

Ukotvení symbolů v kontextu nulového sémantického závazku

Michal Vavrečka

Katedra psychologie, Fakulta sociálních studií Masarykovy univerzity, Joštova 10, 602 00 Brno
vavrecka@fss.muni.cz

Abstrakt

Článek se zabývá přehledem a analýzou současných přístupů k ukotvení symbolů. Kromě historických kořenů problému, jsou probírány jednotlivé části procesu ukotvení, včetně posouzení použitelnosti základních architektur. Hlavní část je věnovaná analýze obecných i konkrétních modelů dle kritéria nulového sémantického závazku a následně kritice tohoto kritéria.

1 Úvod

V současné době patří problematika ukotvení symbolů k velmi diskutované oblasti kognitivních věd, přičemž návrhy způsobů řešení začínají v oblasti filosofie mysli a končí konkrétními návrhy architektur. Jen v oblasti obecného způsobu řešení nalezneme několik návrhů, jak problematiku uchopit. Jelikož některé z nich budou v následujícím textu zmiňovány, uvádím pro přehled názvy přístupů, které jsou s otázkou ukotvení symbolů spojeny:

Argument čínské pokoje (Searle 1980), (Cole, 2004)
Problém ukotvení symbolů (Harnad, 1990)
Problém ukotvení znalostí (Chalmers, 1992)
Perceptuální ukotvení prostorové sémantiky (Regier, 1992)
Problém ukotvení konceptu (Dorffner a Prem, 1993)
Fyzické ukotvení (Brooks, 1993)
Past internalismu (Sharkey a Jackson, 1994)
Problém ukotvení (attachment/tethering) symbolů (Sloman, 2002)
Problém zakotvení (anchoring) symbolů (Coradeschi a Saffioti, 2003)

Bereme-li jako výchozí moment Harnadův článek z roku 1990, můžeme říci, že následná odezva související s rozvojem návrhů ukotvení symbolů byla značná. V tomto článku se pokusím posoudit jednotlivé teorie a nalézt společné body, které mohou následně sloužit jako základní prvky pro návrh systému, blížící se svými sémantickými schopnostmi lidskému kognitivnímu aparátu.

V následujících kapitolách budou nejprve připomenuty kořenové teorie vedoucí k ukotvení symbolů, poté

základní definice *symbol grounding* včetně možnosti posouzení pomocí sémantického kritéria. Pro přehlednost jsou uvedeny jednotlivé směry v oblasti kognitivních věd a jejich přínos při řešení otázky ukotvení symbolů. Po krátké rekapitulaci jednotlivých architektur, použitých při modelování reprezentačních mechanismů, následuje posouzení obecných i konkrétních modelů a jejich hodnocení. Závěr je věnován návrhu vytvoření kritérií pro budoucí modely.

2 Historická východiska

Potřeba řešení otázky ukotvenosti symbolů je spojena zejména s rozvojem klasické počítačové architektury během druhé poloviny minulého století. Následkem tohoto rozvoje je prudký nárůst aplikací umělé inteligence, zaměřené na oblast zpracování symbolů. Ať už hovoříme o silné umělé inteligenci, fyzickém symbolickém systému, či reprezentaci pomocí rámců, ve výsledku dochází ke směřování syntaxe a sémantiky. Lépe řečeno používání syntaktické struktury jako sémantické bez hlubší analýzy problému. Jedná se o celý proud autorů, kteří považují pouze symbolickou rovinu za dostačující pro tvorbu reprezentací (Fodor, Newell, Pylyshyn apod.).

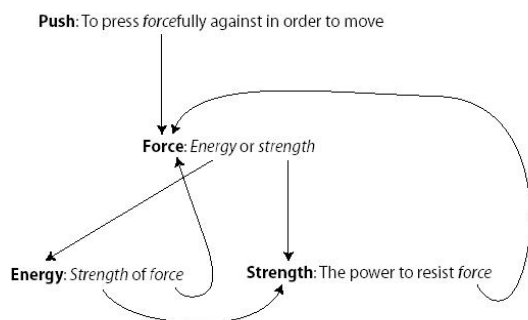
Alternativou k tomuto směru je výzkum Stephena Kosslyna (1980) o povaze mentálních představ. V psychologii velmi známý spor o způsobu mentálního kódování probíhá do dnešních dní (Kosslyn, 2003), (Pylyshyn, 2003), přičemž ani jeden z autorů či stoupenců jednotlivých směrů nebyl schopen nalézt dostatečnou argumentační základnu, popřípadě výstižný experimentální design, pro přesvědčivý důkaz o povaze představ, potažmo mentálních reprezentací. Kosslynova teorie o perceptuální povaze (či perceptuálním podkladu) reprezentací, byla rozvinuta v práci A. Paivia (1986). Autor přichází s teorií duálního kódování. Samotný pojem kódování je zde poněkud nešťastný, jelikož v oblasti kognitivních věd se tento termín spíše spojuje se způsobem elementárních operací jednotlivých architektur než s oblastí reprezentační. Vhodnější je proto hovořit o duálním systému reprezentací – obrazovém a verbálním (propozičním a analogovém, symbolickém a nonsymbolickém, imagenu a logogenu). Autor během dlouholeté experimentální činnosti prokázal, že jednotlivé

systemy vzájemně spolupracují, ale povahu či přesnější formalizaci této součinnosti nikdy nerozvedl. Jsou to právě způsoby propojení dvou rozdílných reprezentačních systémů, jež tvoří teoretické zázemí pro problematiku ukotvení symbolů.

3 Symbol grounding problem

V epistemologické rovině souvisí se zmíněným Searlovým argumentem čínského pokoje, tedy problém vlastní (intrinsic) sémantiky (vlastního významu nebo také intencionality). Namítá se, že pokud je význam symbolů v symbolickém systému nevlastní (oproti vlastnímu významu v našich hlavách), tak to nemůže být přípustný model pro sémantiku podobnou lidské. Kognice nemůže být pouhá manipulace se symboly (Harnad, 1990). Základní otázkou groundingu je, jak vytvořit funkci a interní mechanismus artefaktu (např. jako interní reprezentaci odkazující na externí objekt) tak, aby byla vlastní vzhledem k samotnému artefaktu, místo toho, aby byla závislá na interpretaci externím pozorovatelem (Ziemke, 1999). Searle (1980) a Harnad (1990) zkonstatovali, že čistě tradiční a čistě výpočetní způsob UI, založený na programování, nemůže vést ke vzniku inteligentního systému, protože znalost zůstane pro systém vždy nevlastní, a je znalostí pouze pro externího inteligentního pozorovatele (designéra). Proto je nutné, aby znalost vstupovala do systému z jeho prostředí v ukotvené formě. Otázkami čistě symbolických systémů, či rozdíly mezi designérským a učitelským přístupem se budeme ještě věnovat v dalších částech článku.

Pro vytvoření úvodní metafory snad postačí následující ilustrace. Harnad (1990) popisuje čistě symbolický přístup metaforou učení se čínskému jazyku, jako svému prvnímu (mateřskému), pomocí čínsko-čínského slovníku. Podobně se k problematice vyjadřuje Roy (2004), který problém demonstruje na cyklické slovníkové definici jednotlivých pojmů.



