

LINGUAGEM GRAMATICAL: ANÁLISES E DISCUSSÕES

Vera Aparecida de Figueiredo

Universidade Estadual de Campinas

Faculdade de Engenharia Elétrica e de Computação - FEEC

vera.aparecida@gmail.com

Resumo - Este trabalho tem por objetivo demonstrar alguns experimentos voltados à área de “*evolução de linguagem*” em sistemas inteligentes. Esses experimentos possuem distintos níveis de compreensão e utilização de linguagem gramatical utilizando verbos, substantivos, classes de palavras, cenas visuais, em diversas acepções para o desenvolvimento de robôs que utilizam linguagem. Serão demonstrados e avaliados os experimentos, especificando o uso da linguagem gramatical em cada um deles.

Palavras - chaves: *linguagem gramatical, “evolução de linguagem”, sistemas inteligentes.*

I – INTRODUÇÃO

Qualquer sistema de símbolos que serve como meio de comunicação de idéias (sonoros, gráficos, gestuais) e podendo ser percebido por diferentes órgãos dos sentidos, é o que leva a distinguir várias espécies de linguagem: visual, auditiva, tátil. Os elementos constitutivos da linguagem são: gestos, sinais, sons, símbolos ou palavras usadas para representar conceitos de comunicação, idéias, significados e pensamentos.

A função biológica da linguagem é aquilo que mais profundamente distingue o homem de outros animais. Alguns pesquisadores consideram que o desenvolvimento dessa função cerebral ocorre em estreita ligação com a bipedia e a libertação das mãos, que permitiram o aumento do cérebro, a par do desenvolvimento de órgãos fonadores e da mímica facial. E para esses pesquisadores a principal distinção entre o homem e o animal, se dá através da linguagem.

Segundo Gudwin (2009), a linguagem no ser humano evolui da seguinte forma: até 2 meses a criança reconhece a linguagem materna, de 6 a 8 meses as primeiras

vocalizações, de 8 a 12 meses vocalizações sofisticadas, 1 ano as primeiras vocalizações com significado, 18 meses estágio de 2 palavras (nome-adjetivo, sujeito-verbo, verbo-objeto), 3 anos linguagem gramatical simples e palavras modificadas, plurais e verbos conjugados e aos 6 anos: 14.000 palavras (9 palavras por dia).

Stevan Harnad (2008), em seu artigo (*Why and how the problem of evolution of Universal Grammar is hard*), ele cita que a gramática universal (UG), não é um conjunto de regras que aprendemos na escola, e que a percepção de conhecer essas regras “implicitamente”, nos tornam capazes de falar com todos gramaticalmente.

As regras da UG podem ser aprendidas através de tentativa e erro, com correção de erros, não é que as crianças falam na perfeição a partir do nascimento, mas é a partir da observação, imitação e correção de erros que a criança faz, que passa do estágio de ser incapaz de falar e em ser capaz.

Os experimentos que serão demonstrados são voltados a esse estudo da formação de linguagem gramatical, que compreende a fase dos 3 anos de idade da criança. Serão descritos alguns experimentos de pesquisadores da área de “*evolução de linguagem*”. Dentre eles será citado experimentos de Deb Roy, Michael Fleischman, Peter Gorniak, Niloy Mukherjee e Josh Juster, em questões que envolvem aprendizagem da linguagem.

II - DESCRIÇÃO DOS EXPERIMENTOS

No primeiro experimento que será relatado por Deb Roy e Michael Fleischman presente no artigo, (*Why verbs are harder to learn than nouns*), eles descrevem sobre a realização de um modelo computacional que tem por objetivo o aprendizado situado de palavras. Para os autores, o crescimento em direção a ciência cognitiva é como

modelar palavras que são fundamentadas em representação sensório-motor, provendo explicações sensório-motor, como mapas de palavras para desenvolvimento físico do mundo.

