

UNIVERZITA KOMENSKÉHO
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
ÚSTAV INFORMATIKY



Akvizícia gramatiky v modeli so sémantikou
rozlišovacích kritérií

Diplomová práca

Autor Marek Palanský
Školiteľ RNDr. Martin Takáč

BRATISLAVA

máj 2008

Čestne vyhlasujem, že som diplomovú prácu vypracoval samostatne, len s
použitím uvedenej literatúry a pod odborným dohľadom školiteľa.

Marek Palanský

V Bratislave 1. mája 2008

Abstrakt

Táto diplomová práca skúma proces osvojovania si prirodzeného jazyka v simulovanom prostredí. Jazyk je ukotvený vo vonkajšom svete a tvorí sémantickú štruktúru použiteľnú na zmysluplnú orientáciu v prostredí. Cieľom mojej práce bude obohatiť existujúci model osvojovania si sémantiky pomocou rozlišovacích kritérií o správne používanie viacslovných výrazov s dôrazom na syntaktické pravidlá v jazyku, založené na klasifikovaní a vzájomnom prepojení slov podľa ich funkcie.

Súčasťou skúmania vzájomnej interakcie syntaxe a sémantiky je implementácia navrhnutého modelu akvizície jazyka pričom komunita používateľov bude modelovaná multiagentovým systémom v jazyku Java. V tom budem simulovať vývoj jazyka obsahujúceho sémantické koncepty ukotvené v perceptuálnom priestore agenta.

Vnímaný priestor bude tvorený scénou, popísanou množinami obsahujúcimi páry "atribút-hodnota". Týmto spôsobom je možné popísať umelý svet pre agentov, ktorý je témou rôznych jazykových hier. Dôraz v skúmaní a realizovaní algoritmického návrhu takejto akvizície konceptov zviazaných s bežnými jazykovými výrazmi je kladený na bohatosť zachytiteľných konceptov a schopnosť jednoduchého a spoľahlivého prenosu medzi rôznymi generáciami používateľov jazyka.

Súčasťou skúmania je pohľad na tvorbu kategórií v prostredí rozlišovacích kritérií. V navrhnutej slovnej hre, je gramatika postavená na kombinácii slovných druhov. Priebeh komunikácie zabezpečuje správnu výuku agenta žiaka. Agent učiteľ jednoznačne popíše náhodný objekt a následne sa žiak snaží o identifikáciu. Tento proces necháme viacgeneračne iterovať.

Moja práca sa snaží nevyužívať žiadne vnútorné obmedzenia rôznych konkrétnych kultúr a ich jazykov ale uprednostňuje jednotný prístup k učeniu sa ľubovoľného jazyka založeného na výskumoch získavania lingvistických znalostí deťmi pri ich rannej akvizícii jazyka.

Kľúčové slová: jazyk, gramatika, kognitívna sémantika, rozlišovacie kritérium, bootstrapping

Obsah

1	Úvod	1
1.1	Predhovor	1
1.2	Prirodzený jazyk	2
1.3	Jazykové hry	3
2	Tvorba významov mechanizmom obmedzovania sémantiky	5
2.1	Predpoklad uceleného objektu	5
2.2	Podobnosť objektov	6
2.3	Predpoklad o taxonómii	6
2.4	Predpoklad o vzájomnej výlučnosti	7
3	Tvorba kategórií	7
3.1	Odlíšenie základnej úrovne od iných úrovní kategórií	7
3.2	Všeobecná kategória	8
4	Kognitívna sémantika	9
4.1	Predstavové schémy	9
4.2	Sémantika rozlišovacích kritérií	9
4.2.1	Konštrukcia významov a jej dynamika v procese iterovaného učenia	10
5	Gramatika	13
5.1	Predikát-subjekt	13
5.2	Kirbyho model iterovaného učenia	14
5.3	Konštrukčná gramatika	15
5.4	Konceptualizácia slovných druhov	16
5.4.1	Podstatné mená	17
5.4.2	Prídavné mená	17
5.4.3	Príslovky	18
5.4.4	Slovesá	18

6	Návrh výpočtového modelu akvizície jazyka	19
6.1	Špecifikácia cieľa	19
6.2	Jazyková prax v pomenúvacej hre	20
6.3	Popis experimentu	22
6.4	Náčrt interpretácie priebehu simulácie	24
6.5	Návrh modelu akvizície jazyka a jeho implementácia	26
6.6	Parametre simulácie	31
6.7	Charakteristika takto naučeného jazyka	33
7	Kvantitatívne výsledky simulácií a ich interpretácia	34
7.1	Prvá generácia	35
7.2	Druhá generácia	40
7.3	Stabilita ďalších generácií	41
8	Problematické miesta a budúci vývoj systému	42
9	Zhrnutie a diskusia	45
10	Záver	46
11	O elektronickej prílohe	48

1 Úvod

1.1 Predhovor

Diplomová práca sa usiluje o rozširovaním programového modelu akvizície jazyka. Snažíme sa modelovať sémantiku prirodzeného jazyka, pričom významy zakladáme na tzv. rozlišovacích kritériách. Vychádzame z teoretického návrhu sémantiky od Jána Šefránka ¹ a implementácie od Martina Takáča ².

Pri upresňovaní nášeho skúmania budeme postupovať pozvoľna. Najprv predstavíme jazyk ako ľudský fenomén. Charakterizujeme jeho význačné vlastnosti, pričom budeme postupovať smerom ku konkrétnejšiemu pohľadu na fungovanie jazyka v spoločnosti. Výklad o jazyku sa nezaobíde bez detailnejšej analýzy rôznych náhľadov na sémantiku, ktorá je stavebným kameňom každého jazyka. Rovnako nutná súčasť pohľadu na jazyk je analýza možnosti prepojenia a vzájomného obmedzovania významov slov vo vete použitím vhodnej syntaxe.

Po takomto všeobecnom úvode, analýzach a náčrtoch mechanizmov akvizície jazyka, sa budeme venovať konkrétne jazyku, ktorý je možné vytvoriť pomocou kognitívneho aparátu využívajúceho sémantiku rozlišovacích kritérií. Popíšeme výsledky implementácie existujúceho modelu s takouto sémantikou. Tieto výsledky budú slúžiť ako vstup pre druhú časť textu, v ktorej sa budeme venovať vlastnej časti práce spočívajúcej v snahe o rozšírenie modelu akvizície jazyka o gramatiku, to znamená o objektovo orientovanú programovú implementáciu syntaktických štruktúr zabezpečujúcich prenos kompozičného jazyka.

Predtým však ešte predstavíme experiment, ktorý je vzorom simulácie a podrobne popíšeme ako je funkčnosť programu implementovaná. Potom budeme analyzovať priebeh experimentu a predstavíme jeho kvantitatívne výsledky úspešnosti. Ďalej sa budeme venovať dodatočnej analýze a inter-

¹Kapitola 4.2

²Kapitola 4.2.1

pretácii týchto výsledkom. Navrhujeme aj určité zlepšenia modelu na základe odhalených slabých miest v terajšom modeli.

1.2 Prirodzený jazyk

Zvieracia ríša je bohatá na rôzne formy života, ktoré sú pestré nielen svojim fyzickým telom a svojím spôsobom prežitia v prostredí, ale rovnako aj rôznymi formami dorozumievania sa. V prírode proces komunikácie prebieha v rámci spoločenstva rovnakého druhu ako aj medzi navzájom súperiacimi druhmi napríklad v rámci rolí dravec - korisť. Vo všeobecnosti však tieto zvieracie signály sú skôr reakciami na externé stimuly referujúce k aktuálnej prítomnej situácii bez možnosti vyjadrenia sa v inom, bližšie špecifikovanom čase a priestore. Fakty dôležité pre blízku budúcnosť, ktoré sú predmetom spoločného záujmu sú vyjadrené v primitívnejšej forme jazykom tela ako tomu je pri včelom tanci špecifikujúcom približnú lokáciu zdroja potravy. Jazyk tvorí vrodenuú reakciu na prítomné okolnosti, v rozvinutejšej podobe je úmyselným signalizovaním a porozumením informácii vrámci spoločného zámeru v určitej situácii. Zvieracia komunikácia je postavená na vrodenných inštinktoch súvisiacich so snahou prežiť.

Jazyk je u ľudí naučený kultúrnym prenosom s unikátnou možnosťou vyjadriť konceptuálne zovšeobecnenia, ktoré žiadne iné zviera v takejto dokonalej forme nedokáže. Výrazy v jazyku, ktoré komunita ľudí používa pri komunikácii, sú naučené počas života, prostredníctvom interakcie so svetom v kultúre, v ktorej žijú. Takáto schopnosť naučiť sa asociovať slová s vnútornými významami a tvoriť z nich štrukturované vety bola pre ľudstvo z hľadiska evolúcie komunikácie výhodná.

Prirodzená ľudská kapacita pre jazyk je fundamentálnym znakom pre pochopenie, čo znamená myslieť ako človek. Ale z čoho pozostáva toto nadanie a ako ho mentálne v mysliach reprezentujeme a pracujeme s ním? Zodpovedať otázky tohto druhu nie je ľahké. Dôležitým spôsobom štúdia tejto problematiky je analýza vývoja jazykových schopností detí, mladých žiakov v procese učenia, a odhaľovaním spôsobov rozvíjania porozumenia sveta okolo nich.

Deti sú prakticky bez výnimky schopné akvizície ľubovoľného jazyka už vo veľmi skorom veku v rapídnom tempe i napriek extrémnej komplikovanosti a množstve nuansí. Zrejme je ľudský druh šikovne geneticky naprogramovaný pre porozumenie jazyka, ale mentálna výbava, v podobe schopnosti intuitívne myslieť, prispieva ku zručnosti osvojiť si jazyk vo veľkej miere a umožňuje zachytenie zložitejších konceptov.

1.3 Jazykové hry

Termín význam používame na označenie obsahu alebo zmyslu nejakého zoskupenia slov v jazyku. Odovzdávanie a prijímanie takýchto informácií je účelom komunikácie a jej obsah je predmetom skúmania sémantiky. Pohľady na sémantiku prirodzeného jazyka boli v minulosti preskúvané dvoma hlavnými smermi s navzájom nezlučiteľnými prístupmi.

Jazyk s referenčnou sémantikou stavia svoj model na pevnom prepojení reálne existujúcich objektov s významom slov. Frege ako zástanca tohto smeru neskôr rozlíšil zmysel výrazu a jeho denotát. Každému objektu priradil jedinečnú entitu, ktorú nazval denotát a ten bol identifikovateľný pomocou významu slovného spojenia. Jedná sa teda o nepriamy vzťah medzi symbolom, významom a jeho referentom.

Tento prístup neskôr zožal kritiku u Wittgensteina, ktorý ho zpočiatku zdieľal, ale v neskoršej fáze svojho života sa sústredil na skúmanie množstva bežných jazykových regularít v jazyku, ktoré vybočovali z Fregeho ³ predošlého pohľadu na jazyk. Jeho filozofické skúmania naznačovali, že fregeho model nie je vždy schopný správne zachytiť význam výrazov, ktoré sú v ľudskej komunikácii zrozumiteľné a bežne používané. Vylučoval preto existenciu jednoznačného prepojenia objektov s ich významami či denotátmi a zdôrazňoval pragmatické hľadisko v rečových hrách, kde význam výrazov identifikoval s ich samotným použitím v jazyku spôsobom, ktorý je známy a vznikol na základe spoločnej dohody.

Vo svojej modrej knihe pristupuje k učeniu sa jazyka takým spôsobom,

³Vybočovali aj z jeho vlastného skorého pohľadu na jazyk

že popisuje a analyzuje jednotlivé slovné hry. Skúma komunikáciu podľa zaužívaných pravidiel medzi ľuďmi v nejakej situácii, pri vzájomnej činnosti. Pravidlá jazyka sú analogické k pravidlám stolových hier, vyslovenie myšlienky v jazyku je analogické k vykonaniu ďalšieho ťahu. Iba v rôznych činnostiach, ktoré človek vykonáva majú slová význam. Jazyk je doslova vybudovaný zo slovných hier. Kombináciou jednoduchších hier dostávame zložitejšie formy jazyka, ktoré rozširujú možnosti pochopenia nových významov.

„Language games are the forms of language with which a child begins to make use of words... When we look at such simple forms of language the mental mist which seems to enshroud our ordinary use of language disappears. We see mental activities, reactions, which are clear-cut and transparent. On the other hand we recognize in these simple processes forms of language not separated by a break from our more complicated ones. We see that we can build up the complicated forms from the primitive ones by gradually adding new forms.“ [21]

Na základe termínu jazyková hra, ktorý používal pre označenie ľudskej komunikácie vrámci kontextu sa neskôr postavila teória rečových aktov rozpracovaná Austinom [1]. Hovorca v rečovom akte nie je iba opisovateľom sveta. Samotné používanie jazyka je chápané aj ako činnosť v kontexte situácie.⁴

Keď sa deti učia hovoriť o udalostiach alebo činnostiach, musia byť schopné analyzovať, čo vidia a vedieť ako to označiť relevantným výrazom. Zisťujú, aký druh jazyka sa učia z pohľadu ako hovoriť o vzťahoch typu odosielateľ a príjmateľ alebo predmet a jeho umiestnenie. Musia teda aj zistiť, akým spôsobom sú tieto vzťahy v jazyku gramaticky označené a aké sú možnosti a vlastnosti kombinácií rôznych typov slov. Rovnako je dôležitá aj schopnosť identifikovať predikát-subjektový vzťah. Prvé viacslovné vety však môžu znamenať, že deti sa z počiatku jednoducho sústredia na vyjadrenie najvýraznejších znakov toho, čo zachytí ich pozornosť a vedia to pomenovať.

⁴Rečový akt môžeme skúmať z hľadiska lokúcie, ilokúcie a perlokúcie

2 Tvorba významov mechanizmom obmedzovania sémantiky

V tejto kapitole sa bližšie zaoberám výskumami učenia sa jazyka detí a ich predispozíciami zamietnuť alebo naopak preferovať určité významy slov. Tieto obmedzenia pre význam výrazov slúžia na podstatné redukovanie možných významov a neurčitostí tak, aby problém naučenia sa jazyka čo najviac zjednodušili a umožnili vôbec riešiteľnosť tohto problému. Pri každej komunikácii zdieľame určité predpoklady.

Na nasledovné hypotézy o predpokladoch, ktoré zjednodušujú prenos významu sa nemusíme pozeráť ako na pevne definovaný zoznam. Deti vnímajú svet určitým spôsobom a ich spôsob práce s konceptami a jazykom môže byť podopretý vhodnými intuitívnymi heuristikami vychádzajúcimi z podoby ľudskej kognície. Nie je celkom isté, do akej miery sú tieto predpoklady vrodené a do akej miery sú nadobudnuté vo veľmi rannom vývoji dieťaťa.

