

Processamento co-referencial de nomes e pronomes em português brasileiro

Márcio Martins Leitão*

Resumo – Este trabalho tem como objetivo central investigar a eficiência, em termos de processamento co-referencial, de nomes e pronomes em retomadas anafóricas na posição de objeto direto em Português Brasileiro. Além disso, pretende testar e contrastar a hipótese da penalidade do nome repetido (GORDON; HENDRICK, 1998) e a hipótese da carga informacional (ALMOR, 1999), relacionadas respectivamente à maior eficiência de pronomes em relação a nomes repetidos e de Sintagmas Nominais (SNs) mais gerais (hiperônimos) em relação a SNs mais específicos (hipônimos) no que se refere ao processamento co-referencial.

Palavras-chave – Processamento discursivo. Co-referência. Objeto direto anafórico.

Introdução

O estabelecimento das relações entre elementos co-referenciais é um problema fundamental a ser resolvido pelo sistema de compreensão, seja no âmbito sentencial, seja no âmbito discursivo.

Um dos fenômenos mais comuns da linguagem humana é a referência a uma determinada entidade por meio de mecanismos anafóricos, como, por exemplo, os pronomes. As anáforas¹ são expressões lingüísticas cujo significado depende de uma parte anterior da sentença ou do discurso que se

* Doutorando em lingüística pela UFRJ, membro do LAPEX – Laboratório de Psicolingüística Experimental (Rio de Janeiro, Brasil). E-mail: mmleitao@vetor.com.br.

denomina antecedente. Ambos, anáfora e antecedente, são co-referenciais, já que se referem à mesma entidade do enunciado, embora com termos lingüísticos diferentes na maioria dos casos. Assim, no exemplo (a) tanto o pronome “eles”, como o antecedente “Julio e Edna”, se referem a duas pessoas, enquanto que no exemplo (b) o substantivo “felino”, e o antecedente “o leão”, se referem a um animal.

(a) **Julio e Edna** viajaram para Fortaleza.

Eles aproveitaram bastante a viagem.

(b) **O leão** gosta de caçar em bando.

Por isso esse **felino** vive em família.

A retomada anafórica é um mecanismo importante no estabelecimento da coesão discursiva, que facilita a integração de diferentes partes de uma sentença e/ou de um texto e evita a repetição de determinadas expressões já mencionadas. Dessa forma ajuda no processo de compreensão, reduzindo a carga da memória de trabalho do leitor/ouvinte durante a leitura/audição. Portanto é de grande relevância entender, do ponto de vista cognitivo, como o estabelecimento da co-referência ocorre, e que tipos de princípios e de fatores estão envolvidos no processamento desse fenômeno lingüístico.

Entre os vários fatores que atuam no processamento da co-referência investigaremos, neste artigo, a eficiência, em termos de processamento, de pronomes lexicais (ele e ela) e Sintagmas Nominiais (SNs), seja como nomes repetidos, seja como SNs em uma relação de hiponímia ou hiperonímia com seus antecedentes, na posição e função de objeto direto em português brasileiro (PB).

Estudos em outras línguas, utilizando a técnica experimental da leitura automonitorada (*self-paced reading*), têm encontrado menores tempos de leitura para a co-referência estabelecida por pronomes do que por nomes repetidos (CHAMBERS, SMYTH, 1998; GORDON, HENDRICK, 1998; GARROD,

FREUDENTHAL, BOYLE, 1994; GORDON, GROSZ, GILLIOM, 1993), além disso, estudos utilizando a técnica de rastreamento ocular (*eye-tracker*) também têm encontrado resultados na mesma direção (GARROD, FREUDENTHAL, BOYLE, 1994; KENNISON, GORDON, 1997)².

Os estudos de ALMOR (1999; 2000) referentes à língua inglesa, buscam também entender melhor o processamento da co-referência a partir de formas lingüísticas distintas e, para isso, focalizam retomadas anafóricas estabelecidas por SNs que ora mantêm uma relação de hiperonímia com seu antecedente, ora mantêm uma relação de hiponímia, como nos exemplos (c) e (d), respectivamente, a seguir³:

(c) **O felino** gosta de caçar em bando.

Por isso esse **animal** vive em família.