Baseado nessa visão eles desenvolveram um modelo que é adquirido a partir de um léxico da linguagem humana de participantes interagindo em um ambiente virtual do jogo, onde cada personagem usa discurso espontâneo para guiar as ações do outro. Os autores destacam o reconhecimento intencional, em conhecimento de palavras para formalização de estruturas conceituais de formação intencional e o modelo aprende mapeamento das palavras, observando eventos e objetos nestas inferências de estruturas intencionais.

O jogador tem que responder oralmente para pedidos do outro jogador, através da execução de várias tarefas, a fim de ganhar o jogo.

Um dos participantes diz a palavra: “GROK”, com base nessa informação, é apresentado as seguintes interpretações para a palavra GROK: “Open the door”, “Move to the door”, “Another play into the room”, “Player to go find some objects”.

Os autores citam que o primeiro tipo de ambigüidade que aparece, se refere a ambigüidades verticais, ambigüidade entre “find axe” versus “open the door”, para a introdução da palavra “GROK”.

Nas ambigüidades horizontais, ele descreve a diferença de interpretação entre “Find axe ” e “Let in player”.

Tanto uma como a outra são usadas para representação de ação e serve como base para reconhecimento intencional, que é a capacidade de inferir as razões subjacentes em um comportamento do agente com base em uma seqüência de ações observadas.

Além do reconhecimento intencional os autores usam a “gramática do comportamento”, que é constituída por regras intencionais que descrevem ações intencionais dos agentes (exemplo, “find axe”), que pode conduzir a ações de menor nível intencional, onde cada nó codifica uma forma semântica que contém os participantes da ação e seus papéis temáticos (agent: player, patient : axe e action: find).

O experimento consiste, seqüência paralela de fala e ação, que são registradas a partir de um jogador orientando o

outro. A cada ação observada, o algoritmo analisa as estruturas de inferência que descreve todas as possíveis ordens de ação intencional, que poderiam ter produzido por ele. Como muitas ações são observadas, o algoritmo se concentra em uma única seqüência mais provável de ação intencional. Para encontrar a mais provável árvore de intenção, o algoritmo resolve a ambigüidade horizontal, que descreve uma seqüência de ações observadas, mas ainda subsiste o nível de descrição que o orador tinha em mente para a sua emissão, ou seja, a ambigüidade vertical.

Para resolver essa ambigüidade vertical, ele usa o “mapeamento lingüístico”, onde cada nó infere estruturas de intenção, que consiste em formar uma semântica. O algoritmo de “mapeamento lingüístico” tenta aprender as associações entre palavras em expressões e funções. Foi usado um algoritmo de estimativa EM;

EM tem um conjunto de entradas paralelas e encontra uma ótima localidade condicional de probabilidades distribuídas para a interação entre um passo estimativa (E) e um passo maximização (M).

Para compreender o uso do EM é assumido que já se sabe que o nó vertical está associado com uma expressão (ou seja, nenhuma ambigüidade vertical).

No passo E, uma condição probabilística é usada para coletar uma contagem de quantas vezes uma palavra aparece numa expressão, com função de unir pares semânticos.

No passo M, as contagens esperadas para calcular uma nova probabilidade de distribuição condicional.

Foi acrescentado um papel NULL, de carga semântica a cada palavra que atua como “lixo colecionador”, palavras comuns que não são facilmente mapeadas de objeto ou ação (“a”, “agora”, “OK”, etc).

Algoritmo EM:

- 1 – Seta probabilidades uniformes para todas as expressões/ formas emparelhadas.
- 2 – Para cada par executa padrão EM.
- 3 – Une as distribuições de saídas do EM (ponderação de cada um pela probabilidade emparelhamento).
- 4 – Usa as distribuições unidas para recalculas as probabilidades de todas as expressões/formas

emparelhadas.

5 – Volta ao passo 2.

Por iteração através de todos os possíveis mapeamentos, o algoritmo eventualmente converge para uma probabilidade de distribuição que maximiza a probabilidade dos dados. Para o experimento, foi usada uma média de 130 expressões de formação de dados.