2.1 Predpoklad uceleného objektu

Macnamara [9] argumentuje, že deti prirodzene popisujú svoje prostredie pomocou množiny v ňom obsiahnutých a rozpoznateľných objektov. Keď sa učia nové slová, automaticky predpokladajú, že nové slovo referuje na celý objekt namiesto nejakej jeho časti alebo vlastnosti. Týmto spôsobom je zabezpečená značná redukcia možnosti pochopenia zaumienených významov. Učenie detí dospelými funguje len za predpokladu, že učiteľ vie správne predpovedať interpretáciu významu prirad'ovaného nejakému objektu. Predpoklad o ucelenosti objektu je teda veľmi praktický spôsob obmedzenia sémantiky vysvetľujúci dôležitý základný princíp postupnej svojpomocnej akvizície konceptov.

2.2 Podobnosť objektov

Ďalším významným odhalením vplyvu obmedzovania sémantiky pri učení je objavenie preferencie pre kategorizáciu pevných objektov na základe ich tvaru. Landau a kolektív [8] vykonal experiment, v ktorom deti zaradovali objekty do kategórií a zistilo sa, že aj napriek očividným rozdielom vo veľkostiach, textúre a iných vlastnostiach objektov, boli tieto objekty zaradené do spoločnej kategórie na základe spoločnej podobnosti v tvare. Tento predpoklad úzko súvisí s predpokladom uceleného objektu a spresňuje vhodné spoločné obmedzenie pre klasifikáciu objektov. Uľahčuje proces naučenia sa významu bežných podstatných mien.

2.3 Predpoklad o taxonómii

Taxonomická organizácia je o zoskupovaní objektov rovnakého typu a je v kontraste s tématickou organizáciou, ktorá zoskupuje objekty na základe ich vzájomného vzťahu. Markman a Hutchinson [11] experimentálne ukázali, že deti používajú taxonomickú a tématickú organizáciu rozdielne a to v závislosti od toho, či sa učia slová alebo nie.

Napríklad, pokiaľ sa nepoužili slová, deti bežne zoskupili auto a pneumatiku, ale pokiaľ ukážeme na auto a nazveme ho ako "dax" a žiadame dieťa o nájdenie iného predmetu označeného ako "dax", tak pravdepodobne dostaneme ako odpoveď predmet z rovnakého taxonomického zoskupenia ako napríklad bicykel. Záver z tohto experimentu sa dá zhrnúť ako návrh na tento predpoklad pri učení nových slov - deti uprednostňujú v náhlade na svet taxonomickú organizáciu pokiaľ sa učia nové slová, inak im je bližšia tématická organizácia.

Táto hypotéza naznačuje, že deti majú pri učení sa jazyka špecifické metódy organizácie objektov, ktoré sú odlišné od bežného náhladu. Taxonomický predpoklad je špecifický pre jazykovú doménu a funguje v spojení s predpokladom o ucelenom objekte tak, že deti interpretujú neznáme slová ako referujúce na objekty z rovnakej taxonomickej triedy.

2.4 Predpoklad o vzájomnej výlučnosti

Tento predpoklad funguje na princípe, že kategórie objektov sú navzájom rozpoznateľne rôzne množiny. V nadväznosti na predchádzajúcu kapitolu o taxonómii v tomto prípade tí istí autori pristupujú k vzájomnej výlučnosti aplikovanej konkrétne na základnej úrovni kategorizácie. Na tento predpoklad sa ešte bližšie pozrieme v nasledujúcich podkapitolách venovaných tvorbe kategórií a rovnako budeme venovať pozornosť aj iným súvisiacim predpokladom akým je napríklad princíp dobrého kontrastu medzi dvoma správne utvorenými základnými kategóriami.

3 Tvorba kategórií

Pojmom hierarchická štruktúra kategórií označujem rôzne stupne všeobecnosti, na ktorých ľudia bežne organizujú svoje znalosti. Ide o rozlišovanie medzi rôznymi vhodnými úrovňami tried objektov. Napríklad znalosti môžeme zoskupiť okolo viac všeobecných kategórií vyjadrených označeniami ako "zvieratá", "nábytok", a na druhej strane okolo viac špecifických kategórií aké tvoria označenia "mačky", "stoličky".

Prvú spomenutú skupinu budeme nazývať všeobecná kategória alebo umelá, a druhú označím ako základnú úroveň alebo prirodzenú, pričom platí vzťah inklúzie: základná kategória (prirodzená) je obsiahnutá vo všeobecnej (umelej). Tretiu skupinu kategórií budeme označovať ako podradenú a jej príkladom sú označenia ako "vlčiaky" alebo "hojdacie stoličky". Tento prístup postavený na troch rozličných úrovniach všeobecnosti kategórií bol rozpracovaný ako užitočný spôsob kategorizácie pre orientáciu vo svete. [14]

3.1 Odlíšenie základnej úrovne od iných úrovní kategórií

Základná úroveň poskytuje dobrý kompromis medzi dvoma rôznymi cieľmi kategorizácie: (1) maximalizácia podobností medzi členmi kategórie (2) mi-

nimalizácia podobností s členmi iných kategórií. Táto úroveň je najprirodzenejšia pre označovanie bežných objektov.

Tversky a Hemenway [19] argumentuje, že podstatnou vlastnosťou rozlišovania na základnej úrovni je rozlišovanie spoločných častí objektov a nie len ich jednotlivých vlastností. Tento prístup súvisí s už spomínaným predpokladom o tvare, ktorý obmedzuje tvorbu významov na základe výrazných, preferovaných atribútov charakterizujúcich tvar objektu. Pre spresnenie optiky, nedefinujeme kategórie náhodnou kombináciou atribútov, ale tvoríme ich korelovanými štruktúrami atribútov existujúcich prirodzene vo svete. Napríklad zvieratá, ktoré majú perie, majú zvyčajne aj zobák a krídla - pričom zvieratá so srstou nemajú zvyčajne zobák a krídla. Takáto prirodzená štruktúra atribútov dáva priestor pre vznik prirodzených kategórií. [13]

Ďalší spôsob vedúci k rozpoznaní základnej kategórie funguje sledovaním frekvencie jej používania, pretože je jednoduchšia na naučenie a tiež preferovaná dospelými pri rozprávaní s deťmi.

3.2 Všeobecná kategória

Predchádzajúca kapitola o prirodzených kategóriách naznačovala, že tento spôsob zoskupenia objektov je relatívne jednoducho naučiteľný, pretože všetky objekty sú pre človeka perceptuálne zjavne podobné. Pozrime sa teraz na všeobecné kategórie tvorené zložitejšími súvislosťami medzi objektmi. Tieto súvislosti môžu mať formu nejakej dodatočnej informácie špecifikujúcej kritérium akvizície tohto konceptu predstavujúceho danú kategóriu. Jedným z dôležitých spôsobov špecifikovania kritériá môže byť použitie jazyka na jeho dodatočné vysvetlenie. Napríklad vysvetlenie prečo koala a kengura patria medzi vačkovce je jednoducho také, že obe zvieratá nosia svoje mláďatá vo vaku. Táto nutnosť dodatočnej lingvistickej informácia je jedným z viac iných dôvodov, prečo sa predpokladá, že tieto umelé kategórie sú deťmi získané až po vzniku základnej kategórie. [10]

4 Kognitívna sémantika

Tento druh sémantiky je časťou lingvistického smeru, ktorý definuje gramatiku ako konceptuálnu štruktúru, ktorá je vytvorená na základe ľudíom vlastných schématických konceptov. Znalosť jazyka pritom využíva všeobecné kognitívne schopnosti človeka a nie špeciálny jazykový modul.

4.1 Predstavové schémy

Sémantika postavená na schémach ⁵ skúma mechanizmy konštrukcie významov zo základných jednotiek znalostí alebo skúseností so svetom, ktoré sú v ľudskej kognícii univerzálne prítomné. Lakoff [7] rozpracoval teóriu predstavových schém, v ktorej odmieta paradmigu symbolickej manipulácie so symbolmi, postavenú na oddelenosti tela od myslenia. Vonkajší svet človek vníma cez predstavové schémy, a význam nie je objektívna skutočnosť, ale len subjektívny vnem. Nie je isté, do akej miery sú tieto koncepty vrodené alebo získané v rannom veku.

Skrz schémy vyjadrujú určitým spôsobom intuitívne, o zmysly sa opierajúce koncepty ako trajektória, ohraničenosť, povrch, kontakt a iné. Ich kombináciou sa tvoria zložitejšie koncepty, ktoré sú pri správnej interpretácii tiež intuitívne zrozumiteľné. Význam abstraktných výrazov vychádza z princípu prenášania významu na základe vnútornej (metonymia) alebo vonkajšej (metafora) súvislosti z nejakého súboru základných konceptov.

4.2 Sémantika rozlišovacích kritérii

Kognitívny prístup k jazyku a jeho sémantike využíva predpoklad o individuálnych významoch, ktoré sa tvoria v komunikácii na základe subjektívnej skúsenosti zo svetom cez dostupné vnemy. V nasledujúcom sa budeme zaoberať určitým smerom výskumu pochopenia kognície a jej dynamiky, pričom neskôr popísaný prenos jazykom bude nadstavbou nad týmto typom kognície.

⁵Image schemas

Táto sémantika modeluje kognitívne schopnosti organizmu prostredníctvom zachytenia významov pomocou jednoduchého predpokladu. Tým je schopnosť živého organizmu rozlišovať, čo je vlastné každému reaktívnemu správaniu a nevyžaduje prítomnosť žiadnych zložitých mentálnych procesov. Význam je stotožnený s rozlišovacími kritériami, abstrakciami schopnosti rozlišovať. Rozlišovanie sa v elementárnej forme deje medzi triedami predmetov alebo indivíduami a ich vlastnosťami alebo vzájomnými vzťahmi. Zachytenie situácii je možné rozlišovaním zmeny v po sebe nasledujúcich časových krokoch. Rozlišovacie kritéria ponúkajú riešenie problému uzemnenia symbolov v tom zmysle, že sémantiku nekonštruujeme prekladom z jazyka do jazyka, ale symboly sú uzemnené ⁶. Takéto pochopenie redukované na rozlišovanie spolu so schopnosťou komunikácie, prináša bohaté možnosti konštrukcie významov v procese modelovania ľudskej komunikácie. Kritériá rozlišovania tvoria jednotiaci rámec pre reprezentáciu rôznych typov konceptov pričom koncept je definovaný ako funkcia miery dôveryhodnosti predloženej inštancie nejakého perceptu. [15]

4.2.1 Konštrukcia významov a jej dynamika v procese iterovaného učenia

V tejto kapitole popíšem implementovaný model konštrukcie významov za pomoci učenia, skúmajúci dynamiku prenosu významov v iterovanom medzigeneračnom prenose. [17] Významy nie sú dané vopred externe, ale sú individuálne vytvárané.

Individuá, agenty v architektúre multiagentového programovania, sa nachádzajú v prostredí, na scéne. Tieto agenty pripisujú významy aspektom v prostredí, ktoré sú relevantné ich cieľom. Využívajú na to možnosti sémantiky, postavenej na rozlišovacích kritériách a navyše obsahujú jazykovú úroveň, ktorá im dovoľuje asociovať slová s konceptmi. Cieľom pre agenty je spolu komunikovať a maximalizovať vzájomnú mieru porozumenia, čo závisí od reprezentácie správnych konceptov a ich asociovanie so správnymi slova-

⁶Symbol grounding problem [5]

mi.

Metóda iterovaného učenia slúži v takomto systéme na modelovanie ľudského jazyka ako komunikačného systému zameraného na externalizáciu a prenos vnútornej reprezentácie konceptov. Nutnou podmienkou úspešnosti je, že agenty zdieľajú spoločné prostredie, rovnaké učiace mechanizmy a podobné interakčné skúsenosti. Navyše situácia, ktorá je kontextom komunikácie, pomáha za pomoci vhodnej pragmatickej inferencie v prekonávaní nejasnosti pri výmenách významu. Samotné subjektívne významy rôznych agentov sa koordinujú v procese samoorganizácie. Jazyk vzniká lokálnou interakciou medzi používateľmi jazyka na základe pozitívnej spätnej väzby, ktorá je merítkom úspešnosti. Významnú úlohu majú v tejto simulácii aj neverbálne prostriedky komunikácie ako napríklad ukazovanie a zdieľanie pozornosti.

Jazykový experiment je zameraný na skúmanie vplyvu verbálnej inštrukcie (pomenovania) na process formovania kategórií. Dva agenty situované v prostredí by sme bežne označili ako učiteľa a žiaka. Vzájomná komunikácia má formu pomenúvania rôznych aspektov momentálne vnímanej scény tak, že učiteľ učí žiaka a žiak odvodzuje významy učiteľových slov na základe podobností medzi referentami rovnakých slov v rôznych situáciách. Vzťah medzi konceptami a ich pomenovaniami využíva predpoklad, ktorý je kľúčový v opísanom prenose významov. Predpoklad pozostáva z takého obmedzenia, že rôzne slová majú rôzne významy a všetky referenty toho istého slova sa považujú za príklady jedného konceptu.

Simulované prostredie v experimente pozostávalo z plošných geometrických tvarov charakterizovaných piatimi atribútmi: počet vrcholov, súradnice stredu a veľkosť opísaného obdĺžnika. V časových krokoch sa prostredie dynamicky menilo, objekty mohli získať, zmeniť alebo stratiť svoje vlastnosti. Agent učiteľ bol vybavený predefinovanou slovnou zásobou, ktorú používal na označenie objektov na scéne prídavnými menami a podstatnými menami. Zmenu v dvoch po sebe nasledujúcich časových krokoch mohol označiť slovesami. Tieto lingvistické prostriedky majú však formu súboru

jednoslovných pomenovaní.

Po odsimulovaní prvej generácie je agent učiteľ nahradený žiakom, takže nová generácia preberá svoje znalosti od vyučených žiakov a tí už nepoužívajú predefinovanú slovnú zásobu ako pôvodní učitelia, ktorí slúžili len na naštartovanie tohto jazykového procesu. Výsledky simulácií ukázali, že učiaci sa agent môže na základe pomenovaní skonštruovať významy dostatočne podobné učiteľovým. Učenie na základe medzi-situačných podobností je rýchle a spoľahlivé. Žiak si veľmi skoro osvojí ontológiu a lexikón dostatočnej kvality, ktoré ostanú stabilné počas ďalšieho behu simulácie. Takýto vertikálny prenos jazyka medzi generáciami sa vytvára opakovanými transformáciami medzi osobnou jazykovou kompetenciou a externalizovanými výpoveďami - opakovanou akvizíciou a produkciou. V procese iterácie však neostanú výrazy úplne nezmenené a vyvíjajú sa, čo je v súlade s pozorovaním, že ľudský jazyk podlieha historickým premenám bez toho, aby stratil svoju vhodnú dorozumievajúcu funkciu.

Význam v experimente použitých adjektív ("malý", "veľký") bol založený na originálnom vzťahu k nejakým konkrétnym konštantám (priemerná veľkosť). Toto nie je realistická požiadavka, pretože sémantika adjektív je obvykle závislá od podstatného mena, ktoré modifikuje. "veľký zajac" odkazuje na inú veľkosť ako "veľký slon". Výskum postavený na tomto modeli naznačuje, že sémantická závislosť adjektív môže byť dôsledkom dynamiky selekčného procesu v rámci iterovaného prenosu, v ktorom sú významy založené na štrukturálnych vzťahoch oveľa stabilnejšie.