(d) Os **felinos** gostam de caçar em bando.

Por isso os **leões** vivem em família.

Os resultados encontrados por Almor indicam que retomadas anafóricas estabelecidas por SNs hipônimos (mais específicos) são processadas mais lentamente do que retomadas anafóricas por SNs superordenados (ou hiperônimos e mais gerais), o que mostra que dentro da categoria de constituinte SN não existe uma homogeneidade em termos de processamento da co-referência. Não parece haver, portanto, uma dicotomia entre pronomes e SNs como muitas vezes os estudos que comparam o processamento apenas entre pronomes e nomes repetidos fazem crer (por exemplo, GORDON, HENDRICK, 1997, 1998). O que há são diferentes graus de eficiência no estabelecimento da co-referência que varia com a forma lingüística da retomada anafórica.

Para explicar esses resultados ALMOR (1999) propõe a Hipótese da Carga Informacional (*Informational Load Hypothesis*), que é motivada psicologicamente e se baseia fundamentalmente na noção do custo operacional da

memória de trabalho e na funcionalidade da retomada anafórica processada, isto é, quanto maior a distância semântica entre o antecedente e a forma de retomada anafórica, mais carga informacional terá que ser processada na memória de trabalho para que se identifique a representação desse antecedente mantida na memória, e quanto mais carga informacional mais custoso o processamento co-referencial.

Por isso, termos mais específicos, como o hipônimo **leões** do exemplo (d), são processados mais custosamente do que termos superordenados mais gerais, como **animal** do exemplo (c), já que termos mais específicos têm mais traços semânticos a serem processados (maior carga informacional) no estabelecimento da co-referência do que termos mais gerais (menor carga informacional).

As investigações que serão explicitadas neste artigo caminham nessa direção, buscando, com base em dois experimentos de leitura automonitorada, observar e analisar se, em português brasileiro e em posição de objeto, diferentes formas lingüísticas são processadas também distintamente no estabelecimento da co-referência no âmbito discursivo. No primeiro experimento comparamos o processamento co-referencial entre nomes repetidos e pronomes, e no segundo experimento comparamos o processamento entre SNs superordenados e SNs hipônimos semanticamente relacionados aos seus respectivos antecedentes.

1. Experimento 1

Este experimento tem como objetivo comparar, a partir de uma tarefa de leitura automonitorada (*self-paced reading*), o processamento do pronome lexical com o de nome repetido em posição de objeto em PB. Ambas as formas lingüísticas retomam

anteriores nomes localizados em sentença anterior também na posição de objeto.

Nesse experimento está sendo testada a eficiência de pronomes e nomes repetidos, em termos de tempos de leitura, para verificarmos se pronomes são lidos mais rapidamente do que nomes repetidos no estabelecimento da co-referência, como prevê a hipótese proposta pela teoria da centralização, *Centering Theory* (GORDON; HENDRICK, 1998), com base no princípio de que pronomes são os veículos naturais para o estabelecimento da co-referência e por isso devem ser as formas lingüísticas preferidas.

Quando esse princípio não é seguido, ao se usar, por exemplo, um nome-repetido co-referencial, a teoria da centralização advoga que ocorre uma penalidade em termos de processamento, denominada penalidade do nome-repetido (*Repeated-Name Penalty*), ou seja, nomes repetidos co-referenciais são processados mais lentamente do que pronomes (GORDON, CHAN, 1995; GORDON, HENDRICK, 1998; CHAMBERS, SMYTH, 1998, entre outros).

Além disso, estamos testando a hipótese de que pronomes em posições não proeminentes, como a de objeto direto em PB, estabelecem de maneira natural e eficiente a co-referência com um antecedente também em posição não proeminente, convergindo com os achados de CHAMBERS, SMYTH (1998) para o inglês e com os achados de STREB, RÖSLER, HENNIGHAUSEN (1999) para o alemão, e divergindo dos princípios e conceitos expressos pela teoria da centralização, que propõe a existência da penalidade do nome repetido apenas para retomadas em posição de sujeito co-referentes a antecedentes também em posição mais proeminente (GORDON, HENDRICK, 1997, 1998).

