

# PROCESSAMENTO DA REFERÊNCIA EM ORAÇÕES RELATIVAS: CUSTO COMPUTACIONAL E COMPLEXIDADE SINTÁTICA

**Aluna: Caroline Kitzing**  
**Orientadora: Erica dos Santos Rodrigues**

## Introdução e objetivos

O presente trabalho está vinculado ao Projeto de pesquisa “Complexidade computacional em uma visão integrada processador/gramática: definição de uma métrica de avaliação de custo computacional na produção e na compreensão da linguagem” [1], cujo objetivo geral é estabelecer uma métrica de avaliação do custo de processamento de sentenças na produção e no *parsing* de enunciados lingüísticos, à luz de um modelo integrado da computação on-line [2, 3].

O projeto, que teve início em agosto de 2008, contou inicialmente com bolsa da Coordenação Central de Pós-Graduação e tinha como objetivo a investigação do custo computacional do processamento de orações relativas, com foco em potenciais demandas geradas por DPs complexos como expressões referenciais. A partir de outubro de 2009, a pesquisa passou a contar com bolsa de IC da FAPERJ (processo número E-26/102.174/2008) e teve seu escopo ampliado para o estudo da interferência de fatores sintáticos associados à estrutura interna de DPs complexos tanto no estabelecimento da referência quanto no processamento da concordância sujeito-verbo.

Os objetivos específicos da pesquisa são os seguintes:

(i) verificar se efeitos de distância linear e hierárquica reportados no estudo do processamento da concordância verbal na produção de sentenças [4], também são obtidos em experimentos de compreensão, e avaliar como tais efeitos seriam explicados à luz de um modelo integrado da computação *on-line* [2, 3]

(ii) verificar se existe relação entre os tipos de expressão referencial utilizados em sentenças complexas (em particular, orações relativas) e custo de integração de informações durante a compreensão de enunciados.

Na próxima seção, apresentaremos uma breve resenha de trabalhos que exploram questões associadas ao processamento da concordância e ao processamento da referência. Na seção seguinte, iremos reportar a metodologia adotada e os procedimentos de construção dos experimentos. Conclui-se com uma breve síntese do texto, apontando o atual estágio da pesquisa.

## 1 Breve resenha da literatura

### 1.1 Processamento da concordância

Um grande conjunto de trabalhos vem explorando, desde a década de 80, como se dá o processamento da concordância entre sujeito e verbo durante a produção de sentenças [4, 5]. Esses estudos têm apontado para a ocorrência de efeitos de atração provocados por DPs complexos como *A chave dos cofres desapareceram*. Observa-se que a presença de um elemento interveniente plural afeta o processamento e traz um custo adicional para a computação da concordância, levando à produção de erros de atração no processo de formulação de sentenças. Nesses experimentos, emprega-se uma tarefa de produção induzida de erros, em que o falante deve formular uma sentença completa iniciada por um DP

complexo que lhe é apresentado oralmente. A sentença pode ser construída livremente ou pode ser solicitado que o falante empregue um verbo que é projetado na tela de um computador. Manipulam-se nesses testes fatores sintáticos, semânticos e morfofonológicos de modo a verificar se podem gerar interferência no processamento da concordância. A variável dependente, isto é, a variável que permite verificar se houve efeito dos elementos manipulados, é o total de erros produzidos pelos sujeitos. Note-se que o erro nesse contexto é entendido como uma falha de processamento, determinada pela interferência da(s) variável(is) manipulada(s).

Em experimentos conduzidos com falantes de português [4], além do número dos elementos nominais do DP complexo (*A(s) tinta(s) do(s) cartucho(s) da(s) impressora(s)*) e da distância linear entre o núcleo do sujeito e o verbo a ser produzido (próxima – *A tinta dos cartuchos* + verbo vs. distante – *A tinta dos cartuchos da impressora* + verbo), também foram manipuladas outras variáveis, como tipo de modificador (sintagma preposicionado – *O técnico dos jogadores de futebol* vs. oração – *O técnico que treinou os jogadores*), status argumental do modificador (argumento – *A construção das casas* vs. adjunto – *O telhado das casas*), distributividade do sujeito (distributivo – *O volante dos carros* vs. não distributivo – *O cofre das jóias*) e fatores morfofonológicos como a forma de singular/plural do substantivo (forma variável – *O volante do(s) carro(s)* vs. forma invariável – *O volante do(s) ônibus*).