5 Gramatika

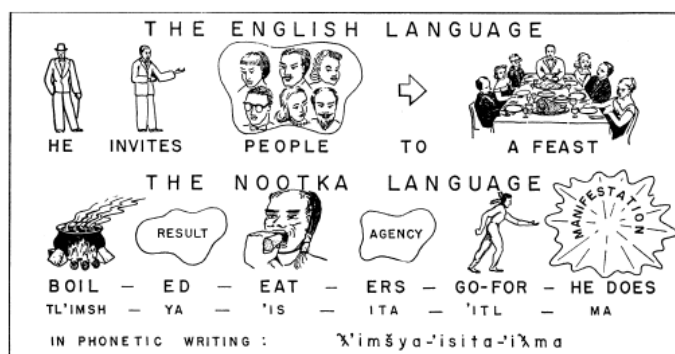
Gramatika opisuje vzťahy, ktoré používame v jazyku na tvorenie viet a to najmä morfológiu a syntax. V každom prirodzenom jazyku sú nejakým spôsobom definované vetné členy, teda existuje nejaký spôsob, ktorým sa určuje funkcia slova vo vete. V slovenčine a iných takzvaných syntetických jazykoch sa využíva viac morfológia, čiže flexia - ohýbanie slov, skloňovanie. V japončine slúžia častice na určenie pozície slova vo vete, teda na určenie vetných členov. Angličtina je naopak príklad analytického jazyka, preto kladie väčší dôraz na poradie slov a význam je viac závislý od kontextu situácie. K akvizícii tohoto typu jazyka spejeme.

5.1 Predikát-subjekt

Všeobecne predpokladáme, že jazyky sú založené na paradigme predikát-subjektu. Je rozumné očakávať aspoň túto základnú podmienku, no v skutočnosti sú známe jazyky domorodých amerických indiánov, ktoré pracujú s jazykom odlišne a slovami B. Whorfa, stavba vety skôr pripomína aplikovanie pravidiel syntézy ako pri chemickej reakcii. [20] Uvádza príklad dvoch viet, ktoré majú v angličtine (ako aj v slovenčine) úplne iný význam, no paradoxne v jazyku kmeňa Shawnee znamenajú to isté. "I push his head back / Tlačím mu dozadu hlavu" a "I drop it in water and it floats / Vhodím to do vody a pláva to". Nebudeme však posudzovať platnosť whorfových teórií, ktoré sa stretli so širokou kritikou napríklad Pinkera. [12]

Tu je na mieste poznámka k motivácii vlastnej práce. Bohužiaľ nemôžeme skúmať každú drobnosťku v množstve ľudských jazykov a i keď s vedomým zovšeobecnenia, budeme sa snažiť skúmať jazyk nezávisle od nejakej jeho konkrétnej inštancie, no s dôrazom na dôkladnú špecifikáciu slovnej hry tak, aby sa naučenie jazyka stalo zvládnuteľné a intuitívne modelovateľné v dostupnom modeli sémantiky pomocou jednoduchšej analytickej gramatiky.

Pozrime sa na priložený obrázok iba ako na demonštráciu toho, že rozdiely medzi jazykmi môžu byť priepastné.



Obrázok 1: Porovnanie spôsoby skladania konceptov vo vete "Pozval ľudí na oslavu." v anglickom jazyku a rovnakej vety, z hľadiska zmyslu, v jazyku indiánskeho kmeňa Nootka.

5.2 Kirbyho model iterovaného učenia

Kirby vo svojej práci [6], predstavil model iterovaného učenia za účelom samovoľného vzniku kompozičných štruktúr v jazyku medzigeneračným prenosom. Model simuloval jazykové hry medzi učiteľom a žiakom. Učiteľ nemal v prvej generácii žiadne lingvistické znalosti, ale mal vlastnosť vytvárať náhodné reťazce pre časti vety, ktorej nerozumel. Takto bol umožnený vznik jazyka, ktorý sa odovzdával z generácie na generáciu takzvaným kultúrnym prenosom v rámci populácie. Jazyk bol postavený na predikát-subjektom vzťahu, pričom bola implementovaná bežná generatívna gramatika.

Agent mal schopnosť zjednotiť vytvorené gramatické pravidlá do jedného, ak to bolo možné pre ich podobnosť, ako aj vytvárať nové gramatické pravidlá. Pravidlá vznikali na základe opakovania sa časti reťazca v rôznych dlhších reťazcoch, spolu so snahou o čo najoptimálnejšie (najkratšie) kódovanie významov.

Tento prístup je výlučne symbolický. Uchovávané reťazce znakov nemali žiaden vo svete ukotvený význam a teda významom boli samotné reťazce písmen. Proces medzigeneračného prenosu mal preto "telepatickú" formu, agent spolu so slovami získaval aj ich význam.

5.3 Konštrukčná gramatika

Konštrukcie, základné jednotky konštrukčnej gramatiky, sú ustálené formy vyjadrenia rôznych typov významov pomocou viacslovných výrazov v jazyku [3]. Kombinácie slov ako napríklad modifikátor+objekt, objekt+sloveso, individuum+sloveso+lokácia a iné tvoria šablóny pre vyjadrenie určitej formy významu. Význam jednotlivých zložiek konštrukcie môže byť prepojený rolovým rámcom platným pre túto konštrukciu. Veta "Utekaj sem!" aktivuje predstavovú schému označiteľnú ako zdroj-cesta-cieľ, pričom rolový rámec určuje, že niekde sa nachádzajúce individuum subjektu, sa má presunúť do cieľa, označeného ako "sem". Ide teda o bežne používanú postupnosť kategórii slov so špecifickým významom obmedzeným použitou šablónou.

Každé zoskupenie slov okolo slovesa má vlastné syntaktické pravidlá a vzťahy, prepájajúce sémantiku jednotlivých funkčných zložiek, tvorené na základe interakcie so svetom. Predpoklad, že slovesá určujú štruktúru konštrukcií je predpoklad nazývaný ako hypotéza slovesného ostrova ⁷[18]. Je zameraná na vysvetlenie stavby gramatiky u detí. Vety v jazyku sú deťmi vnímané ako individuálne vnútorne organizované zoskupenia slov, ostrovy, v inak neorganizovanom gramatickom systéme. V skorom štádiu vývoja sa dieťa učí vhodné slová a ich syntaktické pravidlá pre jednotlivé slovesá samostatne, a poradie slov a morfológické značky naučené pre toto jedno sloveso sa automaticky nezovšeobecňujú na iné slovesá. Dôvodom pre to je, že deti, učiace sa prirodzenému jazyku, nemajú žiadne syntaktické kategórie slov alebo univerzálne pravidlá tak ako dospelí, ani nemajú žiaden typ triedy sloves, ktorá by umožnila zovšeobecňovať medzi slovesami. Metódy symbolického začleňovania slov pri tvorbe viet sú špecifické a fungujú zo začiatku pre rôzne slovesá zvlášť. Čo deti majú k dispozícii v tomto štádiu sú vedomosti o úlohách, ktoré zohrávajú rôzne entity v špecifických situáciách, spolu so syntaktickými prostriedkami, ktoré ich pomáhajú rozlišovať. [18]

Učenie prebieha induktívne na základe príkladov, metódou zovšeobecňovania. Namiesto kombinovania základných syntaktických

⁷Verb island hypothesis

jednotiek používame vhodné komplexné šablóny, gramatické konštrukcie. Stelesnená konštrukčná gramatika ⁸ [2] sa sústreďuje na konštrukcie definované určitým spôsobom prepojenými schémami významu, vydedukovanými zo senzomotorickej skúsenosti agenta.

Využívanie gramatiky v jazyku poskytuje hovoriacemu schopnosť produkovať nekonečné množstvo rôznych výrazov. Napriek tomu je veľká pravdepodobnosť, že poslucháč porozumie. Spôsob kódovania významov v jazyku pomocou viet je unikátny a vlastný iba ľuďom. Využíva princíp kompozicionality a rekurzivity. Skladanie významov má za následok, že navzájom štruktúrou podobné vety v jazyku budú mať obvykle v sémantickej reprezentácii k sebe blízko. Tento vzťah je v kontraste s použitím holistického jazyka. Takýto jazyk je tvorený len jednoduchým asociovaním medzi slovami (signálmi, znameniami) a významami a preto podobné slová majú bežne úplne iné významy. [16]

5.4 Konceptualizácia slovných druhov

V tejto kapitole ponúkame výber niektorých známych gramatických kategórií slov v jazyku. Sústredíme sa najmä na ich aspoň základnú charakterizáciu v pojmoch teórie rozlišovacích kritérií a budeme sa zaoberať možnosťami vyjadrenia rôznych typov významov v rámci každého slovného druhu.

Dieťa má pri jazykovej výuke najprv fázu kedy nerozoznáva jednotlivé slová vo vete a rozumie im iba ako celistvému označeniu významu bez vnútornej syntaktickej štruktúry. Neskôr s vývojom jazykových zručností prichádza obvykle schopnosť pomenovať viacerými slovami určitý koncept akoby vymenovaním jednotlivých slov priamo súvisiacich s kontextom. Nato nasleduje u dieťaťa bežná viacslovná fáza zanedlho nato.

⁸Embodied construction grammar

5.4.1 Podstatné mená

Význam podstatných mien tvoria v teórii rozlišovacích kritérií kritériá objektov, založené na rozlišovaní podstatných atribútov a ich veľkostí. Je možná rôzna úroveň všeobecnosti, od individuí až po všeobecné koncepty. V kapitole o tvorbe kategórii sme už analyzovali niektoré pohľady na podstatné mená. Meno kategórie slúži na pomenovanie inštancií v rámci kategórie a rovnako slúžia aj ako koncept, reprezentujúci danú triedu, čím určuje spôsob organizácie znalostí.

Nech „vlčiak“, „chrt“, „pes“ sú rozlišovacie kritériá. V niektorých atribútoch sú veľké rozptyly hodnôt, ako napríklad vo farbe srsti, avšak atribút vyjadrujúci počet nôh má nulový rozptyl. Koncept označený ako „Ferdo Mravec“ má nulový rozptyl vo všetkých atribútoch (pokiaľ neberieme do úvahy nepodstatné atribúty alebo len malé zmeny), reprezentuje individuum.

5.4.2 Prídavné mená

Rozlišovacie kritériá dovoľujú túto skupinu slov rozdeliť pomocou dvoch rôznych spôsobov reprezentácie konceptov. Prvým spôsobom je reprezentovanie konceptov kvantitatívnymi kritériami. Tie sú citlivé na rozpoznanie veľkosti rozdielu atribútov jednotlivých inštancií. Koncept operenosti v má v tomto prípade veľký rozptyl v atribútoch okrem atribútu reprezentujúceho pokrývku tela, ktorá je viacmenej podobná vo všetkých inštanciách. Iným konceptom s kvantitatívnym rozlišovaním je označenie malý. Môže znamenať poukázanie na jeden z menších objektov spomedzi triedy podobných objektov alebo taktiež na objekt, ktorého veľkosť je nezvyčajne malá vzhľadom na doteraz videné inštancie.

Druhý spôsob reprezentácie je pomocou kvalitatívnych kritérií, ktoré sú citlivé na znamienko rozdielu atribútov. Pomocou nich je možné zachytiť vzťahy ako ako napríklad menší.

Označenie „vodný živočích“ odkazuje na typ zvieratá žijúceho vo vode. Pokiaľ nie je priamo atribútom určené a rozlíšiteľné o akého živočícha sa jedná, tak v tomto prípade by bolo dobré rozlišovať na základe prirodzeného

zoskupenia atribútov definujúceho vodné živočíchy tak, ako pojednávala kapitola 3.1 o prirodzenej základnej úrovni kategorizácie.

5.4.3 Príslovky

Príslovky, vyjadrujúce koncepty, ktoré rozlišujeme v modeli rozlišovacích kritérií pomocou kritérií vzťahov. Na tomto mieste iba dodáme, že sa budeme ďalej zaoberať výlučne príslovkovými určeniami miesta ako napríklad vo vete "Mačka vedľa psa.", pričom ich budeme označovať ako príslovky.

5.4.4 Slovesá

Rozlišovacie kritériá citlivé na veľkosť rozdielu atribútov v dvoch po sebe nasledujúcich časových krokoch vieme použiť na reprezentácie konceptov ako napríklad letí. Kvalitatívna verzia umožňuje reprezentovať koncepty ako napríklad narástol.

Komplexnejšiu sémantickú reprezentáciu slovies ukotvenú vo svete agenta by bolo možné vybudovať cez pozorovanie dynamiky vzájomných interakcií dvoch agentov alebo objektov. Na každú interakciu z pohľadu dynamických máp [4] sa môžeme pozrieť z hľadiska troch fáz popisujúcich ich dynamiku. Pred, počas a po vzájomnom kontakte dvoch objektov. Každú fázu môžeme popísať pomocou vzdialenosti objektov, vektorov ich rýchlostí a ich energie. Kvalitatívne môžeme rozlíšiť štyri typy interakcie podľa toho, či sa oba objekty k sebe približujú, vzdávajú, alebo jeden ten druhý nasleduje. Pomocou takéhoto modelu, je možné reprezentovať významy slovies ako napríklad dotknúť sa, potlačiť alebo kopnúť.

6 Návrh výpočtového modelu akvizície jazyka

Naším modelom sledujeme širší cieľ vytvoriť metódu na tvorbu špeciálnej konštrukčnej gramatiky fungujúcej na základe princípu skúmania vzťahov slov tak ako ich vidíme pomocou hypotézy slovesného ostrova.

6.1 Špecifikácia cieľa

Deti používajú slová na komunikáciu o predmetoch, ktoré používajú, ich vlastnostiach alebo aj činnostiach, ktoré s nimi vykonávajú. Náš výpočtový model skúmajúci a napodobňujúci jazykový druh komunikácie sa snaží o vytvorenie umelých agentov, ktoré sú schopné pristupovať k jazyku a vnímať prostredníctvom neho svet podobným spôsobom ako to robí človek. V kognitívnom modelovaní je svet priamo vnímaný pričom vieme akým spôsobom niektoré výrazy interpretovať.

Predpokladom nášho modelu je, že základným mechanizmom uchovania sémantiky, je schopnosť rozlišovať a na základe toho si utvárať vnútorný mentálny priestor významov. Toto rozlišovanie pre agenta zabezpečuje možnosť ukotvenia komunikáciou prenášaných symbolov reprezentujúcich kontextom podmienené významy.