Obr. 1 Sémantika symbolického systému

Toto omezení nás nutí vytvořit systém, který nebude obsahovat pouze symbolickou rovinu. Nové požadavky na konstrukci inteligentních systémů (agentů) leží právě v oblasti propojení symbolického systému s předzpracovanými perceptuálními informacemi. Opuštění "internalistické pasti" je klíčové pro tvorbu inteligentního systému (Law a Miikkulainen, 1994). Stroje musí být *napojeny* na externí svět, například pomocí kauzálních spojů, které nejsou zprostředkované externím pozorovatelem (Sharkey a Jackson, 1996). Obecným modelem pro řešení ukotvení je hybridní systém. Navrhuje jej již Harnad ve svém článku (1990). Jedná se o propojení konekcionistického a symbolického (symbolového) přístupu. Ukotvení probíhá ve třech úrovních:

Ikonizace - převod analogových signálů na ikonickou reprezentaci, která je analogovým ekvivalentem projekce vzdáleného objektu na agentovy senzory.

Diskriminace - posouzení, zda jsou vstupy stejné nebo rozdílné.

Identifikace - proces přiřazení jedinečné odpovědi.

První dvě úrovně jsou subsymbolické, třetí slouží k ukotvení symbolů. Dříve než se pustíme do bližšího popisu a analýzy tohoto systému, bude nutné nadefinovat podmínky a kritéria, která nám poslouží k posouzení, zda lze navrhovaný systém považovat za schopný ukotvení a zda lze takový systém považovat za ukotvený.

4 Sémantické kritérium

Jako velmi slibný se jeví návrh autorů Tadeo a Floridi (2005), kteří přicházejí s podmínkou nulového sémantického závazku. Podmínku dále nazývají *Z kritériem* a snaží se jeho pomocí posoudit současné systémy, které mají ambice být řešením symbol grounding problému. Autoři vycházejí z Harnadovy práce, tzn. že interpretace symbolů musí být vlastní vzhledem k systému (nemůže být nevlastní nebo parazitická). Pro návrh a tvorbu systému z toho plynou následující omezení (Tadeo, 2005):

1. *Není povolena žádná forma vrozenosti, neexistují sémantické zdroje (virtus semantica), které bychom považovali jako předinstalované do autonomního agenta.*

2. *Není povolena žádná forma externalismu - významy nelze do systému vkládat zvenčí pomocí designéra, který má sémantickou znalost.*

Tyto dva body nemají být v rozporu se třetím:

3. *Agent má zdroje a schopnosti (výpočetní, syntaktické, procedurální, perceptuální, schopnosti učení) ukotvit symboly.*

Autorům se během analýzy současných přístupů podařilo prokázat, že ani po patnácti letech (1990-2005) neexistuje systém, který by byl schopen úspěšně vyřešit problém ukotvení symbolů. Jelikož je jejich způsob aplikace *Z kritéria* u některých systémů nepřesný, pokusím se s těmito tvrzeními ve svém článku polemizovat, popřípadě nalézt alternativní způsob interpretace.

5 Teoretická východiska

Protože v oblasti kognitivních věd neexistuje jednota přístupů ke konstrukci umělých inteligentních systémů, a jednotlivé proudy používají rozličná teoretická východiska, bude nutné ještě před samotnou analýzou provést diferenciaci přístupů. Můžeme je rozdělit podle několika kritérií. Z hlediska interpretace a argumentace jsou pro mne důležité následující způsob rozlišení:

Existence reprezentací
Reprezentacionismus
Nonreprezentacionismus

Obecný přístup
Kognitivismus
Konekcionismus
Enaktivní přístup

Způsob tvorby reprezentací
Designérský
Učitel'ský

5.1.1 Reprezentacionismus

Reprezentační přístup vychází z předpokladu, že symboly jsou ukotveny v reprezentacích, které vznikají na základě manipulace systému (agenta) s perceptuálními daty. Obecný postup (Tadeo, 2005) je následující:

- a) *Zachytit podstatné znaky, které jsou společné perceptuálním datům*
- b) *Abstrahovat*
- c) *Identifikovat abstrakci jako obsah kategoriální a konceptuální reprezentace*
- d) *Použít takové reprezentace k ukotvení symbolů*

Již v tomto základním schématu vidí autoři *Z kritéria* obtíže. Hlavním problémem přístupu je použití reprezentací jako tvrzení bez důkazu či jako definicí kruhem („at the price of begging the question“). Sémantické schopnosti jsou podle nich (Tadeo, 2005) dopředu předpokládány, místo aby docházelo k jejich tvorbě. Takto postavené tvrzení je ale na hranici argumentovatelnosti. Pokud chceme, aby nějaký systém

měl sémantické vlastnosti, nemůže pak konkrétní systém považovat za špatný, jelikož jej designér dělá s určitým záměrem. Problematickou pasáží se jeví potřeba tvorby symbolické úrovně (pro následné manipulace s reprezentacemi), jelikož navrhuje systém, který této úrovni musí dosáhnout. Pokud systém integruje (ukotvuje) symboly s perceptuálními kategoriemi na základě učení, nelze jej hodnotit jako neuspokojivý vzhledem k *Z kritériu*. Sporné místo se pravděpodobněji nachází ve vymezení oblastí, které externí pozorovatel (designér) ještě může interpretovat či facilitovat, což se týká způsobu tvorby reprezentací zmiňované níže.

5.1.2 Nonreprezentacionismus

Základem jsou Brooksovy práce z oblasti reaktivní robotiky, vtělenosti a orientovanosti. Podle známého hesla „Svět je svou nejlepší reprezentací“ nejsou reprezentace v takto navrženém systému nutné. V takovém případě nepotřebujeme řešit otázku ukotvení symbolů. Přesto při popisu robotických agentů hovoří Brooks o tom, že obsahují určitý stupeň "fyzického ukotvení", kvůli tomu, že jsou propojeni s prostředím pomocí senzorů a efektorů (Brooks, 1993). To je ale klamné, protože nedochází k ukotvení chování ani interních mechanismů. (Ziemke, 1999). Ukotvení vnitřních mechanismů není možné, jelikož jsou konstruovány jako vzájemně kooperující (z hlediska subsumpce lépe „potlačující se“) deterministické finite state automata a fungují na základě mechanismů, které jsou předprogramované. Nelze je tedy považovat za ukotvené. Jediným emergentním jevem, jenž bychom mohli považovat za vlastní systému, je způsob spolupráce jednotlivých modulů. Hlavní nedostatek, týkající se neschopnosti takových systémů provádět operace, které jsou založené na jeho předchozích zkušenostech, je zmíněn v následující pasáži v souvislosti s enaktivním přístupem.

