**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE**

**CENTRO DE CIÊNCIAS HUMANAS LETRAS E ARTES**

**DEPARTAMENTO DE LETRAS**

**PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS DA LINGUAGEM**

**Elementos cognitivos dos sistemas naturais de linguagem no desenvolvimento de um modelador do discurso.**

**YAN DE LIMA JUSTINO**

NATAL – RN

2011

YAN DE LIMA JUSTINO

**Elementos cognitivos dos sistemas naturais de linguagem no desenvolvimento de um modelador do discurso.**

Projeto apresentado ao Programa de Pós-Graduação em Linguística Aplicada do Centro de Ciências Humanas, Letras e Artes da Universidade Federal do Rio Grande do Norte, como requisito para seleção do mestrado em Linguística Aplicada/2012.

Eixo de pesquisa: Discurso, Gramática e Ensino

Possível orientador: Prof. Dr. Marcos Antonio Costa

.

NATAL – RN

2010

**SUMÁRIO**

RESUMO 1

INTRODUÇÃO 2

OBJETIVOS 3

JUSTIFICATIVA

ESTADO DA ARTE

FUNDAMENTAÇÃO TEÓRICA

METODOLOGIA

CRONOGRAMA DE ATIVIDADES

REFERÊNCIAS BIBLIOGRÁFICAS

**RESUMO**

Este trabalho apóia-se nas abordagens dos estudos da lingüística cognitiva, tendo em vista resultados positivos e promissoras perspectivas que as pesquisas baseadas nos estudos da semântica da Simulação e compreensão da Linguagem oferecem, e tem como objetivo demonstrar, através de simulação computacional, como se dão processos cognitivos envolvidos na construção de sentido a partir do desenvolvimento de um sistema de modelagem do discurso em textos narrativos. Tal pesquisa fará uso de conceitos de *gramática e simulação, semântica da simulação e processamento do discurso* como elementos que possibilitarão uma compreensão plausível para o desenvolvimento de um modelo computacional que possibilite a construção um modelador do discurso. Concomitantemente, ao longo desta pesquisa, provocaremos uma aproximação entre as áreas que circundam o estudo da cognição nos sistemas naturais de linguagem.

**Palavras-chave**: cognição; linguagem; simulação; processamento do discurso

1. **INTRODUÇÃO**

No final do século XIX, a filosofia analítica, investigando as propriedades semânticas da linguagem natural, propõe que todo pensamento lógico poderia ser descrito em forma de símbolos. Esses símbolos seriam manipulados por meio de regras, sem a interferência da intuição dos sujeitos. Dessa forma, seria possível construir uma linguagem artificial livre dos problemas da linguagem natural. O advento das ciências computacionais segue, de certa forma, essa tendência analítica. O matemático Alan Turing demonstrou que qualquer função matemática, descrita por meio de uma formalização, poderia ser executada por uma máquina[[1]](#footnote-2). Desse modo, qualquer função poderia ser mecanicamente processada. Tal princípio apontou para a possibilidade da construção de uma máquina que simularia a lógica subjacente à linguagem natural.

Uma das abordagens do estudo das Máquinas de Turing ou de máquinas universais em geral é como reconhecedores de linguagens, ou seja, dispositivos capazes de determinar se uma dada palavra sobre o alfabeto de entrada pertence ou não a certa linguagem (DIVÉRIO, 2008:87).

Mesmo diante desta possibilidade, muitas das funções desempenhadas por sistemas de linguagem natural não puderam ser, até hoje, formalizadas e processadas mecanicamente. Isso porque, em relação aos sistemas naturais de linguagem, os sistemas físicos de manipulação simbólica não apresentam a mesma flexibilidade quanto à capacidade de adaptação/criação. De fato, os sistemas de linguagens formais, apoiados sob estruturas simbólicas, tecem em escopos sintáticos um conjunto de regras e procedimentos lógicos que tentam sugestionar algum tipo de representação semântica.

O problema na recorrente ideia de desenvolvimento de sistemas mais “inteligentes”, do ponto de vista interacional, é que tal empreendimento muitas vezes não pôde traspassar os conceitos de mente sobre os quais os computadores e, conseqüentemente, as linguagens formais foram desenvolvidos.

A própria percepção da indústria informática do que é inteligência é problemática. Em última instância, as pesquisas de inteligência artificial respondem a uma saliente influência da tradição behaviorista. Em informática, onde a relação entre inputs e outputs está no centro das atenções, o par estímulo-resposta, parece coerente (PRIMO, 2008:159).