Hlavným prínosom implementačnej práce je vytvorenie štruktúry uchovávaní významov pre slovné výrazy s gramatikou. Správnou kombináciou významov slov gramatickými pravidlami sa význam celej vety interpretuje. Dôraz je kladený na akvizíciu jazyka bohatého na druhy konceptov vyjadriteľných podstatnými menami a prídavnými menami. Použitie prísloviak je zaujímavé hlavne z gramatického hľadiska. Podkladom pre uchovanie významu výrazov je snaha o zachytenie ich skutočnej funkcie v pomenúvacej slovnej hre založenej na princípe získania porozumenia ukazovaním a slovným označovaním predmetov. Z tohto hľadiska majú preto príslovky v našom modeli vlastnosť, že menia referenta pri označovaní.

6.2 Jazyková prax v pomenúvacej hre

Agent učiteľ spolu so žiakom je situovaný v prostredí scény, na ktorej sa nachádzajú rôzne objekty. Pod scénou rozumieme množinu týchto objektov, ktorá nieje usporiadaná. Objekty sú popísané pre agentov pozorovateľnými vlastnosťami, vyjadrenými v tvare dvojice atribút-hodnota. Agent dokáže rozlišovať medzi atribútmi a ich kvantitatívnou analýzou na vzorke pomocou aparátu rozlišovacích kritérií je tento agent schopný vytvárať prototypové kategórie.

Učiteľ a žiak spolu hrajú slovnú hru nazývanú pomenúvacia. Učiteľ vyslovuje slovné spojenia v prirodzenom jazyku, ktoré svojím významom referujú, t.j. poukazujú, na objekty na scéne. Žiak má za úlohu tento objekt pre príslušný výraz jednoznačne identifikovať - ukázať na neho. Pomenovanie objektov, ktoré sú si podobné môže zapríčiniť nejasnosti vo vyhodnocovaní jedinečného referenta a potom je potreba ukazovať na viac objektov naraz. Tento prístup však marí účel hry, pretože v nej ide o slovné identifikovanie objektov správnymi výrazmi tak aby nevznikali pochybnosti o objekte, o ktorom sa hovorí. Požadujeme teda od pomenovania objektu dostatočnú špecifickosť vzhľadom na ostatné prítomné objekty.

Pod úspešnou hrou rozumieme správne identifikovanie objektu, na ktorý výraz referuje. Pokiaľ výraz referuje na viac objektov, žiaden alebo na nesprávny, hra je neúspešná.

Takáto učebná metodika zabezpečí vznik porozumenia medzi agentami pri úspešnej hre. Ak sú na scéne dva podobné objekty, predpokladáme, že existuje pre každé z nich jedinečné pomenovanie. Potom schopnosť rozlíšiť tieto objekty pomocou jazyka nevyžaduje iba schopnosť rozlišovať objekty podľa vlastností, ale tiež schopnosť rozlišovať doslova podľa toho, ako sa o nich hovorí, respektíve aké jazykové nástroje sa používajú pri referovaní na ne.

V prípade prenosu porozumenia si úspešný žiak potvrdzuje správnosť pochopenia výrazu a má možnosť aktualizovať svoje jazykové znalosti na základe tejto novej inštancie. V opačnom prípade dostane negatívnu spätnú

väzbu a neučí sa nič.

Komunikácia medzi agentami má teda jednoznačné pravidlá a primárne slúži na výuku "nemého" porozumenia od žiaka. Znalosti vyžadované na úspešné absolvovanie hry musia však zahŕňať vhodné prepojenie medzi jazykom zakódovanými vetami a kategóriami objektov zachytenými štruktúrou rozlišovacích kritérií. V takejto lingvistickej štruktúre je preto očakávateľné, že je možné zkonštruovať pomenovanie pre ľubovoľný objekt (referent na scéne).

Naučenie sa slovnej hre neznamená len vystupovať v pasívnej roli žiaka ukazovateľa, ale byť aj schopný prebrať úlohu učiteľa a učiť, či hrať sa slovné hry s nasledujúcimi generáciami agentov. Rozvíjajúca sa schopnosť pomenovať objekty je u žiaka základom pre splnenie podmienky jedinečnosti referenta.

Pozrime sa bližšie, čo vlastne očakávame od učiteľa pomenúvajúceho jeden objekt na scéne. Predstavme si, ako by zvládol takúto úlohu človek, čo sa snaží vhodne popísať objekt, na ktorý myslí. Zjavne musí preskúmať aj okolité predmety aby našiel tie odlišnosti, ktoré sa jazykom bežne nejak o označujú. Musí dávať pozor, aby jeho výpoveď bola interpretovaná tak, ako ju zamýšľal. V našom prípade umelých agentov je táto požiadavka na "inteligentné" pomenovanie zaťažaná rovnakými problémami. Agent musí vedieť skontrolovať, či vyslovené slovné spojenie je naozaj správne, vedieť ho upresniť alebo sa rozhodnúť pre iný spôsob pomenovania objektu. Je treba si uvedomiť, že hľadáme čo najjednoduchšie jedinečné pomenovanie. To znamená, že každé slovo v takejto vete má dôležitý význam pri určení referenta. Na základe tohto faktu sa možno domnievať, že pre každý takto pomenovaný objekt existuje iný objekt, ktorý mu je pri nepochopení jediného slova veľmi podobný, ba je totožný. Nazvem tento nový objekt ako podobný ⁹. Jeho existencia je nutná na naučenie sa slovnej hry pretože pri neúplnom pochopení vety odvádza žiakovu pozornosť na seba. Takto je zabezpečené, že nenastane aktualizácia významu nesprávnou inštanciou a nespôsobí agentovi

⁹bias object

nesprávnu jazykovú skúsenosť.

Keďže táto hra vystupuje ako analógia komplexného učenia sa jazyka u detí, a preto v návrhu algoritmu preberieme niektoré metodologické praktiky a predpoklady rozobraté v kapitole 2 tvoriacej teoretické pozadie. Učenie sa jazyka zvyčajne prebieha v určitom zmysle celkom kontrolovane. Malé dieťa sa učí jazyk, ktorý má jednoduchšiu formu postačujúcu na interakciu zo svetom. Postupné dospievanie obohacuje dieťa o nové jazykové formy, ktoré stavajú na predošlom porozumení jednoduchších foriem. Nové formy vznikajú na základe nových odlišných slovných hier ako aj na hlbšom porozumení usporiadania sveta.

Náš agent učiteľ môže preto vo výuke neznalého agenta postupovať kontrolovane, aby zabezpečil rýchlu a spoľahlivú naučiteľnosť. Postup učenia by mal byť postupný a primerane intuitívny žiakovým schopnostiam. Určitá jazyková intuícia v podobe obmedzujúcich predpokladov na význam výrazov je očakávaná a pri výuke agentov s ňou preto pracujeme.

Po dokončení výuky agenta je možné ho postaviť na základe predpokladu o jeho naučených schopnostiach do roly nového učiteľa ďalšej generácie agentov a teda opätovnému prenosu významu v medzigeneračnom procese iterovaného učenia. Očakávame, že tento jazykový prenos bude stabilný.

6.3 Popis experimentu

Pozrime sa teraz už konkrétne na našu inštanciu pomenúvacej hry, od ktorej očakávame, že zabezpečí prenos jazykovej zručnosti z agenta na agenta.

Objekty sú v tomto experimente zvieratá. Ich vlastnosti sú zakódované pomocou konceptov ako veľkosť, počet nôh alebo dĺžka chvosta. Na zvieratá môžeme odkazovať pomenovaním ako "pes" alebo všeobecnejším výrazom ako "cicavec". V tomto prípade považujeme za vhodné také obmedzenie významu, že jednoslovné pomenovania považujeme za podstatné mená. Vo výuke budeme najskôr postupovať využívaním iba takýchto výrazov, pričom všeobecnejšie kategórie budeme učiť až po dobrom natrénovaní tých základných.

Ak chceme zachovať pravidlo hry o výlučnosti pomenovania, treba očakávať, že rôznorodosť objektov scény musí byť spravidla vyššia pri jednoduchých vetách.

Pre dlhšie slovné spojenia prichádzajú do úvahy označenia ako napríklad "lietajúci vták/flying bird", čiže použijeme prídavné meno na poukázanie na vlastnosť určitej skupiny objektu.

Výraz "vták nad rybou/bird over fish" dovoľuje rozlišovať niektoré intuitívne priestorové pojmy. Tu už však má použité príslovkové určenie miesta význam prepojenia medzi dvoma objektmi, určenie ich vzťahu, konkrétne polohy referenta jedného z objektov.

Informácia o slovnom druhu slova má v týchto prípadoch dôležité postavenie v určení gramatickej stavby vety. Pri takomto pomenovaní objektov je očakávaná predikát subjektová štruktúra a kompozičný jazyk so základnou sadou slovných druhov, ktorými sú podstatné meno, prídavné meno a príslovkové určenie miesta (pre jednoduchosť budeme nazývať príslovkou).

Druh slova obmedzuje jeho význam. Špecifikuje, akým spôsobom pracovať s inštanciami perceptuálneho vstupu patriacimi tomuto slovu. Je kľúčom k určeniu vzťahov medzi objektmi alebo určeniu vlastností, pretože jednotlivé významy slov sú implicitne správne spracovávané pri interpretovaní slovného spojenia.

Gramatická štruktúra vety je chápaná pomocou použitého druhu konštrukcie, v zmysle určitej zoradenej postupnosti kategórii slov ¹⁰ so špecifickým prístupom k ich významom, ktorými prispievajú k spoločnému významu vety.

Na základe tejto vnútornej reprezentácie je agent schopný správne identifikovať referované objekty zo spôsobu ich označenia. Vnútorne významy však nie sú pevne naučené a zviazané s jednotlivými slovami. Význam slov sa tvorí na základe kontextu ostatných slov vo vete a tiež je závislý na kontexte situácie ¹¹. Vo vyššie uvedenom príklade referenta pre pomenovanie

¹⁰Kategóriou je slovný druh

¹¹V tomto prípade jednoducho iba od všetkých rozlíšiteľných objektov na statickej scéne

”veľký vták”, je referent identifikovaný spomedzi rôznych zvierat nájdením veľkého vtáka, takého, ktorý je dostatočne veľký spomedzi ostatných nato, aby mu toto označenie veľkosti bolo možné prisúdiť.

Členenie zvierat v hierarchii ako napríklad zaradenie tuniaka medzi ryby slúži na skrytie priameho označenia za všeobecné a následného testovania schopnosti modelu správne klasifikovať. Sémantickú štruktúru znalostí je preto nutné obohatiť o možnosť zoskupenia slov označujúcich koncepty do hierarchií.

Učiteľ v experimente generuje vety na základe pravidiel pevne daného algoritmu, navrhnutého pre jednoduchú formu anglického jazyka. Vytvára scénu spolu s pomenovaním. Pridáva postupne objekty s vlastnosťami alebo popisuje vzťahy medzi objektmi. Robí to navyše tak, že spĺňa podmienku výlučnosti, ale tiež generuje podobný (bias) objekt, ktorý kontrastuje s referentom určitým špecifickým spôsobom (vlastnosťou, pozíciou).

Žiak sa učí poznávať okolitý svet zvierat spolu s jeho slovným popisom a je schopný uspieť v pomenúvacej hre, správne identifikovať odkazované zviera a byť schopný preniesť túto znalosť na ďalšie generácie.

6.4 Náčrt interpretácie priebehu simulácie

V tejto kapitole predvedieme interpretáciu začiatku jedného behu simulácie popísaním jej logovacieho súboru. Každá epocha začína vypísaním hlavičky, v ktorej identifikujeme číslo epochy, v niektorom kole a číslo generácie. Kolem nazývame niekoľkoepochový priebeh slovnej hry, taký že sa používajú slovné prostriedky určitého druhu.

— Epoch: 1 / 300, Generation: 1, Round: 1 (Teacher: rule-based, SpeakersLevel: 1) Scene: 1=1.furry: 2.0; id: 3.0; legs: 8.0; posX: 87.0; posY: 85.0; ravenous: 2.0; size: 1.0; teeth: 2.0; , 2=2.aquatic: 9.0; fin: 7.0; id: 6.0; posX: 78.0; posY: 88.0; ravenous: 2.0; size: 8.0; teeth: 3.0; , 3=3.aquatic: 9.0; beak: 1.0; fin: 9.0; id: 4.0; posX: 95.0; posY: 5.0; ravenous: 2.0; size: 12.0; skinny: 9.0; teeth: 4.0;

Na počiatku sú vytvorené dva agenty, pričom prvý je pravidlami ria-

dený agent, ktorý vždy na začiatku epochy vygeneruje scénu tak, aby bolo možné každý objekt jednoznačne popísať jazykom, ktorý používame pri tejto hre. Je to práve tá prvá scéna, ktorú vidíme zo záznamu hry v prvej a potom v každej nasledujúcej epoche tejto slovnej hry. V prvej sa scéna skladá zo štyroch objektov, popísaných atribútmi. Je dôležité povedať, že poradie vyjadrené číslom identifikujúcim objekt, nehrá žiadnu úlohu pri úlohách nájdenia popísaného referenta. Takže aj keď scéna bude vždy usporiadaná tak, že referent je na prvom mieste, agent žiak o tejto informácii nič nevie a nemá možnosť ako ju zachytiť. Táto vlastnosť softvéru len uľahčuje orientáciu pri pochopení záznamu.

V ďalšom kroku je agent žiak vyzvaný, aby pomenoval každý z objektov zo scény. Pochopiteľne, keďže v prvej epoche žiadnymi lingvistickými znalosťami nedisponuje, je jeho odpoveď prázdna. Následne nato, agent učiteľ udá svoj popis. Keďže v prvej generácii máme pravidlami riadeného učiteľa a ten si pri generovaní scény vytvoril správny popis pre jeden z objektov, teraz nám ho iba ukáže. Samotný žiak si túto informáciu nezapamätá, slúži iba pre meranie úspešnosti.

V nasledujúcej časti prebieha hádanie referenta. Učiteľ vysloví meno objektu a žiak háda, ktorý to je. V prvej generácii to bude vždy ten jeden učiteľov názov pre objekt, pomocou ktorého konštruoval scénu. Odhliadnuc od úspešnosti porozumenia, prebehne potom učenie sa vysloveného výrazu na označenie referenta. V tomto prípade slova žaba (frog), referujúceho na prvý objekt. Učenie spôsobuje aktualizácie vrcholov v sémantickom grafe. V zázname vidíme aj pravdepodobnosť zaradenia slova do slovného druhu: frog-Noun-1.0. Jediné podstatné meno sa aktualizovalo použitím učiteľovho referenta. Pre overenie správneho naučenia sa a schopnosti interpretovať naučenú znalosť prebieha opätovné preskúšanie určenia rovnakého referenta. Úspešnosť rozhodne o ponechaní získaných znalostí spolu so zvýšením hodnotenia pre použitý konštrukt a zaradenia slov do daného druhu. Neúspešné absolvovanie tohto testu však spôsobí spätný chod všetkých aktualizácií a potrestanie konštrukt spolu zo zaradením slov do druhu. Konštrukt je for-

ma gramatiky definujúca slovné druhy, ktorá funguje ako prepisovací systém (formálna gramatika) ale obsahuje aj informácie o referentovi a úspešnosti konštrukcie (viac o tom v ďalšej kapitole).

