich štúdium a konštrukciu odštartoval Doyle (1979), ktorý proti atomickej reprezentácii (vety ako izolované atómy) postavil reprezentáciu, kde s každou vetou boli asociované dôvody pre jej platnosť a dôvody proti jej platnosti. Pravda, poznanie, ale ani význam nie sú v tomto prístupe atomizovateľné, dôvody, argumenty ich prepájajú (alebo vzájomne vylučujú).

Zdôrazním ešte jednu dôležitú črtu spomínaného systému: autoreflexivnosť. Je užitočné s poznatkami asociovať naše epistemické postoje k nim. Onedlho sa k tomu vrátíme.

Prejdime teraz k rozumeniu matematických teorém. Nejakej matematickej teoréme rozumiem vtedy, keď si ju viem dokázať (o niečo menej jej rozumiem, ak viem, ako by sa dala dokázať – priestor pre stupne rozumenia tu je evidentný).

Všimnime si teorému “problém existencie stabilného modelu konečného normálneho propozičného logického programu je NP-úplný”.

Podľa tradičného chápania tejto vete rozumiem vtedy, keď poznám podmienky jej pravdivosti (v tomto prípade ide o to, že význam konštrukcie “problém existencie stabilného modelu konečného normálneho propozičného logického programu” je v dobre definovanom zmysle slova subsumovaný významom konštrukcie “NP-úplný problém”).

Ak poznám definíciu každého z použitých pojmov, mám dobrú predstavu, o čom sa hovorí; rovnako dobrú by som však mal aj vtedy, keby v našej teoréme miesto “je” bolo “nie je”.

Rozumenie chcem chápať takto: Až vtedy, keď si z tejto dvojice dokážeme jednu vybrať (na základe pevných argumentov, povedzme, že na základe dôkazu), až vtedy veci naozaj rozumieme.

Prečo?

Pre opis zvieracieho správania, ako sme videli vyššie, nie je dosť dobre možné rozumenie ostro oddeliť od poznania (pravdy). Ešte jeden príklad, rozprávkový. V Tolstého rozprávke sa malá myška bojí kohúta, lebo jej pripadá hrozny, ale mačku považuje za milé stvorenie. Očividne myška nerozumie situácii na dvore.

To, či zviera niečomu rozumie, nemôžeme preverovať vzhľadom na všetky možné svety, ale iba vzhľadom na tento svet (dalo by sa to zjemniť napríklad takto – vzhľadom na jeho epistemicky akceptovateľné alternatívy, vzhľadom na epistemicky alebo akčne preferované svety; nepúšťam sa však do toho). Zámerne chcem, aby zvieracie rozumenie bolo plnohodnotným prípadom rozumenia.

Ďalší argument je praktický, ide mi o vystihnutie určitého používania slova rozumieť. Ak niekto pri skúške z matematiky tvrdí nepravdy a zároveň hovorí, že látke rozumie, skúšajúci mu zrejme príliš neverí.

Príkladom matematickej teórie som chcel ukázať, že o rozumení viet sa dozviem veľmi málo, ak ho ostro oddelím od poznania (pravdy); preto akceptujem biologicky dobre motivovaný prístup k rozumeniu, tesne naviazaný na poznanie, pravdu a usudzovanie.

Hrozí nám však jeden nepríjemný dôsledok – nepravdivým vetám nemožno rozumieť. Preto súčasťou rozumenia bude podľa nás schopnosť autoreflexie (nie schopnosť introspekcie!). Ak rozumieme na základe argumentov (dôvodov, vysvetlení, dôkazov), nie je komplikované zaujať akýsi epistemický (meta-)postoj. Túto črtu možno navrhnúť pre program (tu odhliadam od aspektu výpočtovej zložitosti), preto ju netreba spájať s introspekciou. V prípade, že máme nepravdivú vetu, rozumieme jej – v navrhovanom poňatí – ak vieme zdôvodniť jej nepravdivosť (teda zaujať príslušný autoreflexívny postoj). Samozrejme, od bipolárneho prípadu pravda/nepravda, rozumiem/nerozumiem možno a treba prejsť k mnohohodnotovému, mnohostupňovému prípadu.