A lingüística no século XX apresenta um percurso marcado por diversas teorias que compõem um contínuo de perspectivas cujos extremos são: uma visão formal de língua representada como sistema e outra centrada no caráter sócio-cultural de base cognitivista. Transitam entre estes extremos, teorias advindas da virada pragmática, da percepção da variação social da linguagem, e dos estudos da natureza discursiva e suas condições enunciativas. De modo geral, percebe-se que o projeto científico da lingüística no XX derivou da forma para uma visão sócio-interativa representada pela relação de continuidade entre sociedade e cognição.

Muito embora este quadro epistemológico apresente uma característica multifacetada e plural, esta diversidade não representa uma ação conclusiva. Muitas questões como a atividade referencial, problemas da significação, construção de categorias, problemas dos protótipos, metáforas e demais problemas da área ainda merecem um devido aprofundamento.

Em vista disso e que a lingüística vem se voltando cada vez mais para o estudo das atividades cognitivas sob a hipótese de que a compreensão envolve a ativação de simulações perceptivo-motoras em cenas descritivas e que determinados contextos discursivos impactam sobre o sistema de representação mental, a proposta de um projeto de pesquisa com base nos estudos da cognição, semântica da simulação, processamento do discurso e compreensão da linguagem, oferece uma possível análise das atividades de construção e um suporte para uma visão mais abrangente de como, nas práticas internacionais, emergem propriedades da cognição e processos que originam estruturas conceituais complexas.

Seguindo essa tendência e apoiados no conceito da teoria da linguística cognitiva de que a gramática é na verdade uma grande rede de construções, interessa-nos neste trabalho (i) demonstrar, a partir de simulação computacional, como se dão processos cognitivos dos sistemas naturais de linguagem, envolvidos na construção de uma gramática “on-line”; (ii) e como tal investigação pode, também, colaborar no desenvolvimento pressupostos para criação de modelos computacionais que se baseiem em elementos cognitivos para modelagem da informação a partir do processamento de linguagem natural.

**2. OBJETIVOS**

*Objetivo Geral*:

Demonstrar, através de simulação computacional, como se dão processos cognitivos envolvidos na construção de sentido a partir do desenvolvimento de um modelo computacional de modelagem do discurso em textos narrativos.

*Objetivos Específicos*:

Analisar processos de construção de sentido, utilizando-se de conceitos como *gramática e simulação, semântica da simulação e processamento do discurso*, advindos de experimentos de linhas pesquisa cognitivistas.

Propor modelo computacional de compreensão de linguagem natural a partir de uma perspectiva da simulação mental e processamento do discurso.

Tentar ampliar o campo de discussão de como a abordagem cognitivista pode colaborar no desenvolvimento de novas interfaces computacionais, possibilitando a reavaliação dos conceitos fundamentais da própria computação.

**3. JUSTIFICATIVA**

A hipótese de que a compreensão envolve a ativação de simulações perceptivo-motoras em cenas descritivas, motivam pesquisas que tentam revelar os processos envolvidos em sistemas representacionais. Uma das metas e desafios da abordagem cognitivista da linguagem é descrever tais processos pelo prisma do processamento do discurso. Ao fazer isso, busca-se uma melhor compreensão de quais representações são derivadas quando sentenças são reconhecidas. Experimentos de simulação mental e o monitoramento de ativação de áreas perceptivas e motoras do cérebro – como os realizados no **NTL2** (*Neural Theory of Language*), de Berkeley, Califórnia, e no **Language & Cognition Lab**, de San Diego, Califórnia – revelam o quanto determinados contextos discursivos impactam sobre o sistema de representação mental.

Tais experimentos evidenciam a simulação mental como instrumento promissor para compreender o processamento da linguagem e revelam que hoje há uma tendência nos estudos da lingüística cognitiva de trabalhos que fazem uso da simulação computacional enquanto procedimento de análise do processamento cognitivo de compreensão do discurso.

Tendo em vista os resultados positivos e as promissoras perspectivas que as pesquisas baseadas nos estudos da semântica da Simulação e compreensão da Linguagem oferecem, que propomos este projeto de pesquisa que tem como objetivo demonstrar, através de simulação computacional, como se dão processos cognitivos envolvidos na construção de sentido.

Vislumbramos que ao simular computacionalmente processos de significação, sob a ótica da abordagem cognitiva, podemos reforçar evidências das pesquisas de caráter cognitivista, bem como fornecer pressupostos para construção de modelos computacionais mais eficientes, no que se refere à experiência de interação homem-máquina, sob o aspecto de modelagem de discurso pelo processamento de linguagem natural.