Ako vidíme interpretácia naučeného v prvej epoche prebehla úspešne. Následne sme však požadovali, žiak sám skúsil popísať scénu, ale neodpovedal. V tomto prípade to považujeme zatiaľ za žiakovu plachosť, neskôr sa prejaví a bude schopný pomenovať¹². Ďalšie epochy prebiehajú obdobne a ako sa môžeme presvedčiť aj úspešne, agent sa učí názvy zvierat a až v 54. epoche sám prvý krát správne pomenuje objekt. Do konca epochy zdokonaluje svoje porozumenie a úspešne pomenúva a vie najstť pomenovanie referenta.

V ďalšom kole sa namiesto konkrétnych názvov zvierat používajú aj opisy pomocou všeobecných kategórii. Keďže je žiak už vytrénovaný v pomenovaní zvieratá, zareaguje správnym popisom. Ako si môžeme všimnúť, scénu tvoria rôzne zvieratá, a učiteľ opíše slovom "vták/bird" objekt 1201, ktorý je naozaj vtákom, a navyše, na scéne iného vtáka niet. Učiteľ sa správa korektne, vytvoril pedagogicky vhodnú scénu na učenie tohto konceptu. Všetko je zdá sa v poriadku a žiak sa úspešne učí, ale niekde po 50. epoche začne jeho schopnosť udržať si koncepty z tejto všeobecnej kategórie rapídne klesať. Môžeme si všimnúť, že niektoré kategórie mu nerobia problém a niektoré, ako napríklad zvieratá cicavce, sa jednoducho nedokáže správne naučiť (príliš veľa rôznych inštancií). Spočiatku pojmom aj rozumie a vie ich dokonca použiť. Postupom času však agent získa príliš veľa odlišných konceptov všeobecnej kategórie zachytených jedným rozlišovacím kritériom.

6.5 Návrh modelu akvizície jazyka a jeho implementácia

Budem teraz popisovať časť programu, ktorú som doplnil do programového balíka Jarmila [17] v jazyku Java. Pojmom simulácia budem označovať sa-

¹²V skutočnosti pomenuje až keď bude mať nadprahovú mieru istoty

motný programový priebeh výmeny významov pre popis scény, alebo hádanie referenta, ktorý sa deje podľa pravidiel pomenúvacej hry ako sme ju voľne predstavili.

Do interakcie vstupujú dva agenty, pričom jeden bude môcť vymyslieť scénu, t.j. vygenerovať objekty s vlastnosťami a v nej pomenovať aspoň jeden objekt. Agent definovaný triedou `BootstrappingAgent` učiaci sa jazyk generovaný učiteľom triedy `RulebasedTeacher` si vytvára význam pre slovné popisy sémantickou reprezentáciou pomocou rozlišovacích kritérií.

Pôvodná verzia neobsahovala pre agenty možnosť práce s viac-slovnými výrazmi, pretože neboli vybavené žiadnym gramatickým aparátom. Obohatil som detaily hry ako som popísal v predošlých kapitolách, tak aby som zabezpečil prenos kompozičného jazyka bez dopredu známej gramatiky. Sémantika rozlišovacích kritérií momentálne nemá schopnosť odolnosti voči nesprávnej klasifikácii pozitívnych príkladov a preto bolo nevyhnutné obmedziť nepochopenie medzi agentami, ktoré by viedlo k deštruktívnej aktualizácii stabilných významov novou nesprávnou inštanciou.

Nejednoznačnosť referenta pri učení sa nových významov obzvlášť narastá v súvislosti s využívaním kompozične vytvorených významov. Jednotlivé kombinácie významov môžu referovať na viac objektov. Výsledný referent získaný kombináciou určenou pravidlami významu prísloviak je však to jediné, čo pri učení agent od svojho učiteľa dostane. Túto informáciu použije pri aktualizácii rozlišovacích kritérií pomocou `getConstructReferent()` z triedy `SemSynInference`. Táto funkcia zabezpečuje svojimi parametrami ovplyvňujúcimi logiku schémy programu 3 metódy. Hodnoty funkcie pre všetky metódy sa počítajú rekurzívne a môžu sa navzájom v priebehu rekurzie používať ako ľubovoľná z metód.

Učenie sa výrazu prebieha podľa známeho konštruktú, tak nazývame gramatickú informáciu uchovávanú v triede `Construct`. Táto informácia je dôležitá hlavne preto, lebo obsahuje rekurzívne definovanú postupnosť slovných druhov. Obsahuje tiež ukazovateľ na časť formy, ktorá priamo predstavuje referenta tejto gramatickej formy. Je ňou buď podstatné meno alebo

rekurzívny odkaz na inú formu.

Zoberme si situáciu pri učení s učiteľom, ktorý pre pomenovanie určí referenta na scéne. Žiak pre tohto referenta nájde konštrukt, ktorý môže na dané slová použiť. Potom prebieha rekurzívne parsovanie prvkov ¹³ formy konštruktú do hĺbky a posúvaním indexu momentálne spracúvaného slova, s ktorým prvok formy korešponduje. Posledné volanie v rekurzii `getConstructReferent()` je ukončené terminálom v podobe podstatného mena ¹⁴. Zistili sme teda, že učiteľom ukazovaný predmet je pomenovaný týmto slovom, ktorého rozlišovacie kritérium vieme hľadať v báze znalostí. Môžeme mať istotu, že referent je správny, lebo ho ukazuje učiteľ a preto vrchol aktualizujeme.

Nato nastáva druhá fáza, kde sa postupne vynorujeme z rekurzie, ale zároveň parsujeme zvyšné slovné druhy postupne na každej úrovni rekurzie. Vzťahy platiace medzi týmito slovami nám bližšie špecifikujú význam slovného spojenia a preto môžeme aktualizovať významy pre jednotlivé slová. Spôsob aktualizácie závisí tiež na interpretovanej kombinácii s ostatnými slovami, ale stále len s tými na spoločnej úrovni rekurzie. Napríklad prídavné meno sa aktualizuje tak, že ak bol na tejto úrovni rekurzie referent terminálny NOUN, aktualizuje sa aj hrana medzi týmito slovami a zabezpečí sa zachytenie konkrétnejšieho významu.

Príslovka prepája referent aktuálne spracúvanej formy konštruktú s tým, ktorý je vyššie v rekurzii. Tento nový referent pri vynorovaní z rekurzie je najskôr určený buď priamo, ak je terminálom NOUN alebo rekurzívnym zavolaním na konštrukt, ale s parametrami určujúcimi vykonávanie metódy na zistenie neznámeho referenta zo scény.

Oba získané referenty sú vyhodnotené pomocou `ApplyAdverb()`, či sú pre danú príslovku aktívne. ¹⁵ Ak áno, nový referent môže slúžiť na (rekurzívne) aktualizovanie významov pre príslušné slová, ktoré preparsoval pri

¹³Triedy Entity

¹⁴NOUN

¹⁵V skutočnosti sa porovnávajú dvojice z dvoch množín potenciálnych referentov a zistia sa najlepší, s nadprahovou aktiváciou

jeho zisťovaní. Ak však niečo zlyhá, nič sa neaktualizuje, aby sme zabezpečili dobrú odolnosť voči chybné určeným referentom slovného výrazu.

Posledný spôsob použitia `getConstructReferent()` umožňuje agentovi pomenovať referenta, to znamená vygenerovať slová pre určeného referenta určeným konštruktom.

Funkcie agenta, ktoré používajú `getConstructReferent()` sú nasledovné. `LearnUtteranceWithSyntax()` naštartuje nájdenie najlepšieho konštruktu pre vstupné slová a potom spustí hľadanie referenta určeného učiteľom, pričom sa aktualizuje aj báza sémantických znalostí. Funkcia `referentsOfUtterance()` nájde objekt, na ktorý výraz referuje. Funkcia `name()` zvolí bližšie špecifikovaný konštrukt, ktorým sa rozhodne pomenovať a potom pre daného referenta vygeneruje slová.

`BootstrappingAgent` obsahuje metódy na participovanie v pomenúvacej hre. Vie identifikovať použitú gramatickú formu vety, nájsť jednotlivé slová v grafe bázy sémantických znalostí a potom pomocou metód v triede `SemSynInference` interpretujúcich použité významy nájsť referenta na scéne postupným aplikovaním rozlišovacích kritérií, zužovaním množstva možných referentov. Pri úspešnej hre si ponechá aktualizované kritériá v grafe a zvýši pravdepodobnosť zaradenia slov do týchto druhov, ako aj skóre konštruktu nesúceho gramatickú špecifikáciu vety. Týmto spôsobom nadobúda všetky svoje lingvistické znalosti. Technika prístupu k učeniu sa jazyku pomocou využívania takto špecifickým spôsobom navrhutej sémantiky by sa dala porovnať k postupnému zaväzovaniu si šnúrok pri obúvaní (bootstrapping). Techniky sú založené na postupnej kombinácii syntaktickej informácie so sémantickou pri dedukovaní jednotného významu.

Vstupná veta je interpretovaná ako usporiadaná a uzátvorkovaná množina vrcholov a hrán z grafu sémantických znalostí.

Pre vstupné slová môžeme nájsť viac za sebou nasledovných konštruktov. Potom neparsované slová sú cez `applyConstruct()` doplnené a teda, ak je to potrebné, vytvorí sa nový konštrukt s iniciálnym skóre a je zaradený do slovníka a stáva sa kandidátom na správnu gramatiku v procese samoorga-

nizácie. Agent však zamietá neintuitívne kódovania gramatiky overovanej volaním `isValidForm()`. K takýmto nesprávnym konštruktom zaraďuje tie, ktoré sémanticky nevie interpretovať. Konkrétne, ak sa vo vete vyskytuje príslovka a iba jedno podstatné meno alebo dve podstatné mená a žiadna príslovka.

Zaradenie slov do slovných druhov ako aj správna gramatická forma sú nadobudnuté samoorganizáciou priamim posilňovaním alebo tlmením úspešnosti, skóre, podľa toho, či slovná hra prebehla úspešne. Snažíme sa teda o maximalizáciu úspešnosti zaradenia slov do slovných druhov maximalizáciou premennej `scoreMax` opakovaným volaním funkcie `parse()` funkciou `parseBest()` pre rôzne konštrukty (`c`). Pomocou volania `wordtypesMatch()` maximalizujeme `scoreMax` určené takto:

$$c.score \cdot \prod_{i=0}^{c.wordtypesCount} wordtypeProbability(words[i], c.wordtypeList[i]).$$

`BootstrappingAgent` potom volá `praise()` pre posilnenie novej znalosti a v opačom prípade volá `punish()`.

Nesprávne zaradenie slova do slovného typu spôsobí, že význam slova sa nesprávne interpretuje, čo zapríčiní, že podobný (bias) objekt pre objekt skutočného referenta zmarí požadovanú identifikáciu. Nesprávna gramatická forma spôsobí obdobné problémy.

Nevhodný spôsob prepojenia významov spôsobí zlyhanie korektnej identifikácie. Treba si však uvedomiť, že zaradenie slova do druhu nieje pevné a preto treba zvážiť ako posudzovať, či gramatická forma je vhodná pre tieto slová. Preto pri priradovaní slov do formy pracujeme vždy len s určitou mierou istoty. V modeli sú však implementované prahy istoty zaradenia slova do druhu, alebo jeho vylúčenia.

V triede `Parameters` sú zhromaždené konštanty parametrizujúce priebeh simulácie. Špecifikujú prostredie scény, ale ovplyvňujú najmä správanie sa agenta. Priebeh jazykových hier medzi dvoma agentami zabezpečuje trieda `Simulator`. Metóda `simulate()` spúšťa rôzne experimenty tak, že volá metódu `simGeneration()` s vhodnými parametrami ako napríklad počet epoch alebo maximálna dĺžka vety učiteľa.

Prostredie scény sa generuje pomocou pravidlami riadeného agenta pomocou funkcie `newObjectDescription()`. Podľa požadovanej zložitosti jazyka takýto agent náhodne vygeneruje výraz spolu so samotným objektom. Práca s objektmi špecifikovanými vo vstupnom súbore `animals.csv` sa spracuje v triede `KB`. Tu sa nachádzajú všetky funkcie na získavanie vlastností `getObjectAttr()`, mien a kategórií, ako aj porovnávanie vlastností `compareFeature()` alebo samotné generovanie `makeObject()` a modifikovanie objektov `modifyObject()` využívané na zväčšenie a zmenšenie objektov zapísaných rámcami.

Pravidlami riadený učiteľ v triede `RulebasedAgent` plní jedinú úlohu, generuje objekty na scéne tak aby jeden z objektov bol referentom výrazu, ktorý zároveň konštruuje. Nevie pomenovať ľubovoľný objekt. Je však využívaný ako generátor scény pre ďalšie generácie, pretože scéna, ktorú vygeneruje má už spomínané potrebné vlastnosti jedinečnosti referenta a existencie podobného objektu. Objekty na scéne sa teda nepridávajú úplne náhodne. Keď sa podľa pravidiel vygeneruje objekt pomocou `newObjectDescription()`, zároveň sa vytvára aj podobný objekt s podobným názvom, ktorý má rozptyľovať agenta pri určení referenta. Napríklad, keď sa agent začína učiť príslovky, je dôležité tam mať objekt, ktorý nespĺňa predikátovú podmienku tejto príslovky, aby vždy existovala dvojznačnosť pre toho, kto príslovke nerozumie.

6.6 Parametre simulácie

Trieda `Parameters` určuje parametre simulácie a teraz sa bližšie pozrieme na tie najdôležitejšie. Uvádzané hodnoty boli stanovené heuristicky a je možné ich meniť.

Parameter `useCategories` určuje, či použiť pre objekty mená kategórie pravidlami riadeným agentom pri učení v kole 3 a 4. Nezávislým parametrom `learnCategories` sa určuje, či v druhom kole učiť mená kategórií (bez ohľadu na to, či sa neskôr budú používať vo vetách).

Iniciálne je pre novovzniknuté konštrukty definované skóre úspešnosti `init_construct_score` a úspešnosť zaradenia slova do slovného druhu

init_wordtype_score. Obe sú nastavené na hodnotu 0.5, pretože obe hodnoty fungujú na intervale $<0,1>$ a teda táto hodnota má rovnako ďaleko k obom okrajom intervalu.

Samoorganizáciou aktualizujeme úspešnosť konštruktú vynásobením konštantou **praiseConstruct** a zaradenie do druhu konštantou **praise-Wordtype**. Obe sú nastavené na 1.03. Iné nastavenie má význam iba v prípade potreby zmeniť rýchlosť nájdenia správnej gramatiky alebo slovného druhu, ale hrozí riziko nesprávneho určenia. Konštrukt navyše, ak je hodnota konštanty **ConstructScoreRecursionMultiply** väčšia ako 1, aktualizuje skóre podkonštruktov s redukovanou váhou.