5.2 Obecné přístupy

Přístupy k ukotvení symbolů se dají rozdělit na kognitivistické, konekcionistické, enaktivní a popř. hybridní. Kognitivistické paradigma v zásadě rozlišuje kognici a percepce. Příkladem může být Fodorovo rozlišení na vstupní systémy (*low-level* percepce) a centrální systémy (myšlení, řešení problémů). Kognitivisté se většinou soustředí na vstupní systém a ukotvení atomárních reprezentací pomocí senzorických nebo senzomotorických invariant. To znamená, že kauzální spojení mezi agentem a prostředím jsou tvořena pomocí propojení atomárních interních reprezentací s externími entitami či objektovými kategoriemi. Takto ukotvené

atomy jsou následně použity jako stavební kameny pro tvorbu komplexních reprezentací, ze kterých je konstruován koherentní reprezentační model světa (Ziemke, 1999). Enaktivní přístup se od kognitivního odlišuje nepřítomností centrálních systémů a také jinou formou reprezentací. Podle některých autorů může tato absence klasické formy reprezentací vypadat jako zjednodušování procesu ukotvení. Podle enaktivního přístupu je ale znalost nazírána jako distribuovaně vtělená (rozprostřená a vtělená v senzorech, efektorech, nervovém systému apod.) a částečně také ležící v prostředí (Varela, 1991).

Enaktivní přístup je ale přesto problematický.

Vynecháním centrálních systémů se vlastně elegantně vyhneme problematice manipulace s reprezentacemi. Pokud budeme mapovat senzoričné systémy přímo do motorických, jakkoliv za použití paralelního zpracovávání, jedná se o krok zpátky. Takový systém je paralelou s behavioristickým pohledem či Brooksovou reaktivní robotikou. Zjednodušený systém nemůže dosáhnout vyšších kognitivních schopností než hmyz. Může sice docházet k jisté formě kompozicionality, ale pokud není předchozí zkušenost uložena ve formě manipulovatelné reprezentace, nemůže takový systém používat vyšší kognitivní funkce. Na druhou stranu je přínosem enaktivního přístupu důraz na kontextualitu, paralelní zpracování a interakci agenta z prostředím. Jak kognitivistický tak enaktivní přístup se shodují v tvrzení, navrženém již Searlem (1980), že jsou vlastnostmi strojů (vtělených systémů, které jsou kauzálně spojené se svým prostředím) než nevtělených programů.

Oba přístupy vyžadují plně ukotvené systémy. V kognitivistickém požadujeme ukotvenost propojení vstupních a centrálních systémů. Enaktivní přístup vyžaduje vývoj celku v interakci s prostředím.

5.3 Způsob tvorby reprezentací

Rozlišení způsobů tvorby komponent, které jsou pro systém předprogramované a těch, které systém může sám vytvořit a používat, je otázkou častých sporů v oblasti kognitivních věd. Obecně lze říci, že se jedná o nalezení takového poměru mezi designérským a učitelským přístupem, abychom se ve výsledku dostali k systému, který je schopen smysluplně manipulovat se svými reprezentacemi, ale i modifikovat některé své mechanismy na základě a v závislosti na učení. Samotné *Z kritérium* je ve své podstatě omezováním designérského přístupu až na nejmenší možnou míru.

Stručně lze charakterizovat designérský přístup jako tvorbu agentovy funkce a učitelský přístup je pak založen na agentovi, který se svou funkcí učí. O tom, že učitelský přístup se dá nahlížet různě, bylo již zmíněno

v souvislosti se *Z kritériem* výše. Tadeo (2005) hovoří o nemožnosti procesu učení se symbolům za pomoci externího pozorovatele (učitele). Externí pozorovatel ale nesmí zasahovat pouze do tvorby konceptů. V (čistě) symbolické úrovni již může fungovat jako učitel, protože bez jeho účasti není symbolická rovina přítomna (ve smyslu designérové produkce symbolů). Pokud bychom přesto vyžadovali striktní nepřítomnost sémanticky vybaveného učitele, lze takovému požadavku vyhovět. Alternativou je získávání symbolů z knih či jeho média. Během takového učení však docházíme k paradoxu učení se symbolického systému bez znalosti k čemu přísluší. Kompromisem je návrh systému, který se naučí manipulovat se svým symbolickým systémem (např. přirozeným jazykem), na základě asociálního učení podobného učení se ze slabikáře (asociace obrazu se slovem). V takovém případě je ukotvování symbolů podobné učení se hluchoněmých dětí psanému a čtenému projevu, aniž mají předchozí zkušenost se symbolickým systémem. Většina dětí se ale naučí znakovou řeč jako první, takže se nám nabízí pouze úzké pole působnosti při získávání podrobnějších a přesnějších informací ohledně této problematiky pomocí psychologických experimentů. V oblasti kognitivních věd je míra přítomnosti učitele a omezení jeho působení v procesu učení předmětem diskuse.

6 Procesy ukotvení

Pro lepší pochopení mechanismů podílejících se na ukotvování symbolů se detailněji podíváme na jednotlivé procesy, o kterých jsme v souvislosti s tímto tématem hovořili.

6.1 Kategorizace

Kategorizace je prvním krokem a její výsledek tvoří podklad pro následné mapování do symbolické roviny reprezentace. Pro studium nám nejlépe poslouží lidský kognitivní aparát. Lidé jsou schopni diskriminovat, manipulovat, identifikovat a popisovat objekty, události a stavy věcí a mohou produkovat deskripce a odpovídat na deskripce objektů, událostí a stavu věcí (Harnad, 1990). Cílem kognitivní teorie je zjistit, jak to lidé dokáží. Podívejme se tedy na proces kategorizace detailněji.

Diskriminace je posouzení, zda jsou dva vstupy stejné nebo rozdílné, popřípadě jak jsou rozdílné. Jedná se o relativní posouzení tzn. hledání míry shody. Identifikace je přiřazení arbitrární odpovědi - "jména" - klasifikující vstupu, které následně tvoří kategorii podle shodného prvku. Identifikace je absolutní posouzení, které rozhodne, zda je vstup členem kategorie či nikoli (Harnad, 1990). Schopnost diskriminace je zajištěna

pomocí *ikonických reprezentací* (Harnad, 1987). Diskriminace je tedy nezávislá na kategorizaci, jelikož mohou diskriminovat, aniž bych věděl, co dané objekty znamenají. To platí ale pouze v případě, že předpokládáme jednosměrný proces (pasivní) při zpracovávání a kategorizaci podnětů (nebo v případě, že je podnět prezentován poprvé). Pokud předpokládáme aktivní anticipující subjekt, který diskriminuje na základě předchozí zkušenosti (a v případě existence apriorních schémat vnímání je tomu tak vždy), nemůžeme hovořit o nezávislosti diskriminace na identifikaci. Proto nelze Harnadovo tvrzení zobecňovat na komplexnější modely. Objevují se i otázky, týkající se procesu kategorizace jako celku. Může nám jediná kategorizace, popřípadě kategorizace podle jednoho kritéria sloužit k vytvoření dostatečně robustní reprezentaci entity ve světě? Dorffner (1996) navrhuje ve svém konekcionistickém modelu ukotvení, na základě rozšíření kategorizačních vrstev a jejich následné zpětné propojení, čímž bude docházet ke kategorizaci podle více kritérií, popřípadě k tvorbě kategorií abstraktních. V nonsymbolické úrovni (bez identifikace se symboly) je možné s výslednou kategorií dále pracovat a znovu na ní aplikovat mechanismy, které umožní její diskriminaci podle více atributů, a následnou identifikaci s podobnou přesností. Problematické se jeví vytvoření dostatečně důmyslných mechanismů víceúrovňové, paralelní, heterarchické a zpětnovazebné kategorizace, které dokáží vytvořit robustní reprezentace, jenž jsou dostatečně kontextově ukotvené, pro následnou tvorbu (mentálních) simulátorů. V teoretické rovině o nich hovoří Barsalou (1999) a jeví se jako perspektivní způsob reprezentačního systému.