Foi selecionado o máximo de expressões no sistema como probabilidade de aprendizado para os testes das expressões e foi examinado quantas vezes o sistema mapeia cada palavra daquela expressão para correta função semântica. Palavras mapeadas com precisão foram separadas por classe de palavras (substantivos e verbos). Foram treinados totais de 719 verbos, e foi obtida uma exatidão de 44.0 %, e para 575 substantivos treinados houve uma exatidão de 65.2%.

Em um outro artigo, (*A trainable spoken language understanding system for visual object selection*), Deb Roy, Peter Gorniak, Niloy Mukherjee e Josh Juster, descrevem sobre um treinamento no entendimento de linguagem falada em sistemas de seleção de objetos virtuais.

O sistema adquire uma gramática e vocabulário a partir de “apresentar” e “dizer”, em que cenas visuais é emparelhado com descrições verbais.

O sistema fica sobre um tampo de uma mesa, onde um conjunto de objetos é colocado na frente dele. Usando um apontador a laser, o sistema aponta o objeto de maneira randômica, através de um mestre falando as descrições dos objetos. Uma vez treinado a pessoa pode interagir com o sistema por descrever verbalmente objetos colocados na frente dele. O sistema reconhece e analisa o discurso e aponta em tempo real, o objeto que melhor se ajusta à semântica visual da descrição falada. O sistema processa expressões tais como, “A maçã verde à esquerda do copo” e localiza o objeto em cena visual adequada.

O reconhecimento do discurso é realizado em tempo real usando um vocabulário médio (até 1000 palavras).

O sistema de aquisição de linguagem é distribuído em um grande número de processos separados, cada um responsável

por aspectos de aprendizagem. O sistema inclui as “w-procs” são processos de palavras, que é atribuído a cada palavra única e estima contexto visual semântico para esta palavra, as “c-procs” são processos de agrupamentos, medidas de similaridades entre as “w-procs” e encapsula a saída similar de “w-procs”, para torná-las um processo único, as “g-procs” são processos gramaticais, que combinam o resultado de vários outros processos que abrangem partes adjacentes de uma expressão, e por último as “a-procs”, processos de ação, ouve todos os outros processos ativos ocorrendo em relação a uma determinada expressão e saídas corretas. O artigo relata que a idéia central é ganhar robustez na análise utilizando esse conjunto de processos distribuídos. As “w-procs” são responsáveis por modelar palavras visualmente fundamentadas. Para fazer isso cada “w-proc”, estima uma gaussiana multivariada sobre todas as características oferecidas pelo sistema visual, condicionado sobre a ocorrência da palavra atribuída.

Quando uma palavra associada com a “w-proc” ocorre, o processo dá as características do objeto que Newt esta apontando e publica a gaussiana atual a outros processos em execução. Sempre que este modo é ativado, a “w-proc” também mede o quão importante sua palavra é fundamentada computando a distância *Kullback-Leibler* entre a palavra condicionada gaussiana e um outro modelo gaussiano que é compilado a partir de todos os objetos que foram encontrados.

Junto com cada “w-proc”, um processo agrupamento (“c-proc”), é liberado e anexado ao “w-proc”. Periodicamente esta “c-proc” analisa todos os outros “c-procs” para fundamentar a sua própria similaridade. Esta similaridade é calculada primeiramente pela formação do perfil de semântica da palavra.

$$s(w) = \begin{bmatrix} KL(w) - KL_{\bar{f}_1(w)} \\ KL(w) - KL_{\bar{f}_2(w)} \\ \vdots \\ KL(w) - KL_{\bar{f}_n(w)} \end{bmatrix}$$

Onde $KL_{\bar{f}_i(w)}$ é a distância *kullback-leibler*, calculada após a remoção da característica *i* da gaussiana.

Esse perfil mede o quanto cada característica contribui para o fundamento da palavra. A distância fundamentada é calculada com:

$$d(s(w_i), s(w_j)) = (s(w_i) / |s(w_i)|)^T (s(w_j) / |s(w_j)|)$$

que mede o grau de coincidência entre características de duas palavras. Se a similaridade for superior a um determinado limiar, os dois “c-proc” são combinados dentro de um. Para um “c-proc” com múltiplas palavras fundamentadas associadas, todas as combinações de similaridades entre as palavras são calculadas e é retirada uma média para dar nova similaridade consensual.