Napokon, je tu ešte prípad hypotéz (pokiaľ sme na pôde matematiky, otvorených problémov, zatiaľ nedokázaných tvrdení).³ Všimnime si známy otvorený problém z teórie výpočtovej zložitosti. Označme ho $P \stackrel{?}{=} NP$. Predpokladá sa, že táto rovnosť neplatí, dôkaz však nie je známy. Keďže v navrhovanej koncepcii rozumenia beriem do úvahy aj autoreflexiu, máme pripravenú pôdu aj pre (akési) rozumenie matematickým vetám, ktoré nie sú dokázané ani vyvrátené. Michal Malý však

³ Za tento podnet ďakujem Michalovi Malému.

upozornil na ďalší dôležitý aspekt. Ak predpokladáme nerovnosť $P \neq NP$, sme schopní dokázať mnoho jej cenných dôsledkov, a teda v istom zmysle slova jej lepšie rozumieť. Toto je črta, ktorou možno charakterizovať hypotetické usudzovanie vo všeobecnosti. Vidno, že téma rozumenia je pestrá a komplikovaná, otvára mnohé zaujímavé problémy.

Zhrňme. Oba príklady (matematický i nematematický) ukazujú, že z akýchsi dôvodov je užitočné rozumenie zakladať na argumentácii, dôvodoch, vysvetľovaní, dôkazoch a autoreflexii, ale aj dôsledkoch (a domyslíme si aj iné druhy súvislostí – teraz mám na mysli usudzovanie na základe analógie).

Pripomeňme si, že sme v rámci sémantiky rozlišovacích kritérií. Mojm cieľom je používať ich aj pre porozumenie porozumeniu. Navrhnutá sémantika je zaujímavá (ak teda je zaujímavá) aj (hlavne?) preto, že sa ostro odlišuje od obvyklých logických sémantik v poňatí sémantickej interpretácie viet. Pravdivostná hodnota (alebo funkcia z možných svetov do pravdivostných hodnôt) ako (jediná) sémantická charakterizácia vety je niekedy nepostačujúca. Navyše, rozlišovacie kritérium vety (kontextu, situácie, udalosti) musí mať konštruktívny charakter, ak má rešpektovať požiadavky sformulované vyššie. Rozlíšenie má byť opreté o argumentáciu.⁴

Na záver ešte jeden problém. Rozumenie spájam s usudzovaním (argumentáciou atd.) a to má zmysel vtedy, ak sa sústredíme na vety a kontexty (alebo na situácie či udalosti, ak nechceme zostať v zajatí jazyka).

Preto sa príliš nezaujímam o rozumenie vlastným menám, menám vlastností, vzťahov a pod. (uvedomujem si však množstvo sémantických a pragmatických problémov na tejto úrovni, nie sú to však problémy, ktoré sú z hľadiska našej témy dôležité). Na tejto úrovni a v tomto kontexte iba predpokladám, že mám alebo nemám rozlišovacie kritérium nejakého objektu, triedy, vlastnosti, vzťahu atd.

⁴ Vidno tu istú príbuznosť ku konštruktívnym teóriám intenzie – Moschovakis (1994), Materna (1998), Materna – Duží (2004) alebo van Labmalgen – Hamm (2002). V týchto teóriách sa intenzia definuje ako algoritmus, ktorý počíta denotát výrazu. Ak by som túto neformálnu analýzu mal dotiahnuť do precízneho tvaru, preferoval by som ideu nedeterministického algoritmu, ktorý by pracoval asi takto: po výbere nejakého kandidáta na argumentáciu podporujúcu (vyvracajúcu) vetu (a pod.) by preveril, či argumentácia zodpovedá akýmisi podmienkam (na základe nejakých explicitne akceptovaných štandardov v prijatej koncepcii argumentácie, usudzovania).

Výpočtový model

Zhrňme doterajšie hlavné argumenty. O tom, že X rozumie Y, sa môžeme presvedčiť pozorovaním správania Xa, pričom pri zameraní našich pozorovaní sa opierame o nejakú teóriu, v ktorej sa dajú odvodiť (veľmi široko chápané) dôsledky predpokladu, že niekto (ne)chápe Y. Rozumenie sa dá poňať veľmi široko, dá sa zmysluplne (a empiricky zmysluplne) hovoriť o tom, že zvieratá rozumejú. Rozumenie vete, textu, diskurzu, kontextu, situácii, udalosti súvisí s argumentáciou, poznaním.

Celkom prirodzene sa vynára možnosť a cieľ zostrojiť abstraktnú a neživú entitu, ktorá rozumie vo vyššie uvedenom zmysle slova. Správa sa podľa istých očakávaní, odvodzuje očakávané dôsledky. Takouto entitou môže byť výpočtový model (program, softvérový agent).