Pomocou konštanty **discard_wordtype_threshold** sa snažíme zamietnuť nesprávne slovné druhy nespĺňujúce danú mieru istoty. Stačí, aby bola dostatočne menšia ako iniciálne skóre 0.5. Používame hodnotu 0.25. **WordtypeCertainThr1** naopak určuje, kedy sme si zaradením istý a toto slovo nepoužijeme inak. Nevylučuje to však, že sa zaradenie môže neskôr zmeniť. Používame hodnotu 0.7.

Parameter **SIM_THR**, ktorý bol dôležitým parametrom v simulácii predgramatickej akvizície jazyka, sa v tomto prípade stáva nie príliš významným. Zabezpečuje totiž prah aktivácie rozlišovacích kritérií, ale keďže agent pri hľadaní a pomenovaní referenta berie do úvahy len najlepšie kritérium, ostatné menej aktívne aj tak zahodí. Výnimka nastáva pri interpretácii príslovky ako je uvedené v poznámke pod čiarou pri analýze funkcie **ApplyAdverb()**.

Pri príslovkových pomenovaniach je otázne, či je vo vete referentom prvé alebo druhé podstatné meno, pokiaľ sa to agent ešte nenaučil. Konštanta **bestReferentVariant** určuje, ktorého referenta predpokladať.

Konštanta **notBestReferentVariantPunish** znižuje skóre konštruktov pri pridávaní do slovníka. V simuláciách predpokladáme referenta v druhom substantíve, takže je na agente, aby samoorganizáciou zistilo chybu a prevážili konštrukty s referentom na prvom substantíve tak, ako to generuje učiteľ.

6.7 Charakteristika takto naučeného jazyka

Nemáme veľké očakávania o sile takejto jednoduchej jazykovej formy v porovnaní s bohatosťou prirodzeného jazyka. Pozrime sa predsa len podrobnejšie, aké sú obmedzenia a nakoniec aj vhodné vlastnosti tohto modelu.

V prvom rade je dôležité uviesť, že spôsob organizácie významov v grafe umožňuje homonýmiu iba pokiaľ ide o slová rôzneho druhu¹⁶. Rovnaké slová rovnakého druhu sú v grafe zaznamenané ako vrcholy a v grafe sa vyhľadáva väčšinou práve podľa tejto dvojice slovo, slovný druh. Na druhej strane ale, významy pre slová sú ako napríklad v prípade prídavných mien zviazané s iným vrcholom a uložené v o vzájomnej hrane. Pre prídavné mená je preto prirodzene zabezpečená homonymia v tom zmysle, že agent sa význam prídavného mena učí vždy špecificky pre každé podstatné meno, s ktorým sa v slovnom výraze nachádzalo. Všeobecne však pravú homonýmiu nemôžeme rozlíšiť.

Synonymia v zmysle podobnosti alebo rovnosti významov pre rôzne slová, nám v takomto grafe prirodzene nerobí problémy. Stačí mať správny aparát na detekciu podobnosti významov. Vzhľadom na to, že spôsob práce s kategóriami rozlišovacích kritérií nie je ešte v modeli zahrnutý, ale zrejme súvisí s detekciou synonym, momentálne o synonymách nie je príliš veľa o čom hovoriť. Agentom synonymá nevytvárajú problém, iba o takto špecifikovanom prepojení slov nevedia.

Čo sa týka fráz a výrazov ako "dať hlavy dohromady" je náš model zatiaľ bezmocný. Kombinácie slov, ktoré niesú frazeologické však fungujú. Ak sa obmedzíme len na pomenovanie objektov v našej slovnej hre, jazyk flexibilne zachytáva popis a usporiadanie objektov. Stavba gramatiky umožňuje tvoriť neobmedzene dlhé vety, pričom agent pri tvorbe nových gramatík využíva kratšie formy, ktoré už pozná a osvedčili sa.

Problém nemožnosti iného ako jednoslovného pomenovania podstatným menom je len čisto technický, súvisiaci s implementáciou.

¹⁶Ako vieme, sémantické znalosti vo vrchoch tvoria dvojicu: slovo, slovný druh

7 Kvantitatívne výsledky simulácií a ich interpretácia

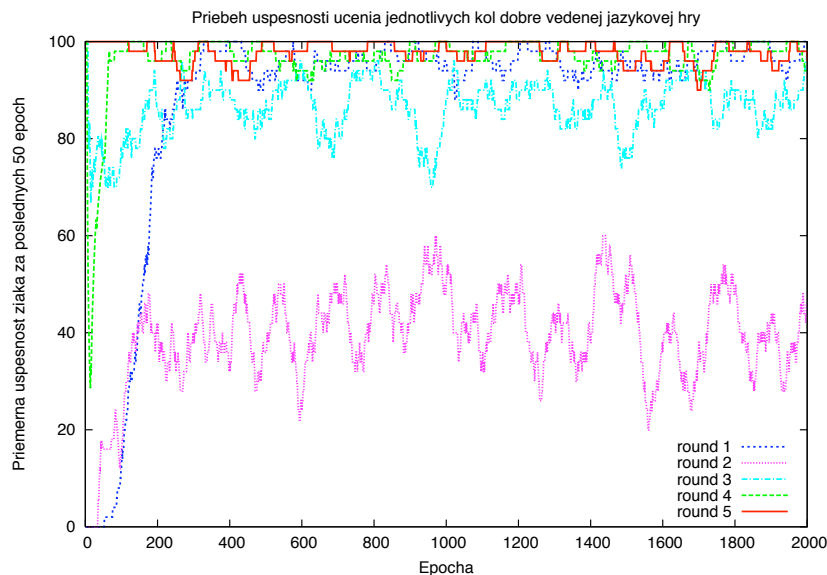
V priebehu simulácie sa zaznamenáva počet úspešných identifikovaní referenta pre pomenovanie učiteľom. Pod úspešnosťou, ktorá bude hlavným merítkom úspešnosti agenta obstať v hre rozumieme priemerné percento správne identifikovaných referentov za posledných 50 epôch, z jedného behu simulácie.

Učenie prebieha v kolách ¹⁷, ktoré sa navzájom odlišujú jazykovou výslovnosťou. Jedno kolo, tvorí niekoľko epôch. V prvom kole sa používajú iba jednoslovné pomenovania objektov, bez kategórií. V druhom kole sa pomenúvajú kategórie. V ďalších kolách sa s pravdepodobnosťou 1/2 využívajú pomenovania aj z predošlých kôl. V treťom kole pribudne pomenovanie jedným alebo dvoma prídavnými menami a vo štvrtom kole pribudne príslovkové určenie miesta. Piate kolo je iba pokračovaním štvrtého a v ňom už agent nie je učení žiadnu novú gramatickú formu.

Pozrime sa teraz na charakteristiku výsledkov z jednotlivých generácií.

¹⁷Rounds

7.1 Prvá generácia



Obrázok 2: Prvá generácia

Obrázok 2 ukazuje rôzne rýchlosti učenia sa v jednotlivých kolách. Ako vidíme prvé kolo začína z nuly, ale postupne vznikne stabilné, takmer vždy správne určenie referenta podľa podstatného mena.

Druhé kolo ukazuje, že kategórie, sa dokáže agent slabo určiť a v priemere v každom druhom príklade zlyhá. Tento výsledok sa však po ďalšej analýze ukázal ako pochopiteľný. Kategória zvierat ako napríklad cicavec, s ktorým mal agent najväčšie problémy, je totiž všeobecná, umelá kategória. Agent ju nemôže zachytiť v jedinom rozlišovacom kritériu, pretože cicavce sú reprezentované viacerými prirodzenými zhľukmi atribútov a ich spoločný prienik nám o cicavcoch nemôže nič povedať. Rozlišovacie kritérium pre cicavce vytvorené na základe inštanície veľryba a inštanície pes neposkytuje žiadny použiteľný význam.

Tretie kolo ukazuje mnohoznačnosť, ktorú prídavné mená vnášajú do pomenovaní. Atribúty sú generované v určitých intervaloch a preto je pre agenta ťažšie si takéto koncepty správne zapamätať, no stále mu to ide veľmi dobre. Pomenovania v tomto kole zpodovedajú podradenej kategórii ako

spojeniu adjektív so substantívom základnej kategórie. Tou skutočnou podradenou kategóriou je len označenie obsahujúce kategóriu zvieráťa. Iba v takomto prípade by prídavné meno vymedzovalo konkrétnu inštanciu spomedzi ostatných. Prídavné mená spolu s podstatnými zo základnej kategórie jednoducho vyjadrujú fakt o vlastnosti zvieráťa, pričom neexistuje rovnako pomenované zviera, ktoré túto vlastnosť nemá.¹⁸ Preto ak chceme byť korektní, na hovorenie o podradenej kategórii je nutné pomenúvať v tomto prípade všeobecnými kategóriami.

Ako môžeme vidieť, štvrté kolo je najúspešnejšie. Časté používanie prísloviak najspoľahlivejšie identifikuje objekt. Je tomu tak preto, lebo výraz referuje na dva objekty a určuje medzi nimi vzťah, čo učiaci sa agent ľahko rozlíši. Môže mať pochybnosti jedine vzhľadom na existenciu bias objektu. Ostatné objekty môže ľahko zamietnuť.

V súbore obsahujúcom slovník slovných druhov a konštruktov¹⁹ nájdeme v danom kole postupne tieto konštrukty s väčšou ako iniciálnou úspešnosťou. Forma konštruktu je vypísaná do hĺbky a táto rekurzia je zaznačená hranatými zátvorkami. V hranatých zátvorkách identifikujúcich konštrukt je číslo konštruktu za znakom # a skóre konštruktu za znakom =. Potom nasleduje rekurzívne vypísanie formy konštruktu. Referent výrazu je určený znakom * na každej úrovni.

1., 2.

[#0=1.00 Noun*]

3.

[#4=0.63 Adjective[#1=1.00 Adjective[#0=1.00 Noun*]*]*]

[#1=1.00 Adjective[#0=1.00 Noun*]*]

[#0=1.00 Noun*]

4.

[#4=0.61 Adjective[#1=1.00 Adjective[#0=1.00 Noun*]*]*]

[#1=1.00 Adjective[#0=1.00 Noun*]*]

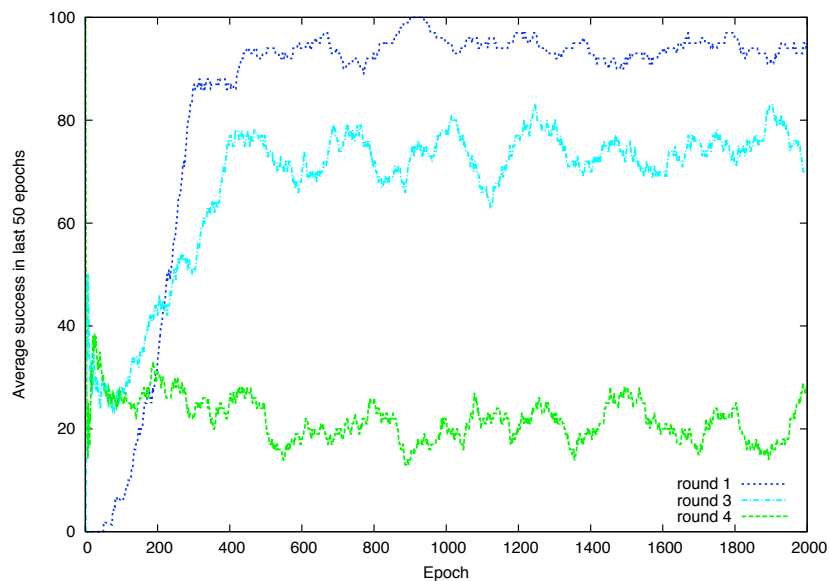
¹⁸Okrem vlastnosti veľký a malý

¹⁹Kapitola 11

[#7=1.00 [#0=1.00 Noun*]*Adverb[#0=1.00 Noun*]]
[#0=1.00 Noun*]

Vidíme, že samoorganizácia gramatiky prebehla úspešne. Vďaka správnej reorganizácii formy sa uchovala vhodná rekurzívna podoba. Referent označený znakom * je tiež na správnom mieste.

Pozrime sa teraz na tak povediac patologický prípad simulácie, ktorý nám ujasní predošlé výsledky o povahe kategórií v takomto jazyku.



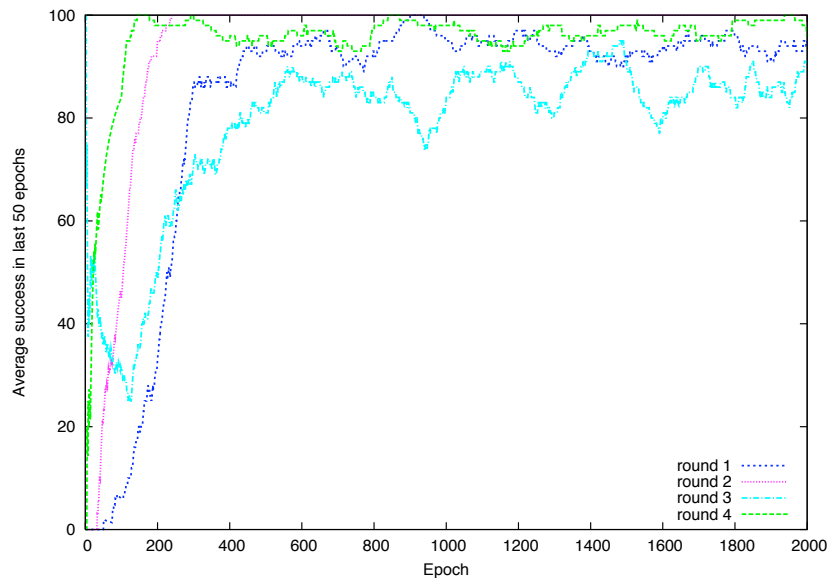
Obrázok 3: Patologický prípad simulácie bez učenia, ale s používaním kategórií

Priebeh výuky agenta bol pre simuláciu na obrázku 3 nasledovný. V prvom kole sa učila ako obvykle základná kategória. Druhé kolo všeobecných kategórií sa neučilo vôbec. V treťom kole dostal agent pomenovania adjektívom a menom všeobecnej kategórie, ktorú nebol učný. V štvrtom kole malo pomenovanie adjektívom tvar všeobecná kategória+príslovka+základná kategória. Ako vidíme, toto kolo dopadlo najhoršie, agent nedokázal odlíšiť objekt od učiteľom nastraženého podobného objektu. Všetky kolá okrem tretieho sú očakávateľné, možno štvrté má prekvapivo nízke hodnoty. Pre vysvetlenie sa pozrime na úspešnú netriviálnu gramatiku v tomto kole.

[#1=1.00 [#0=1.00 Noun*]*Adjective]

Ako vidíme, poradie slov je vymenené voči poradiu akému by mal učiť učiteľ. Pomenovanie kategóriou slúži v takomto jazyku ako prídavné meno, ktoré si agent vie špecificky pamätať pre substantívum. Toto pomenované substantívum je však v skutočnosti prídavné meno - modifikátor, atribút. Vysvetľujeme tento jav ako efekt samoorganizácie, ktorá preferovala významy prídavných mien a preto ich postavila v gramatike na miesto podstatného mena, ktoré má rozhodujúcu váhu v algoritmoch výberu referenta.