Os resultados obtidos indicaram efeito de *marcação* e de *distância linear* entre o núcleo do sujeito e o verbo, com mais erros para núcleo do sujeito não-marcado (singular) e linearmente distante do verbo, e efeito de *posição estrutural* do núcleo interveniente, com maior incidência de erros para os núcleos hierarquicamente próximos do nó mais alto do DP sujeito e núcleos inseridos em PPs argumentos. Um efeito semântico de *distributividade* associado a efeito de *marcação* também foi obtido. Quanto a fatores morfofonológicos, a informação de número no determinante (e não no nome) mostrou-se crucial para a identificação do número do DP sujeito.

Para explicar os erros, assumiu-se um modelo de produção que incorpora um *parser* monitorador, o qual atuaria em paralelo ao processo de formulação de sentenças [6]. Os erros são analisados como decorrentes de interferência de uma representação resultante do *parsing* do DP sujeito, que ocorreria após sua formulação [4]. A representação do DP sujeito, resultante do *parsing*, seria mantida numa memória temporária para que pudesse ser integrada ao restante da sentença que estaria sendo produzida em paralelo. Essa representação do DP complexo estaria sujeita a esvaecimento e poderia ocorrer perda (ou ‘enfraquecimento’) de informação relativa ao traço de número do núcleo do sujeito. Nesse contexto, o elemento nominal plural, contido no sintagma modificador, poderia ser tomado como o definidor do número do DP complexo. O fato de o plural ser um elemento marcado poderia torná-lo mais saliente e mais acessível, o que explicaria por que os erros de concordância ocorrem quando N1 (núcleo do sujeito) é singular e N2 (modificador) é plural. Partindo-se da hipótese que o falante poderia antecipar, em termos de expectativa, o número do verbo a partir dessa representação do DP mantida na memória, os erros poderiam ser resultantes de uma espécie de competição, em um componente morfofonológico de duas informações: a informação de número resultante da concordância na produção e a informação de número “antecipada” a partir de uma expectativa do número do verbo. Quando por questões de acessibilidade houvesse perda de informação relativa ao traço de número do núcleo do sujeito, a “previsão” feita com base na representação do DP sujeito gerada pelo *parser* seria incongruente em relação ao resultado da computação sintática na produção. Nesse caso, dois morfemas abstratos de número não idênticos estariam competindo no momento da codificação morfofonológica do verbo e poderia ocorrer uma associação do verbo ao morfema resultante da “expectativa” gerada no processo de *parsing*, fazendo com que parecesse ter ocorrido um erro de concordância.

Em nossa pesquisa, buscamos verificar se, no caso da compreensão de sentenças, os falantes percebem incongruências de número entre o sujeito e o verbo. Acreditamos que, assim como ocorre na produção, podem ocorrer efeitos de atração na computação da concordância no processo de compreensão de sentenças. Na seção 2, iremos apresentar o experimento construído com o objetivo de verificar essa hipótese.

## 1.2. Processamento de orações relativas e o estabelecimento da referência

A investigação do processamento de relações relativas tem sido conduzida no âmbito das pesquisas acerca dos recursos computacionais empregados no momento do *parsing*. Segundo Gibson [7], dois grandes processos estariam envolvidos no *parsing* de sentenças:

1. Integração de estruturas, que é a conexão de uma palavra a uma estrutura já montada.
2. Manutenção de estrutura na memória, que abrange ainda a manutenção de estruturas ainda incompletas.

Para explicar como esses processos são implementados, Gibson formula a *Dependency Locality Theory* (DLT). Uma das idéias principais é chamada de *localidade*, que pretende explicar como a distância física entre dois elementos que devem ser ligados pode interferir no custo dessa integração. Os custos de integração são calculados em termos do número de referentes discursivos novos que estão entre o núcleo do elemento a ser integrado e o núcleo do elemento ao qual ele está vinculado.