Dá sa tvrdiť aj viac – výpočtový model nielen rozumie (v zmysle uvedenom vyššie), ale sa aj stáva prostriedkom umožňujúcim dôkladnejšie chápať rozumenie. Gruska a Jürgensen (1991) upozornili, že informatika sa stala novou metodológiou vied a podložili to detailnou argumentáciou. V súvislosti s možnosťou výpočtovo modelovať rozumenie vyzdvihujem dva dôležité metodologické aspekty, ktoré rozvíja informatika a dovedla ich do sofistikovaného tvaru. Mám na mysli metodologický prínos výpočtovej simulácie a metodologický prínos intenzívneho štúdia prostriedkov na deskripciu systémov (jazykov, reprezentácií).

Výpočtová simulácia, ak má byť netriviálna, vyžaduje veľmi detailné opisy relevantných objektov, akcií, vzťahov, procesov (veľmi detailnú teóriu oblasti a špecifikáciu problému). Oveľa detailnejšiu, než používali predchádzajúce metódy, ktoré sa mohli spoliehať na ľudskú intuíciu, znalosti a úsudok. Ak meníme niektoré z opisov, môžeme relatívne rýchlo zaznamenať a analyzovať vplyv týchto zmien, čo zosilňuje efektívnosť, operatívnosť a produktívnosť vývoja teórií a hypotéz.

Opisy objektov, akcií, vzťahov, procesov atď. sa stali samy predmetom informatického výskumu – mimoriadna pozornosť sa venuje štúdiu jazykov, reprezentácií. Pri aplikácii informatických nástrojov a metód (od písania programov po uplatňovanie výsledkov jemného teoretického výskumu) na riešenie problémov má kľúčovú úlohu práve vhodná, efektívna a adekvátna reprezentácia.

Ďalej iba načrtnem dôležité vlastnosti výpočtového modelu rozumenia. Takýto model musí mať prostriedky na reprezentáciu správania (aj jazykového, ak je to potrebné). Musí mať prostriedky na priradovanie

nejakých významov k správaniu a, samozrejme, prostriedky na opis sveta významov (musí to byť sémantický model). Čo je dôležité, použitá sémantika bude operovať s pojmom pravdy. Napríklad v experimente z Bräuer et al. (2006) to umožní sémanticky charakterizovať, kedy experimentálne zviera našlo potravu a kedy nie a vo všeobecnosti by sme stratili možnosť sformulovať objektívne kritériá rozumenia (v poňatí, ktoré tu preferujem). Napokon, musí mať prostriedky, ktoré mu umožnia formalizovať teóriu relevantnej domény a odvodzovať jej dôsledky. Simuláciami sa možno presvedčiť, či navrhnuté (reprezentované) správanie je v zhode s očakávaniami, t. j. s dôsledkami teórie danej domény.

Záver

Block (1980) uvažuje o tom, že môže byť užitočné/potrebné zosúladiť náš koncept pochopenia so svetom, v ktorom by roboty boli z toho istého prírodného druhu ako ľudia. Spýtajme sa, prečo. Abstrakcia od partikulárnych aspektov totiž býva plodná, pokiaľ nepresiahne hranicu, za ktorou dostávame samoučelné a neproduktívne hry so slovami alebo symbolmi. Všetci však vieme, že identifikácia tejto hranice je ťažká a že niekedy sú potrebné desaťročia alebo storočia na to, aby sa zdanlivo samoučelné hry stali veľmi užitočnými nástrojmi.

Domnievam sa, že širšie poňatie rozumenia môže byť produktívne. Nielen preto, že môže pomôcť realizovať víziu sémantického webu a aspoň čiastočne predstavu o inteligentných počítačoch, ale aj preto, že má potenciál prispieť k poznaniu sveta. Sveta, v ktorom je spracovanie informácií jednou z fundamentálnych črt. Navyše, nedokážeme spoľahlivo oddeliť naše poznanie sveta od nášho poznania informačných tokov, spracovania informácií.⁵

Z filozofického hľadiska je v pozadí (za námetmi tohto článku) téma externalizmu (ne-psychologizmu), vlastná optika logiky (Frege 1892) a téma mnohosti možností, ako robiť logiku (Makinson 2003). Rešpektujem aj dogmy empirizmu, preto považujem rozumenie za predmet empirického preverovania.

Pokúsil som sa načrtnúť kontext (rámec, „výpočtový model“), v ktorom možno špecifikovať nielen postačujúcu a nevyhnutnú podmienku

⁵ Spracovanie informácií, jeho zákonitosti a obmedzenia, poskytujú kľúč k pochopeniu fyzikálneho sveta, pozri Calude - Gruska (2007).

rozumenia, ale aj iných kognitívnych schopností, ak vyžadujú spracovanie poznatkov a usudzovanie.