Nessa perspectiva, acreditamos que as ciências cognitivas, especialmente a linguística cognitiva, podem fornecer uma imensa contribuição para o avanço nas pesquisas sobre a linguagem, bem como na formulação de conceitos e pressupostos para teoria computacional.

Para tal fim, o suporte teórico e os experimentos advindos dos estudos da semântica da Simulação e compreensão da Linguagem serão fundamentais para cumprimento das metas e objetivos deste projeto.

.

**4. FUNDAMENTAÇÃO TEÓRICA**

O gerativismo chomiskyano inaugurou uma tendência nos estudos linguísticos que contempla os fenômenos de natureza cognitiva, bem como os processos que estão envolvidos na interação mente e mundo, como fundamentais para compreensão da linguagem. Essa perspectiva foi responsável por um movimento revolucionário que estabelece a linguística como uma teoria mentalista, uma vez que se propõe a demonstrar que subjacente ao uma realidade performática existe uma realidade mental sistemática, autônoma e constituída de princípios inatos. COSTA e DUQUE (2009) afirmam que

De acordo com Chomsky, o aspecto relevante da linguagem  é o procedimento recursivo (gerativo), que se sustenta numa gramática universal inata. A Linguística seria, então, uma ciência da mente/cérebro, mais próxima das Neurociências, da Biologia e da Física, do que da gramática estrutural, tradicional ou da Sociologia.

É inegável que o empreendimento gerativista foi significativo para o que ficou conhecido como “virada cognitivista” e que instaura um novo patamar nos estudos da linguagem. No entanto, os gerativistas “privilegiaram em sua análise a busca de aspectos linguísticos universais, deixando de lado, questões sociais e interativas que caracterizam (...) o uso concreto da língua nas situações reais” (MARTELOTA, 2008:177). Além disso, a teoria gerativa apresentou uma destacada negligência no que se refere a um tratamento mais profundo para questões da significação e da idiomaticidade, o que resultou no surgimento de uma segmentação no campo teórico lingüístico que emergia do rompimento com o programa gerativo de Chomsky. Para SALOMÃO (2009), “gênese da ‘linguística cognitiva’ decorre, em parte, desse corte com Chomsky”. Segundo, ela

Os dois grandes cortes com o programa chomskyano – a questão da significação e a questão da idiomaticidade – levam ao desenvolvimento de um campo investigativo bastante heterogêneo nas suas formulações teóricas e preferências analíticas.

Muito embora a lingüística cognitiva apresente essa referida heterogeneidade, ela converge sob três aspectos fundamentalmente unificadores: (i) A cognição lingüística não representa um sistema autônomo aos demais sistemas cognitivos. Ela é contínua, e não modular; (ii) há uma relação contínua entres elementos sintáticos e lexicais determinadas pelo uso lingüístico; (iii) A semântica pelo viés da cognição possui um caráter inferencialista, uma vez que a significação deriva de um processo que se dá pela projeção entre domínios cognitivos.

Nessa perspectiva, a lingüística cognitiva apresenta algo inovador para os estudos da linguagem em relação aos modelos formais. Assumir que os mecanismos cognitivos de linguagem são contínuos aos demais sistemas de cognição possibilita repensar a relação das categorias lingüísticas sob uma ótica diferente do caráter simbólico e arbitrário do signo lingüístico. Segundo SALOMAO (2009),

A crescente expressividade do legado darwinista levou a um rearranjo nos estudos cognitivos de tal modo que pensadores de persuasão tão diversa quanto Bickerton, Pinker, Jackendoff e Fauconnier passaram a tematizar a questão da origem da linguagem, com óbvio impacto nas suas hipóteses sobre a natureza da gramática e das categorias lingüísticas.

Esse é um ponto crucial para o avanço na teoria linguistica, pois é admitindo a existência de tal contínuo entre linguagem e demais experiências de ordem cognitiva que se formulou a hipótese de que a linguagem parece envolver sistemas sensório-motor e as estruturas neurais dedicado a eles. Segundo BERGEN, B. e WHEELER, K. (2005), Experimentos demonstram que

imagens sobre a ação (Porro et al. 1996, Lotze etal. 1999) e percepção (Kosslyn et al. 2001) fazem uso de sistemas, de modo tal que a linguagem de processamento sobre o conteúdo perceptual ou motor resulta na ativação de estruturas neurais espelhadas sobre aqueles que seriam usados para realmente perceber ou executar o conteúdo descrito (BERGEN, B e WHEELER, K., 2005).