6.2 Symbolický systém

Další oblastí, která se v souvislosti s ukotvením objevuje, je symbolická vrstva mentální reprezentace. Před samotným definováním symbolického systému, je nutné provést odlišení symbolů od konceptů.

Koncepty jsou kategorické mentální stavy, tvořené na základě adaptivního kategorizačního procesu. Symboly jsou označení nebo znaky používané agentem, které referují ke konceptu. Dále hovoříme o symbolu z hlediska semioze (jeho odlišení od ikónu a indexu). Symbol jako znak je diskrétní, arbitrární a referuje k objektu nebo stavu světa (Dorffner, 1993). Od symbolu se přesuneme k symbolickému systému. Velmi přesně se daří definovat symbolický systém Harnadovi (1990). Symbolický systém je:

1. Soubor arbitrárních "fyzických znaků", které jsou zaznamenány na papíře, pásce, či jiném médiu, se kterými je

- 2. manipulováno na základě "explicitních pravidel", která jsou*
- 3. podobné znakům, či řetězcům znaků. Pravidly řízená symbolická manipulace je zaměřena*
- 4. pouze na tvar symbolických znaků (ne jejich význam), tzn., je čistě syntaktická a obsahuje*
- 5. "pravidly řízené kombinace" a rekombinace symbolických znaků. Existují*
- 6. primitivní atomické symbolové znaky a*
- 7. kombinované (složené) řetězce. Celý systém a všechny jeho části - atomické znaky, složené znaky, syntaktické manipulace, aktuální i možná pravidla jsou*
- 8. "sémanticky interpretovatelné": Syntax může být systematicky přiřazován významu, např. zastupovat objekty, jevy atp.*

„Sémantická interpretovatelnost, aby byla symbolická, musí být propojena s explicitní reprezentací (2), syntaktickou manipulovatelností (4) a systematicitou (8). Žádné z těchto kritérií není arbitrární, a pokud je zlehčíme, ztratíme možnost pracovat s pojmy jako přírodní kategorie a také možnost operovat na základě formální teorie výpočtu, čímž se stává ze slova "symbolický" nevysvětlená metafora (která se liší interindividuálně). Proto budeme přemýšlet pouze o klasickém formálním systému (tak jak je definován), chceme-li ukotvovat symboly (Harnad, 1990).“ Tuto pasáž lze považovat za klíčovou. Hovoří o tom, že neexistuje alternativa k formálnímu systému, kterou bychom mohli použít jako podklad pro myšlenkové operace. Přesto Paivio (1986) explicitně popírá model lidské psychiky jako počítačového modelu a staví svou teorii jako alternativu k tomuto přístupu. Což je ale nedostačující k tomu, aby jeho teorie sloužila jako argument proti počítačovému pohledu na mysl (Thomas, 1999). To nás utvrzuje v domněnce, že symbolický systém a otázky ukotvení symbolů se jeví jako perspektivní způsob řešení povahy mentálních reprezentací.

6.3 Ukotvení interních mechanismů

Jak již bylo zmíněno výše, je pro ukotvený systém důležité, aby obsahoval nejen vlastní reprezentace, ale aby byl vybaven mechanismy, jejichž užívání je pro systém také vlastní. Tato problematika má paralelu například v logice, ve které je obtížné nalézt způsob, jak sémanticky vyjádřit své logické operátory. Příkladem může být pravidlo IF..THEN. Způsob jeho ukotvení v systému není zdaleka triviální otázkou. Wittgenstein (1953) zdůraznil rozdíl mezi explicitními a implicitními pravidly: je odlišné se pravidlem *řídít* (explicitně) a nebo jednat *podle pravidla* (implicitně). Harnad (1990) vysvětluje rozdíl mezi explicitním pravidlem jako

symbolickým vyjádřením, oproti *zadržovaným omezením* uvnitř systému. Takový rozdíl poukazuje na funkci a její jedinou implementaci v určitém systému, ale nerozlišuje mezi explicitním a implicitním. Rozdíl by měl být založen na zdůraznění pravidel, které systém používá. Pokud jsou pravidla vlastní (implicitní), dokáže je používat smysluplně a kontextuálně, narozdíl od explicitních pravidel, se kterými pracuje pouze jako *searlovský manipulátor*. V úrovni reflexu a nižších kognitivních funkcí (u člověka) nenarazíme na takové rozlišení, jelikož aplikace pravidel u reflexů či pudů je deterministická. Pokud je podnět příliš silný, reagujeme reflexivně a většinou stejně. Aplikace pravidla je ve své explicitní, či implicitní formě stejná, což z této úrovně činí dobrou oblast pro ukotvení atomárních reprezentací. U vyšších kognitivních funkcí už hovoříme o rozdělení na implicitní a explicitní, jelikož se jedná o manipulaci s reprezentacemi (ne s pouhými počítky), které musí být systémem vlastní a kontextuální. Taková implicitní manipulace s reprezentacemi by měla být založená na ukotvenosti obsahu, se kterým je manipulováno (kontext, nondeterminismus). Rozdíl lze demonstrovat na příkladu. Explicitní následování pravidel (např. práce u pásu první den bez znalosti jakýchkoliv podrobností (jako *searlovský manipulátor*) nevyžaduje kontext ani nondeterminismus. U člověka by ale následně došlo k převodu pravidla do implicitní podoby i pokud bychom si k činnosti měli vymyslet kontext, který je subjektivní a nekoresponduje s realitou. Vysvětlení můžeme hledat v klasickém chápání pravidla jako sériové sekvence, přičemž právě sériovost, abstraktnost a akontextuálnost (díky arbitrarnosti symbolů a také univerzálnosti pravidel) vede k neschopnosti převodu takto konstruované reprezentace či pravidla do implicitní podoby. Sériovost může být problematická pro vznik či začlenění do kontextu. Lidská implicitnost pravidel souvisí obecně z robustnosti reprezentace. Požadavek by proto měl být přenesen i na robustnost pravidel. Pro nedostatek teoretických podkladů však nebudu otázkou robustnosti pravidel v tomto článku dále rozvíjet a nechávám ji otevřenou pro další diskusi.

Harnad (1990) považuje za dostačující, pokud je explicitně reprezentované symbolické pravidlo částí formálního systému, je dekomponovatelné (až na úroveň primitiv), jeho aplikace a manipulace je čistě formální (syntaktická, závislá na tvaru) a celý systém musí být sémanticky interpretovatelný.