De fato, “c-procs” funilam as ativações, para múltiplas “w-procs” similares, para torná-las idênticas na origem, para um nível de processo mais alto.

Sempre que dois processos abrigarem expressões próximas uma das outras, o processo libera expressões gramaticais “g-procs”. Cada “g-proc” está ligada a dois outros processos de saída, e realiza funções específicas para combiná-los. Um “g-proc” pode executar agrupamentos logarítmicos sobre gaussianas que recebe.

Para Newt, um processo como esse, contribui para conceitos associados “cone amarelo”, composto de conceitos “amarelo” e “cone”.

Um processo ativo, “a-proc”, ouve a saída de todos os outros processos e tenta contar fragmentos na expressão. Ele calcula a probabilidade para cada objeto na cena com base na saída dos processos e envia uma recompensa para os processos que atribuem a maior probabilidade de ser o objeto realmente indicado e que cobre a maior parte das palavras na expressão. Junto com a recompensa, a “a-proc”, envia de volta o objeto indicado durante o treinamento.

Cada processo que recebe uma recompensa por seu retorno, envia sua recompensa aos processos que são usados para cobrir uma parte das expressões correntes.

Na fase de entendimento da linguagem, se nenhum objeto é indicado corretamente, a “w-proc” não toma nenhum exemplo e não envia nenhuma mensagem de recompensa. De preferência, a “a-proc” envia o mais provável objeto indicado pelo processo que abrange mais palavras na frase para o controlador Newt, que por sua vez aponta para o objeto.

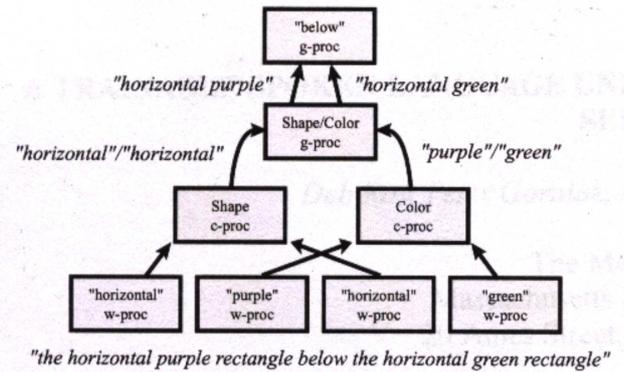


Figura 1- Exemplo simples no processo de entendimento de linguagem

A figura acima mostra um exemplo de análise de expressão por processos, deixando de fora a “a-proc”, que reúne todos os resultados dos processos.

Foi usada uma base de dados a partir de 303 expressões, cada expressão descreve um objeto na cena de quatro objetos. Quando treinado o Newt alcança 82% na precisão escolhendo o objeto correto, em comparação com base aleatória de 25 %.

Em um outro artigo, (*Speech, Space and Purpose: Situated language understanding in Computer Games*), Deb Roy descreve a realização de um experimento por meio de jogos de computador, nos quais a situação é mediada pela “linguagem”. Para este autor, os jogos vêm ganhando cada vez maior importância na vida das pessoas.

Baseado nessa visão de mundo ele desenvolve um experimento (*Bishop*), onde usa pares de participantes, não limitado a “linguagem”, para conseguir uma tarefa, onde um dos participantes fala e outro escuta. Ambos vêem 30 objetos, porém eles não se vêem; existem dois monitores que mostram cenas idênticas. O participante que fala seleciona o objeto e descreve-o ao ouvinte. Este, por sua vez, identifica o objeto e seleciona-o em sua tela. Se caso ambos, falante e ouvinte, escolhessem o mesmo objeto, esse desaparecia e o falante poderia continuar escolhendo outro objeto; caso contrário tinha nova tentativa para descrevê-lo. Nesta tarefa houve um êxito de 65%. A partir de arranjos aleatórios de objetos em duas cores, falantes usavam configurações para

distinguir os objetos, analisando os dados, estratégias descritivas e relações lingüísticas.