A idéia que reside sob este experimento é que o entendimento da linguagem implica na execução de uma simulação mental de uma cena descrita onde quem compreende ativa representações motoras correspondentes as que os participantes na cena podem fazer e representações perceptual de imagens que possam perceber. “Nesta visão, comunicação viável envolve a evocação bem sucedida de um conjunto alinhado de imagens motoras e perceptivas na mente de um ouvinte” (Glenberg & Robertson 1999, Bergen & Chang 2005).

Para uma grande parte dos pesquisadores da compreensão da linguagem há certo consenso de que compreender um texto implica na construção de um modelo de situação ou uma representação mental de uma situação descrita na linguagem. Destacam-se entre estes pesquisadores, Glenberg, Meyer, & Lindem, 1987; Graesser, Millis, & Zwaan, 1997; Johnson-Laird, 1983; Morrow, Bower, & Greenspan, 1990; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998.

Acredita-se que os modelos de situação partilham a mesma forma representacional que outros processos cognitivos não-lingüísticos na compreensão de textos (Kaup, B., Yaxley, R.H., Madden, C.J., Zwaan, R.A., & Lüdtke, J., 2007). Por tanto, essa noção nos leva a hipótese de que modelos de situação não são baseados em um sistema simbólico convencional e que compreender significa construir simulações mentais dos estados de coisas descritas no texto.

**5. METODOLOGIA**

Inicialmente, o trabalho será dedicado à pesquisa bibliográfica. Buscaremos textos teóricos que abranjam temas relevantes aos modelos de processamento de linguagem natural e modelagem do discurso*,*à luz da teoria cognitiva, identificando aspectos relativos à construção dos sentidos. Concomitantemente à pesquisa faremos elaboração de um modelo computacional protótipo de processamento textual de linguagem natural.

Num segundo momento, dedicáramos à coleta de dados para compor nosso corpus de textos. Devido ao seu caráter mais declarativo e para fins de simulação computacional, nos limitaremos aos textos do gênero relatório.

Em seguida, serão submetidos à simulação, pelo modelo protótipo, os dados coletados os quais deverão ser processados, analisados e confrontados com as perspectivas teóricas a fim de demonstrar processos de significação. Por fim, com os processos analisados e descritos passaremos à fase final, de redação, revisão e editoração da dissertação.

**CRONOGRAMA DE ATIVIDADES**

Pesquisa bibliográfica e estudos teóricos:

Leitura dos textos e coleta de dados para compor nosso corpus

Elaboração de protótipo de simulação de processamento de linguagem natural:

Submissão do corpus a simulação do protótipo e análise do processamento

Redação da dissertação:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1º Semestre | 2º  Semestre | | 3º Semestre | | 4º Semestre | |
| Pesquisa bibliográfica e estudos teóricos |  | |  | |  | |  |
| Leitura dos textos e coleta de dados |  | |  | |  | |  |
| Elaboração de protótipo de simulação |  | |  | |  | |  |
| Submissão do corpus a simulação |  | |  | |  | |  |
| Redação da dissertação |  | |  | |  | |  |

**REFERÊNCIAS**

BERGEN, B. and WHEELER, K. 2005. Sentence Understanding Engages Motor Processes. In **Proceedings of the Twenty.** Seventh Annual Conference of the Cognitive Science Society.

EHRSSON, H.H., GEYER, S., and NAITO, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part specific motor representations. **J. Neurophysiology**. 90: 3304-3316.

KOSSLYN, S. M., BALL, T. M. & REISER, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. **Journal of Experimental Psychology**: Human Perception and Performance, 4(1), 47-60.

KAUP, B., YAXLEY, R.H., MADDEN, C.J., ZWAAN, R.A., & LÜDTKE, J. (2007). Experiential simulations of negated text information. **Quarterly Journal of Experimental Psychology**, 60, 976-990.

TANENHAUS, M.K., & TRUESWELL, J.C., (1995). Sentence comprehension. In J. L. Miller and P.D. Eimas (Eds.) Handbook of Perception and Cognition, 2nd edition. Vol. 11: **Speech, Language, and Communication.** NY: Academic Press. Pp. 217-262.

.

1. A máquina de Turing, proposta por Alan Turing em 1936 ([TUR36]), é universalmente conhecida e aceita como formalização de algoritmo. Trata-se de um mecanismo simples que formaliza a idéia de uma pessoa que realiza cálculos (DIVÉRIO, 2008:83). [↑](#footnote-ref-2)