7 Konekcionalismus

Formálními principy, na kterých jsou založené neuronové sítě jsou hlavně asociační a statistické struktury kauzálních vztahů v daných dynamických systémech.

Podle Harnada (1990) jsou neuronové sítě dobrým podkladem pro implementaci dynamických systémů. Existují určité pochybnosti o tom, zda může být konekcionalismus symbolický. Argumenty proti možnostem neuronových sítí v oblasti práce se symboly podali Fodor a Pylyshyn (1988) Podle nich sítě, jako symbolické systémy, nesplňují požadavek složenosti (7) a systematicity (8). Neuronové sítě nepracují na základě formální syntaxe, takže není možná systematická sémantická interpretace.

Přesto už v roce 1985 Touretzky a Hinton vytvořili model, který měl ukázat schopnosti konekcionalistické architektury v oblasti zpracování symbolů. Jejich *konekcionalistický produkční systém* umožňoval reprezentovat a manipulovat s arbitrarním, diskrétními *tokens*. Přestože síť dokáže pracovat s kontinuálním stavovým prostorem (analogově), smysluplné jsou pouze diskrétní *tokens* (Touretzky and Hinton, 1985). Autoři museli takovou síť vytvořit pomocí speciálního nastavení vah a prahů. Mělo jít o důkaz, že lze pomocí neuronových sítí operovat se symboly či na symbolech. Z čehož vyvstává otázka.

Je subsymbolický přístup pouze alternativní implementací symbolického systému?

Objevují se názory, pomíjející rozdíl mezi symbolickým a subsymbolickým, jelikož každá neuronová síť může být implementována do Turingova stroje, který je již z definice symbolický a také lze vytvořit neuronovou síť napodobující svou činností Turingův stroj.

Ohledně omezení neuronových sítí se objevují i jiné námitky. Davidsson vidí tvorbu konceptu jako stupňovitý proces opakované interakce s prostředím. Inkrementální učení kategorizuje data do konceptů. Neuronové sítě jsou sice dobré pro diskriminační učení, nejsou dobré pro inkrementální adaptaci, protože mají "fixní strukturu" (Davidsson, 1993, s. 160) Proto navrhuje použít metody strojového učení.

Toto hledisko poněkud popírá teoretická východiska jednotlivých architektur. Obě obsahují implicitní předpoklad „univerzální“ zpracovatelnosti (jsou schopné řešit libovolnou úlohu v rámci kritérii, které jdou za rámec tohoto článku). Je však obecně známo, že jsou to právě neuronové sítě, které mají podobně jako klasické počítačové zařízení fixní strukturu, ale jsou vybaveny redundantním počtem propojení, jejichž pomocí můžeme aplikovat způsoby řešení úloh, které u klasického strojového učení nenacházíme.

Nutno podotknout, že se rozdíly ve způsobu práce jednotlivých architektur objevují již v člancích průkopníků konekcionalismu. V souvislosti s otázkami kontextu bývá často citován výrok:

„V symbolickém přístupu je kontext symbolů manifestován okolo a sestává z dalších symbolů. V

subsymbolickém paradigmatu je kontext manifestován uvnitř a skládá se ze subsymbolů (Smolensky, 1988,s.17).“

Clark (1993) přichází s otázkou, zda je konekcionistický přístup kontextově závislý ve všech úrovních, nebo jsou jeho systémy v mikroúrovni kontextově nezávislémi nositeli obsahu? Existuje kontextově nezávislá úroveň reprezentace? Jako kontextově nezávislý atom považujeme syntaktickou položku (individuovanou nesémanticky), která hraje fixní reprezentační roli v symbolickém systému (Clark, 1993). Autor následně argumentuje, že problém kontextu je v konekcionismu chybně položen. Je nazírán z hlediska symbolického systému (z hlediska syntaxe), ale nic podobného v neuronových sítích neexistuje. Právě pro tyto nesrovnalosti, zůstávají otázky povahy neuronových sítí, jejich použitelnosti a také rozdílnosti oproti klasické počítačové architektuře, stále v rovině diskuzí.

8 Obecné způsoby řešení

Nyní se podíváme na samotné způsoby ukotvení symbolů. Mezi základní otázky, objevující se v Harnadově článku (1990) patří následující: Jak učinit sémantickou interpretaci formálního symbolického systému vlastní systému? Jak může být význam bezvýznamných symbolických prvků, se kterými je manipulováno pouze na základě jejich (arbitrárních) tvarů, ukotven pouze v dalších bezsmyslných symbolech? (viz. Obr.1) Postup, který má tyto problémy řešit, je popsán následujícím způsobem. Symboly musí být ukotveny na základě (Harnad, 1990):

- a) *Ikonických reprezentací* - vznikají převodem vzdálených objektů či událostí na blízké senzorní projekce.
- b) *Kategoriálních reprezentací* – vznikají na základě vrozených a naučených mechanismů a dokáží ze senzorních projekcí abstrahovat společné vlastnosti objektů či událostí - kategorie. Symboly jsou následně názvy, přiřazené těmto objektovým nebo událostním kategoriím, které vznikly nonsymbolicky.

Takto navržený systém musí být hybridní, zčásti konekcionistický a zčásti symbolický. Systém neobsahuje pouze symbolický modul, protože symboly jsou ukotvené pomocí kategoriálních reprezentací a tedy vlastní pro daný systém (Harnad, 1990). Někteří autoři navrhuji extrémnější přístup, nazývaný radikální konekcionismus, založený na použití neuronových sítí v celém systému

(Dorffner, 1993). Radikální konekcionismus je cestou ke tvorbě komplexních konceptuálních struktur, ale není tak systematický, jako kombinování externích symbolických znaků. Tento přístup je subkonceptuální, jelikož stavební bloky leží o úroveň níže a také nonsymbolický, protože stavební prvky nejsou arbitrární. *Symbol grounding* se tak stává *concept groundingem* - modelováním formací konceptů ukotvených ve zkušenosti (Dorffner, 1993). Takové vysvětlení nám však neposkytuje odpověď na Harnadovu otázku, pouze přikládá důležitost formování konceptů, tvorbě konceptových formací a nechává otázku následné manipulace v symbolické úrovni nevyřešenou. Objevují se i přístupy opačné. Rozšíření a doplnění hypotézy ukotvení symbolů následovalo v letech 2000-2002, opírajíc se o články Harnada a Cangelosiho (2001) a Cangelosiho (2000,2002). Přicházejí s hypotézou *symbolické krádeže*.

Jedná se o mechanismus transformace kategoriální percepce do ukotvených *low-level* označení a následně do *high-level* symbolů. Proces *grounding transfer* slouží k akvizici nových symbolů na základě kombinace starých, již ukotvených, přičemž může být implementován do neuronových sítí.