Tais estratégias distintivas realizavam-se através de categorias como cor, regiões espaciais, agrupamentos, frases, analogias (ponto referência).

O objetivo do Bishop é modelar estratégias descritivas, utilizando uma semântica básica para entender a linguagem em termos semânticos relacionados ao mundo. Ele distingue entre posições relativas e absolutas. Exemplo: “verde na esquerda”, para significar o objeto verde mais a esquerda independentemente da posição horizontal absoluta. O meio é tomado a partir do centro da cena. O Bishop seleciona o alvo correto em 80% das vezes na base de dados.

No artigo, (*Learning Visually Grounded Words and Syntax of Natural Spoken Language*), Deb Roy apresenta um modelo CELL (*Cross-channel early lexical learning*), um modelo de aprendizado de palavras sensório-fundamentadas, que aprende segmento de discurso, categorias visuais e adquire semântica adequada entre palavras faladas e categorias visuais. Foram feitas gravações, com discurso de 6 mães, brincando com seus filhos, usando brinquedos comuns.

Estas gravações foram emparelhadas, com os mesmos objetos de vídeos gravados por um robô, proporcionando entradas multisensoriais para o modelo. O modelo usa uma rede neural recorrente (RNN), que extrai características fonêmicas do discurso de entrada. O modelo também gera histograma, de compreensão fonêmica, onde colunas com múltiplas faixas brilhantes indicam confusões fonêmicas.

Características visuais são usadas para analisar a forma da figura. A distância d , o ângulo θ , formados para cada par pontos ao longo da imagem são medidos. Todos os pares são analisados e agregados em um histograma bidimensional.

A fala e as imagens são analisadas juntas para encontrar segmento de discurso similar ao contexto visual.

A memória de curto prazo (STM) armazena as últimas cinco expressões faladas pareadas com a co-ocorrência de formas.

CELL sistematicamente compara todos os pares de segmentos falados, sobre todos os pares de expressões na

STM. Quando um segmento recorrente do discurso é encontrado, as formas dos histogramas são comparadas. Se os contextos visuais são semelhantes uma recorrência crossmodal, é detectada.

A recorrência segmento falado e a forma são analisadas como adivinhação de uma palavra e seu significado, a “candidata léxica”.

Ao longo do tempo as candidatas léxicas encontradas são acumuladas na memória de longo prazo (LTM). Um exemplo de candidatas léxicas é o segmento falado, retornando o formato de uma bola.

Candidatas léxicas não são confiáveis por duas razões. Como o processo sensorial é propenso a erros, muitas recorrências detectadas levarão a candidata léxica incorreta. Em segundo lugar, a análise de recorrência leva a candidatas semânticas inapropriadas.

Por exemplo, se a mãe na gravação repetir “yeah”, enquanto está brincando com brinquedo cachorro, há a hipótese de ocorrer semântica incorreta para “yeah” pareado com cachorro.

Uma informação mútua (MI), é utilizada para medir a associação entre o discurso e as categorias visuais resultantes. O MI, para um número de possibilidades de limites visuais e auditivos é computado, originando uma superfície MI. Uma superfície (MI), para o “yeah” emparelhado com cachorro é baixa, mas “shoe” emparelhado com sapato produz emparelhamento de alto pico. Candidatas léxicas que levam a valores de picos altos como “shoe”, constituem resultado final do modelo.

CELL aprende pequeno vocabulário de nomes como bola, sapato e associa alguns termos semânticos para cães como “ruf-ruf”, para automóveis “vroom” e conduz a 50% de palavras com precisão.

No mesmo artigo, (*Learning Visually Grounded Words and Syntax of Natural Spoken Language*), Deb Roy, apresenta DESCRIBER (*Learning Word and Grammar Semantics in a Spatial Description Task*), é o aprendizado palavras e semântica gramatical numa tarefa de descrição espacial. A imagem gerada pelo computador é um conjunto

de dez retângulos sobrepostos. O “sistema perceptivo” constitui de um conjunto de funcionalidades extratoras que opera sobre a imagem.