As dificuldades encontradas no processo de compreensão de sentenças podem ser investigadas a partir de estruturas encaixadas. Mesmo sem qualquer ambigüidade, estas estruturas são de difícil processamento:

- (1) O repórter não gostava do editor.
- (2) O repórter que o senador que João conheceu atacou não gostava do editor.

Em (1), a compreensão se dá de maneira fácil porque não há estruturas entre o verbo e o sintagma nominal que a ele se liga. Já em (2) a estrutura é bastante complexa por haver duas orações entre o sintagma nominal “o repórter” e o verbo “gostava”. A dificuldade de compreensão das estruturas encaixadas provavelmente é causada pela quantidade de recursos exigidos durante o processamento, já que não há ambigüidade e todas as palavras utilizadas são de uso corrente, não apresentando, portanto, dificuldades no nível lexical.

Entre as orações encaixadas, as relativas de objeto são as que apresentam maior custo de processamento, visto que envolvem um maior número de expressões referenciais intervenientes.

Pesquisas têm investigado a existência de uma relação entre complexidade sintática de orações relativas de objeto e o tipo de expressão referencial acessada no processamento dessas estruturas [8, 9, 10]. Verifica-se que, quando a expressão referencial de uma oração encaixada é um pronome de 1ª ou 2ª p. (*O técnico que eu critiquei expulsou a imprensa dos treinos*), é mais fácil processar o período do que quando a expressão referencial é um DP definido (*O técnico que o torcedor criticou expulsou a imprensa dos treinos*).

Em nosso estudo, explora-se a interferência de DPs do tipo D + N + PP, com foco no *status* argumental do PP modificador. Deseja-se avaliar se há diferenças no custo de processamento de orações relativas quando a expressão referencial é um DP que contém um PP adjunto (*A menina que o palhaço de macacão divertiu brincou com as irmãs*) ou um DP que contém um PP argumento (*A menina que o diretor de cinema elogiou ganhou um prêmio*).

## 2- Metodologia

A pesquisa que estamos conduzindo é de natureza experimental e envolve a aplicação de técnicas específicas. Foram planejados dois experimentos de compreensão, relacionados, respectivamente aos objetivos (i) e (ii) da pesquisa.

No primeiro experimento, busca-se verificar, especificamente, a interferência de fatores associados à distância linear e hierárquica na computação da concordância na compreensão. O experimento tem duas variáveis independentes: número do N2 (singular ou plural) e número do verbo (singular ou plural), às quais correspondem quatro condições experimentais, abaixo exemplificadas:

- a) O pediatra da criança da creche receitou o remédio.
- b) O pediatra das crianças da creche receitou o remédio
- c) O pediatra da criança da creche receitaram o remédio.
- d) O pediatra das crianças da creche receitaram o remédio.

A previsão é que, se efeito de atração também for observado na compreensão, os participantes da pesquisa tenderão a perceber como gramaticais sentenças do tipo (c) – em comparação com as do tipo (d), pois naquelas há um elemento interveniente plural que pode ser equivocadamente tomado como o elemento definidor do número do sujeito.

Serão utilizadas duas técnicas distintas para realizar esse experimento – a técnica de julgamento de gramaticalidade em tempo real e a técnica da leitura auto-monitorada.

No segundo experimento, pretende-se avaliar se propriedades estruturais do DP complexo, em especial a natureza argumental do PP modificador inserido no DP, podem afetar o processamento de orações relativas no estabelecimento da referência. Contrastam-se PPs adjuntos e PPs argumentos, com vistas a verificar se diferenças estruturais relacionadas ao status argumental do modificador podem afetar o processo de integração de estruturas. Partindo de resultados de trabalho sobre a interferência desse mesmo fator no processamento da concordância [4], em que PPs argumentos geraram maior interferência do que PPs adjuntos, prevê-se que sentenças do tipo (a) terão maior custo de processamento do que sentenças do tipo (b):

- (a) O escultor que o criador dos desenhos conheceu bateu com o carro. (PP argumento)
- (b) A empresa que o prefeito da cidade contratou blindou os carros da prefeitura (PP adjunto).