Takto budovaný systém je pokusem o hierarchickou stavbu základních atomů do vyšších celků, přičemž atomy jsou dostatečně elementární (a zde také ukotvené), aby se staly univerzálními prvky pro následnou tvorbu komplexních reprezentací a pro efektivní operace s těmito reprezentacemi. Ve své podstatě se jedná o kompozicionalitu (schopnost skládat prvky), což je základní vlastnost symbolického systému. *Symbolická krádež* si jí vypůjčuje do *nižší úrovně*. Ve svém původním článku (1990) Harnad ještě považuje tuto schopnost za vlastní pouze symbolickému systému. V Cangelosiho přístupu nacházíme snahu o rozlišení způsobu řešení symbol groundingu. Pokouší se nalézt odpověď na otázku, který subsystém by měl být během groundingu převažující. Zda použít *senzomotorickou dřinu*, tedy velmi robustní mechanismy práce se senzorními vstupy a předzpracování na mnoha úrovních (což je výpočetně náročné), nebo použít systém, ve kterém jsou ukotvené symbolické atomy následně využívané symbolickým systémem, jehož činnost je daleko méně náročná, transparentní a teoreticky zakotvena. Cangelosi se v podstatě snaží určit, ve které části převodu senzorních vstupů použít výhody symbolického systému. Sám ve svém modelu navrhuje větší využití symbolického systému a ukotvení již ve *základní úrovni*.

8.1 Analýza přístupu

Podle Tadeo (2005) nesplňuje hybridní forma systému *Z kritérium* (nulový sémantický závazek). Problém spočívá

ve způsobu, kterým hybridní systém nachází neměnné vlastnosti senzorických projekcí, umožňujících kategorizaci a identifikaci objektu. Problematickou se jeví neschopnost takových systému dojít k označení třídy (například zvíře) bez předchozí znalosti. Pokud se jedná o učení s učitelem (*supervised*), dodá označení třídy designér. Pokud ale probíhá učení bez učitele (*unsupervised*), není jasné, jak vznikne symbolické označení třídy. Lze jej vysvětlit pouze sociálním kontaktem a možnostmi během komunikace. Což v důsledku znamená, že by takové vymezení *Z kritéria* neumožnilo jeho splnění žádnému systému (včetně člověka). Jedná se o podobný problém, se kterým jsme se setkali už v části věnované reprezentacionismu. Pokud hovoříme o vzniku symbolické roviny, musí mít tato rovina nějaký účel. A tou je zejména intersubjektivní komunikace. Jestliže neexistuje možnost sociální komunikace, je zbytečná symbolická úroveň, jelikož nenachází v systému opodstatnění a nemůže ani dojít k jejímu vzniku. Požadavek autorů je paradoxní: vyžadují emergenci symbolické roviny, přičemž se ale v prostředí, ve kterém se agent pohybuje, žádná forma symbolů neobjevuje. V této interpretaci nemá symbolická úroveň možnost vzniknout. Při vzniku kategorií, během učení bez učitele, se jedná o podobný problém. Neexistuje-li komunikace, probíhá kategorizace na základě diskriminace a identifikace (způsob tvorby kategorií je subjektivní), ale nemůže dojít k integraci se symbolickou vrstvou reprezentace, jelikož není přítomna. Pro Tadeo (2005) je také problematické vysvětlit, jak systém pozná, na které vlastnosti či vzorce se má zaměřit, které má považovat za invariantní. Taková možnost musí být dopředu napevno zakódovaná, což vyžaduje účast designéra. Navržený postup je podle autorů *Z kritéria* nevlastní.

S problémem týkajícím se schémat použitých při vnímání a následné kategorizaci se setkáváme již v oblasti filosofie mysli. Dotýká se ještě obecnějších témat než problém ukotvení a nazývá se konceptový empirismus. Jedná se o aplikaci klasického empirismu do oblasti tvorby konceptů, které jsou tvořeny na základě interakce s prostředím. Ale pokud má systém získávat své koncepty pouze empiricky, musíme předpokládat apriorní schémata pro jejich tvorbu. Na základě tohoto zjištění, vyvrátil Kant (2001) teorii konceptového empirismu. Nelze získat koncepty pouze ze zkušenosti, protože mít zkušenost znamená použití konceptů (např. koncepty časoprostorých vztahů nebo koncept příčinnosti). Proto musejí být koncepty vrozené. Chomsky (1960) došel ke stejnému zjištění v oblasti učení se jazyka, přestože se detaily jeho teorie později změnilly. Harnadův článek je ale pouze obecným přístupem k problematice. Zaměřuje se na jednoduchou formu kategorizace a následného ukotvení. Je samozřejmé, že

pokud se vydáme směrem k systémům, které se pohybují v dynamickém a proměnlivém prostředí, je nutné vybavit základní systém dalšími komponentami, jako je pozornost či krátkodobá paměť. Psychologická zkoumání poukazují na to, že tyto systémy jsou z části *napevno zadrátovány*, ale že dochází zároveň k jejich modifikaci v závislosti na vnímaném podnětu. Rozšířením Harnadova modelu o zmíněné komponenty získáváme systém, který je schopen pomocí apriorních schémat vytvořit prvotní kategorie a koncepty, na jejichž základě je schopen ukotvovat další koncepty již autonomně. Detailnější popis je nad rámec tohoto článku.

Pokud shrneme uvedené argumenty, lze říci, že požadavek nulové účasti designéra při tvorbě systému by znamenal přílišnou restriktivitu. Ve svém důsledku popírá existenci vrozených (v případě umělého systému předdefinovaných) mechanismů, např. že se dítě rodí bez reflexivní výbavy či implicitních protomechanismů vnímání. Striktní aplikace kritéria vlastně popírá jeho třetí dodatek, který umožňuje systému mít zdroje (výpočetní, syntaktické, procedurální, perceptuální, učení) být schopen ukotvit symboly. Jelikož tyto zdroje nelze v případě umělých systémů získat dědičně, nemůžeme úlohu designéra zcela eliminovat.

8.2 Konkrétní modely

Pro lepší ilustraci uvádím několik konkrétních návrhů řešení *symbol grounding*, obsahující následnou analýzu daného přístupu. Jedná se o rozšířenou polemiku nad interpretací *Z kritéria*.

Cangelosi (2000) vytvořil neuronovou síť, která měla jako vstupy umělou retinu a lingvistický vstup. Na vstup prezentoval základní geometrické útvary a jejich názvy. Proces tréninku byl následující:

1) Síť vytváří prototypy základních geometrických útvarů s pomocí externího agenta se sémantickou znalostí (člověka).

Podle Tadeo (2005) dochází tímto krokem k porušení sémantického kritéria. Pokud ale autoři chápou kritérium ve verzi, kterou prezentují, je nutné zopakovat, že v takovém případě nedojde k ukotvení symbolů (nebo k možnosti užívat symbolický systém) ani u člověka. Důkazem jsou například kasuistiky dětí vychovávaných zvířaty (Singh, 1966), které se nebyly schopny po návratu k civilizaci naučit klasickou formu řeči, a jejich myšlení zůstávalo na prelingvistické (lze říci i subsymbolické) úrovni. Že je problematika vzniku řeči ještě omezenější demonstrují teorie, které hovoří o kritických obdobích ve vývoji jedince (Piaget, 1999), při jejichž zanedbání dochází k celoživotní neschopnosti učení se přirozenému jazyku. Tím se ale dostáváme nad rámec tohoto článku. Pro úplnost ještě uvádím následné fáze tvorby Cangelosiho modelu.