Aprender no DESCRIBER consiste em seis fases:

Na primeira fase, formação de classes de palavras, 2 métodos de agrupamento de palavras foram avaliados, primeiramente palavras que co-ocorrem em uma descrição é improvável que pertençam a mesma classe, segundo método clusteriza palavras que co-ocorrem em contexto similar visual. Um método híbrido que combina os dois métodos leva a um agrupamento ótimo de palavras. Na fase 2, subconjunto de características visuais são selecionadas e associadas a cada palavra. Na fase 3, para cada palavra, um modelo de distribuição gaussiana é calculado usando as observações que ocorrem com aquela palavra. Na fase 4, uma classe-base, modelo estático de linguagem bigram é aprendido e modela a sintaxe do substantivo nas frases. Na fase 5, usa modelo de linguagem bigram, para identificar substantivos no corpo da frase. Expressões que contem 2 substantivos são usados como entrada para esta etapa e estágio 6. Na fase 6, formação de frases com múltiplos substantivos são usados como uma base para estimar um modelo de frase baseado no modelo de linguagem bigram. Modelo linguagem substantivos adquiridos no estágio 5 são incorporados modelo linguagem nessa fase.

Treinar Describer, um participante humano foi convidado a descrever verbalmente 500 imagens, cada descrição falada foi manualmente transcrita, resultando em um corpo de formação de imagens emparelhado com expressões transcritas. O modelo tem estruturas de 3 camadas, o nível alto nível de abstração é modelado pelo modelo Markov que especifica possíveis seqüências de substantivos nas frases e conectores de palavras, a maioria termos espaciais. Dois nodos na frase gramatical designa substantivos nas frases [TargetObject e LandMarck Object] e são diagramaticalmente ligadas por linhas tracejadas para o próximo nível no modelo a fim de distinção alvo e marca divisória na frase. O sistema “sabe” que o primeiro substantivo na frase descreve a meta e o segundo o alvo.

Cada palavra no modelo de linguagem substantivo frase

está ligado a um modelo visual associado. As palavras “escuro”, “claro” e “branco”, foram agrupadas em classes de palavras na fase 1. Os componentes de cor “azul” e “verde”, foram formadas na fase 2. Palavras condicionadas no verde, azul, yellow, na fase 3. O modelo para 'escuro' especifica baixa valores de ambos os componentes azuis e verdes, enquanto que "light" e "brancos" especificam valores elevados.

Um sistema de síntese do discurso é usado para converter a seqüência de texto à fala. O resultado final da saída do sistema são descrições faladas do objeto alvo na voz de um professor humano. O sistema planejador funciona da seguinte forma: fase 1, criação de substantivos na frase, usando um modelo de substantivo na frase como um gerador estocástico, a mais provável seqüência de palavras, é gerada para descrever o objeto alvo, na fase 2 é calculada uma ambigüidade para o alvo do substantivo na frase. O algoritmo Viterbi (Rabiner, 1989) é utilizado para calcular a probabilidade de que cada objeto na cena corresponde à frase meta, na fase 3, geração de cláusulas espaciais. Um objeto marco objeto é selecionado automaticamente que pode ser usado para identificar inequivocamente o alvo.

Testes com o DESCRIBER, ele apresenta um êxito de 81.3% na seleção de objetos.

III - DISCUSSÃO

No primeiro experimento proposto por Deb Roy, o autor descreve sobre o reconhecimento com base no aprendizado de palavras situadas, o autor não deixa claro quando ele descreve sobre ambigüidades verticais e horizontais, em relação ao aprendizado de verbos e substantivos, ele cita verbos como (encontrar – find, abrir – open e mover -move), onde ocorre uma vez ao longo do caminho vertical, da raiz a folha, mas por exemplo, o substantivo porta – “door”, ele ocorre várias vezes ao longo do caminho, isso deixa uma questão em aberto, ou seja, talvez isso poderia influenciar para melhor aprendizagem de substantivos do que de verbos. Neste nível, ele poderia aprofundar mais sobre o estudo, ou apresentar alguma explicação.