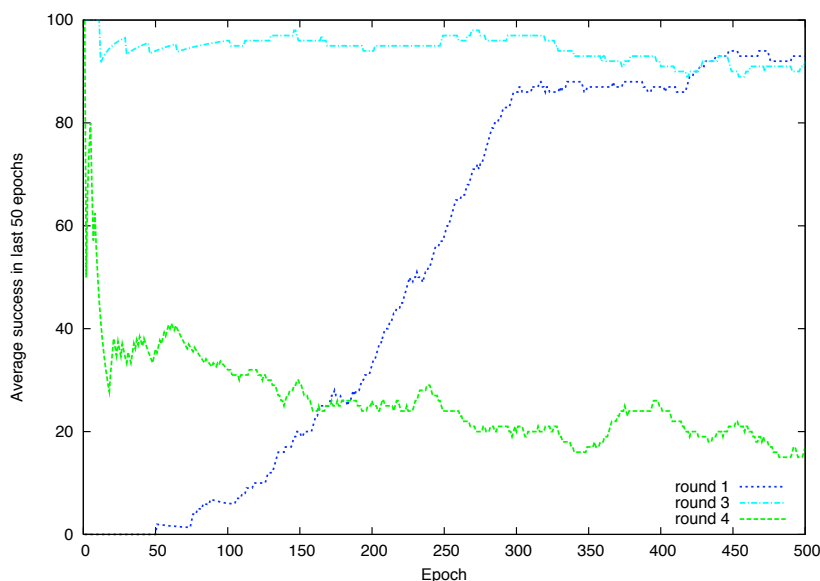
Čo sa týka štvrtého kola, v ňom agent nenadobudol žiadnu použiteľnú gramatiku pre pomenovania použitím prísloviak, pretože kombináciou dvoch prídavných mien príslovkou nejde uskutočniť, preto bola jeho úspešnosť nízka. Príkladom je veta "Veľký zelený vpravo", ktorá sa zdá v poriadku, ale používa zamlčaný podmet. Agenty nemajú implementované algoritmy interpretácie významu vety bez podstatného mena, i keď ich architektúra im to výslovne nezakazuje.



Obrázok 4: Prvá generácia s identifikovateľnými kategóriami

Skúsme teraz použiť nasledujúci trik, ktorý by nám mal umožniť naučenie všeobecnej kategórie. Uvidíme tak, čo vnesie takéto umelé naučenie sa ka-

tegóriám. Tie budú v jazyku fungovať ako synonymá, agent ich jednoznačne rozlíši aj keď ich podoba je mnohorodá. Súbor **animalsID.csv** obsahuje atribúty zvierat upravené o číselný atribút zaradenia zvieratá do kategórie. Rozlišovanie hodnôt v tomto atribúte stačí na odhalenie, ktoré zviera patrí do ktorej kategórie. Inak povedané, atribút, keďže je v rámci inšancií kategórií invariantný, bude rozhodujúcim atribútom požadovaným pre rozlišovacie kritérium. Ako môžeme vidieť na Obrázku 4, kategórie sa naučili v druhom kole najlepšie a najrýchlejšie, pretože ich určenie je jednoznačné. V treťom kole vidíme, že oproti predošlému grafu mierne klesla úspešnosť tretieho kola. Stalo sa tak preto, lebo na scéne zrazu mohli existovať zvieratá, ktoré boli síce jednej kategórie, mali však rôzne vlastnosti na rozdiel od prípadu, keď pomenovanie základnou kategóriou spôsobilo jednoznačnejšie identifikovanie objektu, ktorý mal všetky vlastnosti dané. Výsledok je však stále veľmi uspokojivý.

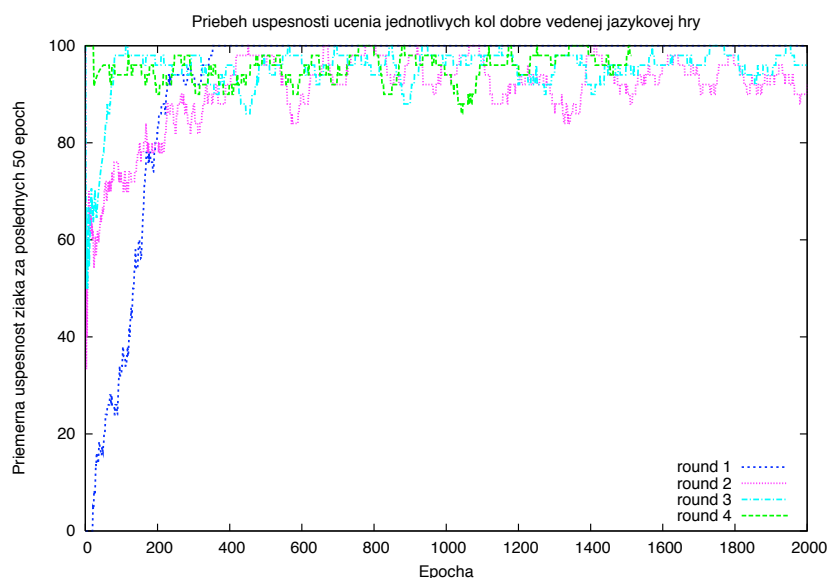


Obrázok 5: Identifikovateľné kategórie, nedostatočne utvorená scéna

Pozrime sa ešte na výsledky na Obrázku 5, ktoré znázorňujú úspešnosť učenia sa bez kategórií ale pravidlami riadený agent, negeneroval pri tvorbe pomenovania pre objekt na scéne podobný (bias) objekt. Agent nebol

schopný uspieť v štvrtom kole. Potvrdil sa teda pôvodný predpoklad, že správne utvorená scéna je pre akvizíciu konceptov v našom modeli nutná podmienka úspešnej samoorganizácie.

7.2 Druhá generácia



Obrázok 6: Druhá generácia

Obrázok 6 z druhej generácie naznačuje, že prenos medzi agentmi s rovnakými lingvistickými zručnosťami je bezproblémovjší. V tejto simulácii bol použitý opäť štandardný súbor bez jednoznačnej identifikácie kategórie. Druhé kolo, ktoré má byť o kategóriách, je však v tomto prípade len pokračovaním prvého kola v pomenovaní základnými slovami pre zvieratá. Deje sa tak preto, lebo agent sa v prvej generácii naučil kategórie veľmi zle a nemá ich preto dôvod v jazyku používať. Ak by sme mu to aj chceli programovo nanútiť, nemáme aparát na rozlišovanie medzi jednotlivými druhmi kategórií, takže pre agenty sú to všetko jednoducho podstatné mená. Významy, ktoré sú ním učené, patria medzi tie, ktoré si najlepšie osvojil a preto ich preferuje v popisovaní objektov.

Zaujímavú zmenu úspešnosti oproti prvej generácií vidíme v treťom kole, pomenovanie použitím prídavných mien. Pôvodne v prvej generácií priemerná úspešnosť oscilovala niekde nad 80%, no v druhej generácií nastalo 10% zlepšenie. Tento jav si vysvetľujeme tak, že v procese učenia ďalších generácií sa na učenie ďalších generácií generovali prídavné mená, ktoré čo najviac obmedzili význam slov pre referenta²⁰ a preto sa minimalizovala šanca na nepochopenie. Pravidlami riadený agent generoval prídavné mená náhodne, až na už spomínané podmienky kladené na rozdielnosť s podobným objektom.

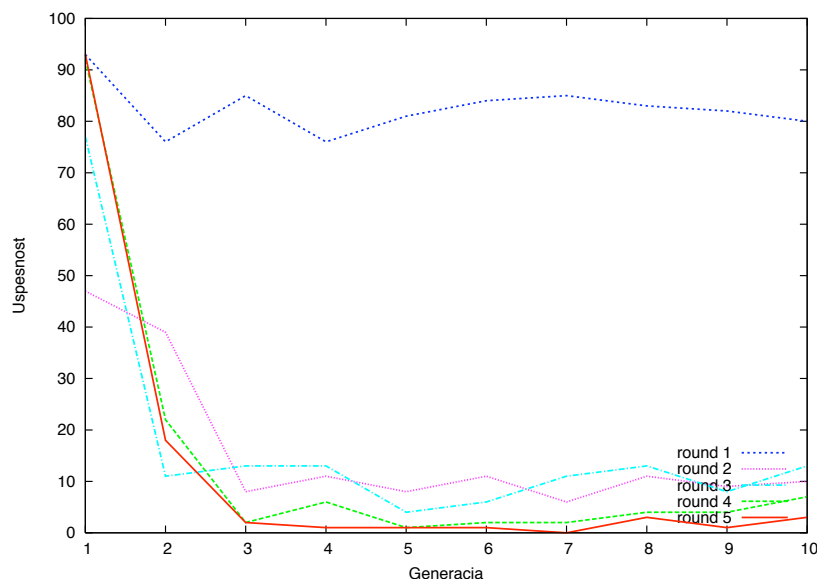
7.3 Stabilita ďalších generácií

Skúmame teraz stabilitu takéhoto jazyka v medzigeneračnom prenose iterovaného učenia. Bude nás zaujímať priemerná úspešnosť každého kola jazykovej hry po tom, čo je agent natrénovaný v dobre vedenej hre. Pravidlami riadený agent bude skúšať žiaka každej generácie 100 náhodnými vetami pre príslušné kolo a počítat' priemernú úspešnosť určenia správneho referenta.

Najskôr sme urobili simuláciu s učením a používaním kategórií v slovných frázach. Výsledok môžeme vidieť na obrázku 7. Okrem prvého kola pomenovaní využívajúcej základnú kategóriu, je prenos pôvodného jazyka neúspešný. Ako sa ukázalo aj v predošlej simulácii, abstraktné kategórie zvierat nie sú schopné prenosu pre neschopnosť agenta správne s nimi pracovať a preto sa ďalšie generácie nemajú ani možnosť takéto výrazy naučiť.

Obmenili sme preto predošlý experiment. Preskočili sme učenie kategórií v druhom kole a taktiež používanie kategórií vo vyšších kolách pri označovaní. Obrázok 8 už ukazuje úspešný prenos jazyka na ďalšie generácie, pričom jazyk je stabilný v tom zmysle, že agent rozumie všetkým svojím predkom v priemere rovnakou mierou.

²⁰Volaním funkcie `getMostConstrainingAdjective`

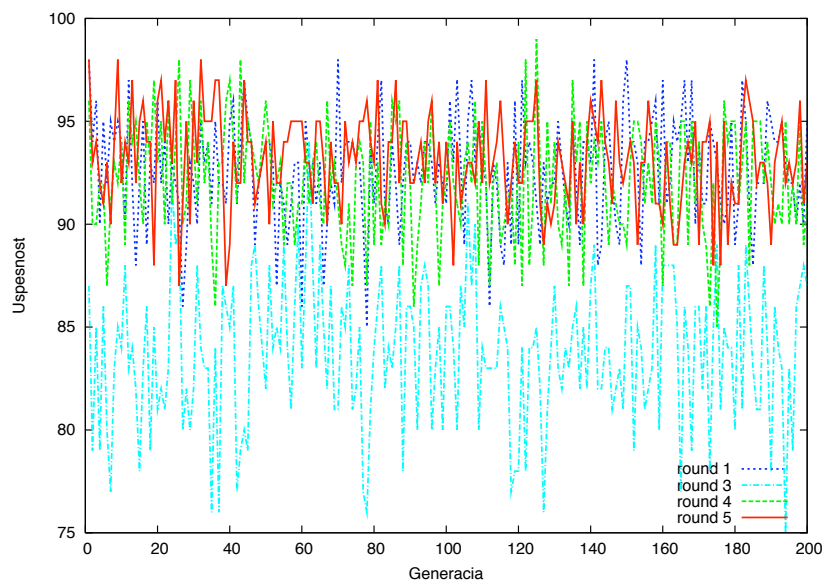


Obrázok 7: Stabilita prenosu významu za použitia všeobecných kategórií

8 Problematické miesta a budúci vývoj systému

Navrhnutý model jazyka stavia gramatické štruktúry na kombinácii slovných druhov. Je teda schopný pochopiť primeranú množinu viet, pre ktorej pochopenie stačí pracovať so slovami takto. V plne rozvinutom jazyku však potrebujeme gramatikou zachytiť zložitejšiu štruktúru, ktorá určuje komplikovanejšie vzťahy medzi slovami. Označenie "Miska na stole" využíva príslovku na označenie miesta objektu, ale označenie "Miska na stôl", má iný význam. Z určitého hľadiska pomenúva už akési miesto a rovnako aj misku, ale vplyvom prítomnosti gramatickej značky pádu v slove "stôl" rozumieme touto vetou niečo iné.

Z tohto pohľadu je voľba slovného druhu ako statického vplyvu na kombináciu slov nedostačujúca. Zdá sa však, že ak by sme vedeli rozlišovať aj ďalšie identifikátory okrem slovného druhu, gramatika by bola bohatšia o schopnosť zachytiť nové významy. Tak ako vzniklo zaradenie do slovného druhu samoorganizáciou, aj v prípade iného spôsobu kategorizácie slov, by



Obrázok 8: Stabilita prenosu významu s použitím základných kategórií

mohol tento proces fungovať. Bolo by však potrebné urobiť niekoľko zmien v architektúre agenta. Prostredie a pravidlá hry by mali vytvárať podklad na vytvorenie porozumenia komunikovaným konceptom.

Pozrime sa však všeobecne na typy alebo kategórie významov. Predpokladajme, že máme možnosť zoskupovať objekty do kategórií a určovať mieru podobnosti medzi nimi. Vertikálna hierarchia kategórií by mohla poskytnúť rôznu stupeň abstrakcie, ak by tieto kategórie slov boli prvkami formy konštruktu, gramatiky. Avšak správne naučenie sa takejto gramatiky by bolo oveľa náročnejšie a použitý algoritmus učenia sa by musel byť odolný voči chybám. Výuka agenta by opäť musela prebiehať postupne od jednoduchších po zložitejšie druhy výpovedí pre vhodne a nie príliš zložité scény. Pre príslušnosť do kategórie by mohla hovoriť aj prípona alebo predpona, čo by však vyžadovalo agentovu citlivosť na takéto informácie. Problémy s udržiavaním kategórie objektov, a vôbec výber charakteristík kategórie sú nemalé. Udržiavané kategórie by induktívne vytvárali hierarchické usporiadanie, ale každým novým príkladom by sa mohla vyvolať reťaz zmien usporiadania pre príbuzné kategórie. Aj takto by mohlo vyzeráť rozšírenie modelu, i

hrubý náčrt je komplikovaný a preto sa nebudeme kategóriami viac zaoberať s necháme stranou túto možno všeobecnej alternatívy k slovným druhom.

Prednosťou grafovej reprezentácie aj je to, že si ľahko zaznačíme vzťahy medzi slovami, prípadne frekvenciu ich používania. Ľudia pri učení sa jazyka intuitívne získavajú schopnosť pamätať si vzťahy pre bežne používané frázy, sú schopní asociovať, nájsť slovo pre kontrastný význam a tak ďalej.