2) Dále jsou asociovány tvary se jmény a dochází k ukotvení symbolů perceptuálně.
3) Probíhá učení se kombinacím ukotvených jmen tvarů s novými arbitrárními jmény (symetrický x asymetrický). Je použito imitační učení pro tvorbu kombinací jmen (např. čtverec je symetrický). Tak dochází k *transferu groundingu* z již ukotvených jmen (čtverec) na nové (symetrický). Nové jména jsou však ukotvená pouze pro jeden prototyp. Pokud by vlastnost symetrický platila pro ukotvený symbol s jinými vlastnosti, systém by si s tím neporadil.

Jiným příkladem konkrétní aplikace v oblasti *symbol groundingu* je Regierova *perceptuálně ukotvená sémantika*. Regier (1992) trénoval neuronové sítě v označování dvourozměrných scén, obsahujících základnu (např. subjekt) a objekt, u něhož byl popsán prostorový vztah vzhledem k základně (na, pod, nad apod.) Podle Regiera se model naučil *perceptuálně ukotvenou sémantiku*. Autor nazývá svůj systém *preadaptovaným strukturálním nástrojem*, který plní podobné úkoly jako malé dítě v průběhu učení se lexikální sémantice prostorových vztahů. Odlišný názor má Ziemke (1999):

„Hlavní rozdíl je v tom, že podobný subsystém u lidí byl vytvořen a testován během evoluce, a proto můžeme říci, že je vlastní lidskému genotypu, a následně individuum (fenotyp) je jeho instanciací. Taková preadaptace je odlišná od Regierovy ... Shrnutí, lze považovat Regierovy transduktéry za umělé srdce. Je vlastní celému systému (nabízí stejnou funkcionalitu jako umělé), ale samo sobě nikdy vlastní být nemůže.“

V tomto tvrzení můžeme identifikovat chyby v interpretaci u Tadeo i u Ziemkeho. Pokud je u lidského mozku jeho funkčnost při zpracování tvořena a testována evolucí, nelze mít podobné požadavky na umělý systém. Znamenalo by to čekání, až se metodou pokusů a omylů vytvoří kognitivní aparát podobný lidskému. Čímž ale rezignují na konstruktérský přístup, pomocí kterého je kognitivní aparát modelován na základě částečných znalostí o jeho struktuře a funkci. Konstruktér nemá možnost, podobně jako evoluce, vyvíjet v nekonečném počtu kroků pomocí jemných zpětnovazebních modifikací. Takový proces je z hlediska efektivity mimo designérskou oblast zájmu. Ještě přesněji jsou chyby v nazírání problematiky u Ziemkeho patrné v pasáži o umělém srdci. Zde již autor dochází na samotnou hranici argumentace při hledání rozdílu mezi přirozeným a umělým. Zjednodušeně lze jeho názor parafrázovat následovně - umělý systém nemůže být nikdy podobný lidskému protože je umělý.

Ziemke ve své argumentaci ale pokračuje dále. Pokud bychom připustili, že (kognitivistický) centrální systém obsahuje proces transdukce vlastní pro celý systém

(skládajícího se z centrálního systému a vstupního transduktivního systému), můžeme poté hovořit o plném ukotvení systému? Bohužel, opět ne, jelikož Regierovy transduktéry (jejich struktura, organizace, vnitřní mechanismy), tedy vše kromě vah neuronových sítí, jsou ukotveny pouze do Regierova výzkumného designu (přestože je psychologicky a biologicky plausibilní) (Ziemke, 1999). Jedná se stejný požadavek jako u Tadeo, tedy aby systém byl vytvořen na zelené louce, ale přitom do něj nikdo nezasahoval. Tento extrémní požadavek *groundingu* hraničí s nemožným. Navrhovaný systém je neuskutečnitelný, jelikož se setkáváme s následujícími omezeními při tvorbě systému:

a) *Reprezentace prostředí vzniká bez interpretace zvnějšku.*

b) *Vnitřní mechanismy, které reprezentaci tvoří, nejsou navrženy, ale samy se ustaví.*

Pokud bychom aplikovali takové postupy na normální vývoj dítěte, nedošlo by u něj k *symbol groundingu*, jelikož by mu symboly nikdo neinterpretoval, a muselo by se pouze spoléhat na to, že se mu pomocí asociálních principů spojí symboly s kategoriemi. K takové situaci ale nemusí dojít. Extrémní požadavky na ukotvení systému jdou za hranice, kterých dosahuje člověk v této oblasti, přičemž jej (přesněji jeho kognitivní schopnosti) považujeme při návrzích systémů za základní měřítko úspěšnosti.

9 Shrnutí

Zaměříme-li se na plně ukotvené systémy, je nutné, aby každý aspekt či interní mechanismus byl vlastní celému systému (celku), musí se tento celek vytvořit během interakce se svým prostředím.

UI a kognitivní věda se zaměřila na modelování inteligentního chování použitím *high-level* abstrakcí odvozených z biologických základů (metafora informačního zpracování). V případě obecné interpretace procesu ukotvení je nutné se soustředit na následující aspekty (Ziemke, 1999):

Musíme být velmi opatrní v případě abstrakcí, protože znamenají vždy nevlastní (což nemusí znamenat špatné) omezení tvorby systému.

Musíme znovu prozkoumat některé detaily, které jsme pravděpodobně předčasně abstrahovali.

Jedním z takových detailů může být fyziologické ukotvení, které se vyvíjelo během evoluce, při interakci agenta s prostředím.

Poslední bod je zdůrazněním enaktivního přístupu k ukotvování symbolů, přičemž sporná místa tohoto přístupu již byla zdůrazněna. Konkrétnější požadavky se objevují u Tadeo (2005):

1. *Bottom-up, senzomotorický přístup*
2. *Top-down zpětnou vazbu, která umožní sladění nejvyšší vrstvy ukotvených symbolů a spodní vrstvy, zajišťující senzomotorickou interakci s prostředím.*
3. *Schopnost agenta vytvářet reprezentace*
4. *Schopnost agenta kategorizovat a abstrahovat*
5. *Schopnost agenta komunikovat s ostatními, tedy ukotvovat symboly diachronicky, čímž předejdeme Wittgensteinovu problému "soukromého jazyka"*
6. *Evoluční přístup během vývoje 1-5*
7. *Uspokojení Z kritéria během vývoje 1-6*

Jelikož se během analýzy *Z kritéria* ukázalo, že některé jeho aplikace nejsou zcela přiléhavé, a že argumentace pomocí evolučního přístupu může jít za hranice designérských možností, navrhuji podrobit body 6 a 7 hlubší analýze, která povede k jejich reformulaci.