Pretende-se aplicar, inicialmente, a técnica de leitura auto-monitorada na realização desse experimento, mas está sob consideração o emprego de técnica em que o sujeito deve identificar um dado som, localizado após o estímulo teste (por exemplo, palavra começada com fonema b).

Os dois experimentos encontram-se em fase de implementação, o que tem envolvido encontros para estudo do programa computacional *PsyScope* [11], a ser utilizado em especial nos testes em que será empregada a técnica de leitura auto-monitorada. Nesse paradigma experimental, segmentos de sentença são apresentados na tela de um computador, a partir da ação do participante de pressionar algum item do teclado do computador. A cada pressionar de tecla, computa-se o tempo de leitura do segmento de sentença. A partir de uma relação entre tempo de execução da ação e demanda da tarefa, pode-se estabelecer que estruturas possuem maior custo de processamento. Nos meses de agosto e setembro, os experimentos serão aplicados.

## 3- Considerações finais

A pesquisa teórica permitiu fazer um levantamento de fatores relevantes para a identificação do custo de processamento de DPs complexos na computação da concordância

verbal e no estabelecimento da referência em sentenças relativas. Os encontros de discussão de metodologia e as sessões de treinamento em *PsyScope* possibilitaram verificar as amplas possibilidades de uso dessa ferramenta na programação de diferentes tipos de experimento psicolinguístico, em especial daqueles envolvendo a técnica de leitura auto-monitorada. Como fase final da pesquisa, está prevista a aplicação dos experimentos e a análise dos resultados.

### Referências

- 1 – RODRIGUES, E. dos S. Complexidade computacional em uma visão integrada processador/gramática: definição de uma métrica de avaliação de custo computacional na produção e na compreensão da linguagem. Projeto de Pesquisa, Depto. de Letras, PUC-Rio.
- 2- CORRÊA, L. M. S.; AUGUSTO, M. R. A. Computação lingüística no processamento on-line: soluções formais para a incorporação de uma derivação minimalista em modelos de processamento. **Cadernos de Estudos Lingüísticos**, v. 49, p. 167-183, 2007.
- 3- RODRIGUES, E. dos S.; CORRÊA, L. M. S.; AUGUSTO, M. R. A. Concordância sujeito-verbo em um modelo integrado misto (*top-down/bottom-up*) da computação *on-line*. **Veredas on-line**, v. 2, p. 76-91, 2008.
- 4- RODRIGUES, E. dos S. **Processamento da concordância de número entre sujeito e verbo na produção de sentenças**. Tese de Doutorado, Depto.de Letras, PUC-Rio, 2006.
- 5- VIGLIOCCO, G.; HARTSUIKER, R. J. The interplay of meaning, sound and syntax in sentence production. **Psychological Bulletin**, v. 128, n. 3, p. 442-472, 2002.
- 6- LEVELT, W. J. M. **Speaking: From Intention to Articulation**. Cambridge, Mass: MIT Press, 1989.
- 7- GIBSON, E. The dependency locality theory: A distance-based theory of linguistic complexity. In MARANTZ, A., MIYASHITA, Y. & O'NEIL, W. (Eds.), **Image, Language, Brain**. Cambridge, MA: MIT Press, 2000.
- 8- WARREN, T.; GIBSON, E. **The effects of discourse status on intuitive complexity: Implications for quantifying distance in a locality-based theory of linguistic complexity**. Poster apresentado no 12th annual CUNY conference on Human Sentence Processing. New York, N Y, 1999.
- 9- WARREN, T.; GIBSON, E. The influence of referential processing on sentence complexity. **Cognition**, v. 85, p. 79-112, 2002.
- 10- GORDON, P.C.; HENDRICK, R.; JOHNSON, M. Effects of noun phrase type on sentence complexity. **Journal of Memory and Language**, v. 51, p. 97-114, 2004.
- 11 - COHEN, J.D. et. al. PsyScope: A new graphic interactive environment for designing psychology experiments. **Behavioral Research Methods, Instruments, and Computers**, v. 25, n. 2, p. 257-271, 1993.