Po získaní lepšieho aparátu na porovnávanie významov by si graf v hranách a vrcholoch mohol všetky informácie zapamätávať a samoorganizáciou by bolo potrebné zabúdať na nepodstatné vzťahy a posilovať úspešné. Kategória prísloviak určenia miesta by napríklad pri učení sa jazyka so skloňovaním mohla obsahovať informáciu o páde zviazaného slova, pričom to zviazané slovo je opäť z určitej špecifikovanej kategórie. Aj pre tvorbu kategórii je nutné dbať na to, aby sa kategorizovalo na základe toho, či je potenciálna inštancia nejakej kategórie používaná rovnakým spôsobom (spája sa s rovnakými slovami) ako všetky ostatné inštancie kategórie. Sémantika jazyka môže byť obohatená o nové heuristiky bootstrappingu postavené na teórii kategórií.

Navrhnutý model je zdá sa schopný ďalšieho vývoja pre akvizíciu zložitejších konceptov. Je potrebné pridať slová pre ďalšie slovné druhy zo slovesného ostrova a vhodne modifikovať jazykovú hru.

9 Zhrnutie a diskusia

Výsledky prenosu významu kompozične vytvorenej vety za použitia základnej kategórie dopadli dobre. Samoorganizácia takto navrhnutých gramatických foriem prebehla úspešne. Základná kategória pomenovaní podstatnými menami mala najlepšie výsledky a pre podradenú a všeobecnú kategóriu dopadli výsledky v súlade s očakávaním teórie.

Prenos jazyka pri pomenovaní objektov pomocou všeobecných kategórií prebiehal v ďalších generáciách úspešne, ale jazyk sa nezachoval v pôvodnej podobe. Toto správanie však potvrdzuje, že jazyk je dynamický systém, v ktorom súťažia medzi sebou významy v medzigeneračnom prenose. Agentov zúžený profil významov²¹ je tvorený snahou o čo najpresnejšie pomenovanie. Neúspešné lingvistické prvky nie sú prenášané do ďalšej generácie.

Potvrdenie našej domnienky o nutnej existencii podobného (bias) objektu pre simuláciu jazykového prenosu s použitím nami navrhnutého pravidlami riadeného učiteľa nás neprekvapili. Spôsob jazykovej výuky agentov tak, že generujeme scénu zároveň so slovným popisom sa ukázal byť plodný najmä preto, že uľahčil samotný proces implementácie agenta učiteľa. Ako sa ale potvrdilo, je nutné dbať na prítomnosť konceptov, ktoré majú podobný význam a zabránili zlej interpretácii a následného učenia sa. Navrhovaný model akvizície je možné chápať aj ako mechanizmus prenosu jazyka medzi dvoma odlišnými druhmi reprezentácie sémantiky a syntaxe cez riadenú výmenu informácií.

Agent bol úspešný v akvizícii správnych gramatických foriem. Výsledky z patologickej simulácie bez dobrého naučenia základných konceptov ukazujú, že algoritmus môže mať problémy s lokálnymi minimami tvorenými neoptimálnou gramatikou, ktorá sa nevie zlepšiť. Slovné druhy sa určili síce neoptimálne, ale dostačujúco nato aby v jazyku plnili čiastočne správnu úlohu uchovania významu. Podstatné mená ako aj prídavné mená sú zachytené rozlišovacími kritériami objektov.

²¹Meaning bottleneck

10 Záver

V diplomovej práci sme postupne načrtli teoretické pozadie akvizície jazyka u malých detí. Zámerom bolo zoskupiť dostatočné množstvo informácií na navrhnutie vhodného modelu akvizície kompozičného jazyka, ktorý je schopný prenosu medzi generáciami.

Paralelne sme tiež aplikovali teóriu sémantiky rozlišovacích kritérií na niektoré dôležité špeciálne prípady výrazov v jazyku a hľadali vysvetlenia, ako tieto javy zachytiť v báze sémantických znalostí. Navrhovaná sémantická grafová štruktúra splnila očakávania a osvedčila sa v reprezentácii významu v jazyku s gramatikou ako kombináciou slovných druhov. Hrany grafu poskytli vhodné miesto, pre uloženie informácie o špecifickom význame určitej dvojice slov.

Akvizícia ako aj medzigeneračný prenos gramatiky a významov prebehol úspešne ako sme na výsledkoch v grafoch ukázali a tiež bližšie dokumentovali okomentovaním samotného priebehu slovnej hry. Ukázali sme aj rozdiel v rôznej metodike priebehu učenia jazyka a vplyve na tvorbu významov ako aj gramatiky.

V jednoduchej forme sme skúmali aj tvorbu kategórií, s výsledkami, ktoré súhlasia s teóriou o všeobecnej a základnej kategórii slov tak, ako sme vzhľadom na mechanizmus fungovania rozlišovacích kritérií aj očakávali.

V texte sme načrtli niektoré fundamentálne vylepšenia práce s gramatikou, ktorá by potom nadobudla schopnosť zachytiť mnoho nových významov. Namiesto rozlišovania pevných slovných druhov ako prvkov skladby vety by sa použili všeobecné kategórie. Tento spôsob si však okrem schopnosti kategorizovať vyžaduje aj zabezpečenie odolnosti algoritmu indukívnej tvorby rozlišovacích kritérií voči zlým príkladom. V našej práci sme preto boli nútení spresniť pravidlá slovnej hry tak, aby tento neželaný jav nenastal no v prípade takejto gramatiky by to zrejme ani nebolo možné.

Výsledky stability jazyka sa ukázali byť iné ako v popisovanom experimente nekompozičného jazyka v [17]. Domnievame sa, že tento rozdiel je spôsobený presnou špecifikáciou slovnej hry a premysleným generovaním

scény učiteľom. Môže to znamenať aj to, že kompozičný jazyk je pre prenos generáciami stabilnejší, pretože redukuje množstvo významov správnym zakódovaním pomocou gramatiky a uľahčuje pochopenie aj učenie sa.

Záverom môžeme konštatovať úspešné splnenie stanoveného cieľa, obohatenie už existujúceho modelu o spôsob uchovania kompozičných významov.

11 O elektronickej prílohe

Priložený kompaktný disk obsahuje adresár **/akviziciaGramatiky**, ktorý je potrebné pred spustením experimentu prekopírovať na lokálny, zapisovateľný disk. Sú pod ním uložené nasledujúce adresáre:

Kópia tejto práce je vo formáte PDF v adresári **doc**.

Dodatočná HTML dokumentácia vygenerovaná pomocou Javadoc pre zdrojový kód je k dispozícii v adresári **doc/javadoc**.

Samotné zdrojové kódy implementácie sú v **src**. Balík pripravený na spustenie je uložený ako java archív spustiteľný príkazom **java -jar akviziciaGramatiky.jar**. Statická metóda `main()` spúšťajúca experiment je v triede `Jarmila.akviziciaGramatiky`. Pre úspešné spustenie je potrebné mať verziu Java 1.5 alebo vyššiu. Adresár **lib** obsahuje programom využívanú knižnicu `JGraphT`.

Priebeh experimentu sa loguje do rôznych súborov v adresári **logs**. Názvy sú zakódované nasledovne: **gX_rY_Z.txt**, kde **X** je číslo generácie, **Y** je číslo kola slovnej hry a **Z** sú reťazce identifikujúce typ logovanej informácie.

Súbory obsahujúce názov **dict** obsahujú slovník úspešného zaradenia slov do slovných druhov aj s ich úspešnosťou. Obsahujú aj gramatický slovník konštruktov.

Spolu uchovávajú záznam jazykovej hry medzi pravidlami riadeným agentom a poslednou generáciou v procese iterovaného učenia.

Súbory obsahujúce názov **ILM_stability** uchovávajú záznam jazykovej hry medzi pravidlami riadeným agentom a poslednou generáciou v procese iterovaného učenia.

Adresár **logs/graphs** obsahuje bash skript **mkgraphs.sh** na vygenerovanie grafov pre všetky súbory ***.p** použitím nástroja `gnuplot`.

Do súboru **g1_rules.txt** vypíše pravidlami riadený agent svoju bázu znalostí. Tú získal zo vstupných dát a číta ju zo súborov formátu CSV **animals.csv** alebo **animalsID.csv**.

Popíšme si formát bázy znalostí pravidlami riadeného agenta, ktorý číta zo súbore `animals.csv`. V prvom riadku sa nachádza hlavička, z ktorej sa

využívajú názvy vlastností zvierat. Je rozdelená znakom ; ktorý rozhoduje o tom, ktoré prídavné mená, t.j. vlastnosti zvierat, sa budú používať pri tvorbe viet. Atribúty pred znakom sa budú používať, atribúty za znakom slúžia iba ako doplnková informácia, ktorá algoritmu pomáha tvoriť vhodné rozlišovacie kritériá. Mená pre doplnkové atribúty agent v hre nevyslovuje.

Prvý stĺpec umožňuje zakomentovať riadok znakom #.

Druhý stĺpec je kategória zvierat, ktorého meno je v treťom stĺpci. V ďalších stĺpcoch sú zadané hodnoty atribútov tak, že pred znakom @ je hodnota, ktorá pripočítaná k náhodnej hodnote z intervalu medzi 0 a reálnym číslom za týmto znakom tvorí numerické vyjadrenie hodnoty atribútu tohto stĺpca. Hodnota -1 znamená, že tento atribút sa nebude do rámcov generovať.

Obsah v experimentoch použitej bázy znalostí sa týka nasledujúcich zvierat zaradených v udanej kategórii.

(amphibian, frog), (bird, eagle), (bird, parrot), (bird, ostrich), (bird, chicken), (bird, penguin), (bird, sparrow), (insect, ladybird), (insect, butterfly), (insect, ant), (insect, spider), (mammal, human), (mammal, cow), (mammal, bear), (mammal, mouse), (mammal, monkey), (mammal, dog), (reptile, snake), (reptile, turtle), (fish, tuna), (fish, carp)

Atribúty zvierat sú beak, teeth, legs, tail, fin, fledged, size.

Atribúty, ktoré učiteľ generuje aj pri pomenovaní ako prídavné mená sú ravenous, domestic, skinny, flying, furry, aquatic.

V prvom riadku sa nachádza hlavička, z ktorej sa využívajú názvy vlastností zvierat. Je rozdelená znakom ; ktorý rozhoduje o tom, ktoré prídavné mená, t.j. vlastnosti zvierat, sa budú používať pri tvorbe viet. Atribúty pred znakom sa budú používať, atribúty za znakom slúžia iba ako doplnková informácia, ktorá algoritmu pomáha tvoriť vhodné rozlišovacie kritériá. Mená pre doplnkové atribúty agent v hre nevyslovuje.

Prvý stĺpec umožňuje zakomentovať riadok znakom #.

Druhý stĺpec je kategória zvierat, ktorého meno je v treťom stĺpci. V ďalších stĺpcoch sú zadané hodnoty atribútov tak, že pred znakom @ je hod-

nota, ktorá pripočítaná k náhodnej hodnote z intervalu medzi 0 a reálnym číslom za týmto znakom tvorí numerické vyjadrenie hodnoty atribútu tohto stĺpca. Hodnota **-1** znamená, že tento atribút sa nebude do rámcov generovať.

Obsah v experimentoch použitej bázy znalostí sa týka nasledujúcich zvierat zaradených v udanej kategórii.

(amphibian, frog), (bird, eagle), (bird, parrot), (bird, ostrich), (bird, chicken), (bird, penguin), (bird, sparrow), (insect, ladybird), (insect, butterfly), (insect, ant), (insect, spider), (mammal, human), (mammal, cow), (mammal, bear), (mammal, mouse), (mammal, monkey), (mammal, dog), (reptile, snake), (reptile, turtle), (fish, tuna), (fish, carp)

Atribúty zvierat sú beak, teeth, legs, tail, fin, fledged, size.

Atribúty, ktoré učiteľ generuje aj pri pomenúvaní ako prídavné mená sú ravenous, domestic, skinny, flying, furry, aquatic.

Prídavné mená veľkosti, "big / veľký" a "small / malý" rovnako aj príslovkové určenie miesta: "over / nad", "under / pod", "left_of / vľavo_od", "right_of / vpravo_od" sú generované priamo zdrojovým kódom.

Referencie

- [1] J. L. Austin. *Ako niečo robiť slovami*. Kalligram, Bratislava, 2004.
- [2] Chang N. Benjamin, K. B. Embodied construction grammar in simulation-based language understanding, 2003.
- [3] Maia T. Chang, N. *Grounded Learning of Grammatical Constructions*. AAAI Spring Symposium on Learning Grounded Representations., 2001.
- [4] P. Cohen. Dynamic maps as representations of verbs. In *European Conference on Artificial Intelligence*, pages 145–149, 1998.
- [5] Stevan Harnad. *The Symbol Grounding Problem*, volume 42, pages 335–346. *Physica D* 42: 335-346., 1990.
- [6] Hurford J. Kirby, S. The emergence of linguistic structure: an overview of the iterated learning model. *Computational Approaches to the Evolution of Language and communication.*, pages 121–148, 2001.
- [7] G. Lakoff. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind* Chicago. University of Chicago Press, 2006.
- [8] Smith L. B. Jones S. S. Landau, B. The importance of shape in early lexical learning. *Cognitive Development*, 3, 299–321., 1988.
- [9] J. Macnamara. *Names for things: a study of human learning*. MIT Press, Cambridge, MA, 1982.
- [10] E. Markman. *Categorization and Naming in Children: Problems of Induction*. MIT Press, Cambridge, 1989.
- [11] Hutchinson J. E. Markman, E. M. Children’s sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology*, 16, 1–27., 1984.
- [12] Steven Pinker. *The language instinct. How the mind creates language*. William Morrow, New York, 1994.

- [13] Mervis C. B. Rosch, E. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605., 1975.
- [14] Mervis C. B. Gray W. D. Johnson D. M. Boyes-Braem P. Rosch, E. *Basic objects in natural categories*. *Cognitive Psychology* 8: 382-439, 1976.
- [15] J. Šefránek. *Kognícia bez mentálnych procesov*. Kalligram, Bratislava, 2002.
- [16] K. Simon. *Learning, Bottlenecks and the Evolution of Recursive Syntax*. *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press., 1998.
- [17] M. Takáč. *Konštrukcia významov a jej dynamika v procese iteratívneho učenia*. Kelemen, J., Kvasnička, V., Pospíchal, J. (eds.): *Kognice a umělý život VII*. Slezská univerzita, Opava, pp. 341-347., 2007.
- [18] M. Tomasello. *First verbs*. Emory University, Atlanta, 1992.
- [19] Hemenway K. Tversky, B. Objects, parts, and categories. *Experimental Psychology: General*, 113, 169-193., 1984.
- [20] B. L. Whorf. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT Press, Cambridge, 1956.
- [21] L. Wittgenstein. *The Blue and Brown Books*. Oxford: Blackwell, 1958.