10 Závěr

V tomto článku bylo poukázáno na problematiku oblasti, se kterými se během procesu ukotvování symbolů potýkáme. Mnohé návrhy kritiků současných přístupů byl shledány jako velmi plodné pro další vývoj v této oblasti. Přesto se ukazuje, že některé aspekty kritiky jsou příliš restriktivní pro možnosti ukotvování pomocí dosavadních přístupů a že se je musíme pokusit reinterpretovat v širším kontextu. Tento článek je pouze elementárním shrnutím pro následnou diskusi. Celkovým výstupem je disertační práce zabývající se otázkami *symbol grounding*, jež se bude snažit danou problematiku nazřít více detailněji a s důrazem na přesnou identifikaci sporných míst.

Literatura

- [1] L. W. Barsalou: Perceptual symbol systems. *Behavioral and Brain Sciences* 22 (1999) 577-609
- [2] R. A. Brooks: The Engineering of Physical Grounding. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society* (1993) 153-154. Hillsdale, NJ: Lawrence Erlbaum.
- [3] A. Cangelosi, A. Greco and S. Harnad: From Robotic Toil to Symbolic Theft: Grounding Transfer from Entry-Level to Higher-Level Categories. *Connection Science* 2 (2000) 143-162
- [4] A. Cangelosi, A. Greco and S. Harnad: Symbol Grounding and the Symbolic Theft Hypothesis. In: *Simulating the Evolution of Language*. (A. Cangelosi and D. Parisi, Eds.), London, Springer, 2002: 191-210.
- [5] A. Cangelosi and S. Harnad: The Adaptive Advantage of Symbolic Theft over Sensorimotor Toil: Grounding Language Perceptual Categories. *Evolution of Communication, special issue on Grounding Language*. 4 (2001) 117-142
- [6] D. J. Chalmers: Subsymbolic Computation and the Chinese Room. In: *The Symbolic and Connectionist Paradigms: Closing the Gap*, (J. Dinsmore, Ed.), Hillsdale: Lawrence Erlbaum, 1992: 25-48
- [7] A. Clark: *Associative Engines*. MIT Press, Cambridge, MA, 1993
- [8] D. Cole: The Chinese Room Argument, *The Stanford Encyclopedia of Philosophy (Fall 2004 Edition)*, (Edward N. Zalta Ed.), URL = <http://plato.stanford.edu/archives/fall2004/entries/chinese-room/>.
- [9] S. Coradeschi and A. Saffioti: An Introduction to the Anchoring Problem. *Robotics and Autonomous Systems* 43 (2003) 85-96
- [10] P. Davidsson: Toward a General Solution to the Symbol Grounding Problem: Combining Machine Learning and Computer Vision. In: *AAAI Fall Symposium Series, Machine Learning in Computer Vision: What, Why and How?*, 1993: 157-161.
- [11] G. Dorffner and E. Prem: Connectionism, Symbol Grounding, and Autonomous Agents. In: *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, 1993: 144-148.
- [12] G. Dorffner, E. Prem, H. Trost: Words, Symbols, and Symbol Grounding, *Oesterreichisches Forschungsinstitut fuer Artificial Intelligence, Wien TR- 93-30* (1993)
- [13] G. Dorffner, M. Hentze, G. Thurner: A Connectionist Model of Categorization and Grounded Word Learning, In: *Proceedings of the Groningen Assembly on Language Acquisition*: (Koster C., Wijnen F. eds.), (GALA '95)

- [14] J. A. Fodor, Z. Pylyshyn: Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, (1988) 3-71.
- [15] S. Harnad, (Ed.): *Categorical Perception: the Groundwork of Cognition*, New York: Cambridge University Press, 287-300, 1987
- [16] S. Harnad: The Symbol Grounding Problem. *Physica D* (1990) 335-346.
- [17] S. Harnad: Grounding Symbols in the Analog World with Neural Nets – a Hybrid Model. *Psychology* 12 (2001) 12-78.
- [18] N. Chomsky: *Syntactic Structures*. The Hague: Mouton, 1957
- [19] I. Kant: *Kritika čistého rozumu*. Praha: Oikomenh, 2001
- [20] S. M. Kosslyn: *Image and Mind*. Cambridge: Harvard University Press, 1980
- [21] S. M. Kosslyn, G. Ganis and W.L. Thompson: Mental imagery: Against the nihilistic hypothesis. *Trends in Cognitive Science* 7 (2003) 109-111.
- [22] D. Law, and R. Miikkulainen: Grounding Robotic Control with Genetic Neural Networks. *Tech. Rep. AI94-223*. Austin: Dept. of Computer Sciences, The University of Texas at Austin: 1994
- [23] A. Paivio: *Mental representation: A dual coding approach*. New York: Oxford, 1986
- [24] J. Piaget: *Psychologie intelligence*. Praha: Portál, 1999
- [25] T. Regier: The Acquisition of Lexical Sémantice for Spatial Terms: A Connectionist Model of Perceptual Categorization. *PhD Thesis / Tech. Rep. TR-92-062*. Berkeley: Dept. of Computer Science, University of California at Berkeley: 1992
- [26] D. Roy: Grounding Language in the World: Schema Theory Meets Semiotics. (2004) URL: <http://web.media.mit.edu/~dkroy/papers/pdf/aij.pdf>
- [27] J. Searle: Minds, Brains, and Programs. *Behavioral and Brain Sciences* 3 1980 417-458
- [28] N. E. Sharkey and S. A. Jackson: Three Horns of the Representational Trilemma. In: *Symbol Processing and Connectionist Models for Artificial Intelligence and Cognitive Modeling: Steps towards Integration*, (V. Honavar and L. Uhr, Eds.), Cambridge, MA: Academic Press, 1994: 155-189.
- [29] N. E. Sharkey and S. A. Jackson: Grounding Computational Engines. *Artificial Intelligence Review* 10 (1996) 65-82.
- [30] J. A. L. Singh, R. M. Zingg.: *Wolf-Children and Feral Man*. Shoe String Pr Inc, 1966
- [31] P. Smolensky: On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11 (1988) 1 - 74.
- [32] M. Tadeo, L. Floridi: The Symbol Grounding Problem: a Critical Review of Fifteen Years of Research. *Journal of Experimental and Theoretical Artificial Intelligence* (2005) 419 – 445.
- [33] N. J. T. Thomas: Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science* 23 (1999) 207-245.
- [34] D. S. Touretzky, G. E. Hinton: Dymbols Among the Neurons Details of a Connectionist Interference Architecture. In: *Proceedings of the 9th International Joint Conference on Artificial Intelligence (IJCAI-85)* Los Angeles, Kalifornia
- [35] F. Varela, E. Thompson, E. Rosch: *The Embodied Mind - Cognitive Science and Human Experience*. Cambridge, MA: MIT Press / Bradford Books, 1991
- [36] L. Wittgenstein: *Philosophical investigations*. New York: Macmillan, 1953
- [37] T. Ziemke: Rethinking Grounding. In: *Understanding Representation in the Cognitive Sciences*, (A. Riegler, M. Peschl and A. von Stein, Eds.), New York: Plenum Press, 1999: 177-190.