Em segundo experimento, proposto por Deb Roy, deixa

mais claro como foi feito o processo para aquisição de linguagem, onde o autor cita sobre os processos separados, cada um responsável pela análise de linguagem, “w-procs”, processos de palavras, “c-procs”, processos de ação, “g-procs”, processos gramaticais, “a-proc”, processos de ação, vantagem de se estruturar o sistema sobre esses processos, tanto que o sistema alcança 82% de precisão, escolhendo o objeto correto. O autor não cita sobre a compreensão de verbos “acontecer”, “dar”, interrogação “porque” e negação “não”.

No terceiro experimento, também proposto pelo mesmo autor, ele cita sobre experimento por meio de jogos de computador, nos quais a situação é mediada pela “linguagem”, baseada na linguagem infantil, mas não deixa claro como se baseou esse desenvolvimento. Do ponto de vista lingüístico, seu embasamento teórico é bastante vago para explicar o experimento, ele não faz uma descrição precisa tanto conceitual quanto técnica em relação aos agentes envolvidos no entendimento da linguagem. É neste nível que o autor poderia aprofundar uma discussão a respeito do processo envolvido na aquisição de linguagem para o sistema.

No quarto experimento, o autor cita sobre o experimento CELL, o autor apresenta uma clareza da compreensão de linguagem falada, pois as entradas sensoriais dão origem as representações lingüísticas. Podemos perceber que ele usa a representação icônica - “*o signo por si só possui tudo necessário para representar o objeto*” - que vem da semiótica, permitindo relacionar a linguagem natural para o objeto visual e isso também está relacionado com a “aprendizagem”, onde as crianças relacionam o objeto à descrição do mesmo.

No quinto e último experimento, o autor cita sobre experimento DESCRIBER, sobre a formação de “classes de palavras”, que gera descrições verbais para cenas, mas não deixa claro sobre o processo envolvido nessa construção de classes de palavras, por exemplo, o algoritmo envolvido.

IV – CONCLUSÃO

Podemos dizer que o estudo envolveu aprendizado

linguagem falada, verbos, substantivos, formação de classes de palavras, linguagem fundamentada a partir cenas visuais, esteve presente nos 5 trabalhos à medida que os autores descrevem sobre linguagem.

Entretanto, aproximamos onde o autor cita sobre aquisição do discurso por meio de “mostrar e dizer”, onde o sistema usa um apontador a laser, onde a aprendizagem ocorre em processos separados, cada um responsável pela formação gramatical da expressão e também do experimento CELL, onde podemos perceber o uso da representação icônica, vindo da semiótica. Pode ser avaliado, que com base no desenvolvimento de linguagem gramatical em criaturas artificiais teremos um ganho tecnológico e poderemos dar saltos qualitativos na área de “*evolução de linguagem*”.

V - REFERÊNCIAS BIBLIOGRÁFICAS

- Harnad, S. (2008) “Why and How the Problem of the Evolution of Universal Grammar (UG) is Hard”; Vol. 31 No 5 PP (524-525)
- Gorniak, P., Orkin, J., Roy, D. (2006) " Speech, Space and Purpose: Situated Language Understanding in Computer Games"; Twenty-eighth Annual Meeting of the Cognitive Science Society Workshop on Computer Games
- Roy, D. (2005) "Grounding words in perception and action: Computational insights"; TRENDS in Cognitive Sciences, Vol. 9 No 8
- Fleischman, M., Roy, D. (2005) “Why Verbs are Harder to Learn than Nouns: Initial Insights from Computational Model of Intention Recognition in Situated Word Learning”; Proceedings of the 27th Annual Meeting of the Cognitive Science Society
- Roy, D., Gorniak, P., Mukherjee, N., Juster, J. (2002) “A trainable Spoken Language Understanding System For Visual Object Selection”; In Proceedings of the International Conference of Spoken Language Processing
- Roy, D. (2001) “Learning Visually Grounded Words and Syntax of Natural Spoken Language”; *Evolution of Communication*, Vol. 4 No 1 PP(33-56)