LITERATÚRA

- BLOCK, N. (1980): Komentáre k Searle (1980). *Behavioral and Brain Sciences* 3, 417 – 457.
- BRÄUER, J. et al. (2006): Making Inferences about the Location of Hidden Food: Social Dog, Causal Ape. *Journal of Comparative Psychology* 120, No. 1, 38 – 47.
- CALUDE, C. S. – GRUSKA, J. (2007): Quantum informatics and the relations between informatics, physics and mathematics: a dialogue. *CDMTCS Research Reports Series*.
- COLE, D. (2004): The Chinese Room Argument. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), <http://plato.stanford.edu/archives/fall2004/entries/chinese-room/>.
- DOYLE, J. (1979): A Truth Maintenance System. *Artificial Intelligence* 12, 231 – 272.
- FREGE, G. (1892): Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100, 25 – 50.
- GRUSKA, J. – JÜRGENSEN, H. (1991): Maturing of Informatics. In: Bjørner, D. – Kotov, V. (eds.): *Images of programming*. Elsevier (North Holland).
- KOVÁČ, L. (1986): Úvod do kognitívnej biológie. *Biologické listy* 51, č. 3, 172 – 190.
- MAKINSON, D. (2003): Ways of Doing Logic: What was Different about AGM 1985? *Journal of Logic and Computation* 13, 3 – 13.
- MATERNA, P. (1998): *Concepts and Objects*. Acta Philosophica Fennica 63, Helsinki.
- DUŽÍ, M. – MATERNA, P. (2004): A Procedural Theory of Concepts and the Problem of Synthetic A Priori. *Korean Journal of Logic* 7, No. 1, 1 – 22.
- MOSCHOVAKIS, Y. (1994): Sense and denotation as algorithm and value. In: Oikonen, J. – Vaananen, J. (eds.): *Lecture Notes in Logic, #2 (1994)*. Springer, 210 – 249.
- RETOVÁ, D. – ŠILLOVÁ, J. – ŠEFRÁNEK, J. (2007): Opice, psy, sémantika a logika. *Kognice a umělý život* VII.
- 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? *Scientific American* 262, 26 – 31.
- SEARLE, J. R. (1999): The Chinese Room. In: Wilson, R.A. – Keil, F. (eds.): *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge: MIT Press.
- ŠEFRÁNEK, J. (2000): *Inteligencia ako výpočet*. Bratislava: IRIS.
- ŠEFRÁNEK, J. (2002): Kognícia bez mentálnych procesov. In: Beňušková, L. a kol. (eds.): *Kognitívne vedy*. Bratislava: Kaligram, 200 – 256.
- ŠEFRÁNEK, J. – FARKAŠ, I. – TAKÁČ, M. (2007): Vznik inteligencie v umelých systémoch. In: Magdolen, D. (ed.): *Hmota, život, inteligencia: Vznik*. Bratislava: Ve-

da. V tlači. Rozšírená verzia na:

<http://kedrigern.dcs.fmph.uniba.sk/reports/display.php?id=16>

- TAKÁČ, M. (2006a): Cognitive Semantics for Dynamic Environments. In: Hitzler, P. – Schärfe, H. – Øhrstrøm, P. (eds.): *Contribution to ICCS 2006 - 14th International Conference on Conceptual Structures*. Aalborg: Aalborg University Press, 202 – 215.
- TAKÁČ, M. (2006b): Kognitívna sémantika rozlišovacích kritérií. In: Kelemen, J. – Kvasnička, V. (eds.): *Kognice a umělý život VI*. Opava: Slezská univerzita, 363 – 372.
- TAKÁČ, M. (2006c): Categorization by Sensory-Motor Interaction in Artificial Agents. In: Fum, D. – Del Missier, F. – Stocco, A. (eds.): *Proceedings of the 7th International Conference on Cognitive Modelling*. Trieste: Edizioni Goliardiche, 310 – 315.
- TAKÁČ, M. (2007): Kognitívna sémantika komplexných kategórií založená na rozlišovacích kritériách. In: Kvasnička, V. – Trebatický, P. – Pospíchal, J. – Kelemen, J. (eds.): *Myseľ, inteligencia a život*. Bratislava: Vydavateľstvo STU.
- VAN LAMBALGEN, M. – HAMM, F. (2002): Moschovakis' Notion Of Meaning As Applied To Linguistics. In: Baaz, M. – Krajíček, J. (eds.): *Logic Colloquium '01*. Natick (Mass): A. K. Peters.