A variável independente manipulada no experimento é o tipo de retomada anafórica: retomada utilizando pronome

lexical (PR) e retomada utilizando nome repetido (NR). A variável dependente é o tempo de leitura aferido após a apresentação das retomadas anafóricas (pronome ou nome repetido) relacionadas aos seus respectivos antecedentes em estruturas sentenciais coordenadas.

Para que não houvesse possíveis influências do fator extensão do constituinte, nós controlamos o tamanho dos nomes em termos de letras (por exemplo: Ivo, Ana), adequando-o ao tamanho dos pronomes lexicais (ele/ela). Com isso, permitimos que os resultados realmente reflitam o custo de processamento de pronomes ou nomes, focalizando a diferença de categoria gramatical e dos traços semântico-lexicais relacionados a cada uma dessas formas lingüísticas.

O Quadro I, na página seguinte, fornece um exemplo de cada uma das condições experimentais que foram testadas no experimento.

1.1 Método

Participantes:

Foram voluntários deste experimento 22 participantes, falantes nativos de português, com idade média de 20 anos, alunos de graduação da UFRJ, sendo 21 do sexo feminino e apenas um do sexo masculino.

Material:

O material consistiu de 2 conjuntos de 8 frases experimentais. Cada informante foi exposto a um desses conjuntos experimentais, embutidos em um conjunto extra de 20 frases distratoras. Cada conjunto experimental é composto de 2 condições com 4 frases por condição. Em uma condição, a retomada anafórica de um antecedente nome é feita por um

Quadro 14 – Exemplos das condições do experimento 1

(1) Retomada anafórica de antecedente nominal com pronomes lexical (PR)																		
Seg. 1	/	Seg. 2	/	Seg. 3	/	Seg. 4	/	Seg. 5	/	Seg. 6	/	Seg. 7	/	Seg. 8	/	Seg. 9	/	Seg. 10
Os vizinhos	/	entregaram	/	Ivo	/	na polícia	/	mas	/	depois	/	absolveram	/	ele	/	no	/	júri.
(2) Retomada anafórica de antecedente nominal com nome repetido (NR)																		
Seg. 1	/	Seg. 2	/	Seg. 3	/	Seg. 4	/	Seg. 5	/	Seg. 6	/	Seg. 7	/	Seg. 8	/	Seg. 9	/	Seg. 10
Os vizinhos	/	entregaram	/	Ivo	/	na polícia	/	mas	/	depois	/	absolveram	/	Ivo	/	no	/	júri.

pronome lexical (PR), enquanto que na outra condição a retomada anafórica de um antecedente nome é feita por um nome repetido (NR) (veja Quadro I). As frases experimentais são formadas por duas sentenças coordenadas e foram divididas em 10 segmentos, sendo que o segmento crítico é o segmento 8 em que se localiza a retomada com o pronome ou com o nome repetido. O tempo de leitura desse segmento é que vai ser levado em consideração na análise dos resultados, já que nele é que a co-referência se estabelece. Apesar das frases serem muito parecidas, diferindo apenas quanto à variável testada, a sua distribuição em dois conjuntos permitiu que todas as frases fossem comparadas e que cada participante fosse exposto a apenas uma frase de cada tipo.

O aparato experimental consistiu de um iMAC, G3, 233Mhz e uma caixa de botoeira conectada ao computador, específica para utilização com o programa *Psyscope*.

Procedimento:

O experimento, elaborado por meio do programa *Psyscope*, utilizou uma técnica *on-line* de leitura automonitorada (*self-paced reading*) em que os participantes monitoram sua própria leitura em frente à tela do computador e a uma caixa de botoeira, em uma sala isolada (sala do LAPEX – Laboratório de Psicolinguística Experimental). A tarefa consistia em ler, em velocidade natural, frases divididas em dez segmentos, como pudemos observar nos exemplos do Quadro I.

Os participantes foram testados individualmente. Todos foram primeiramente orientados oralmente pelo experimenter e depois por instruções que apareciam na tela do computador. O início da tarefa consistiu em ler o primeiro segmento e, ao apertar o botão amarelo da caixa de botoeira à sua frente, o participante fazia com que esse segmento

sumisse. Automaticamente, o segundo segmento aparecia e assim seguidamente, até o término do último segmento (final da frase), sinalizado com um ponto final. Logo em seguida, aparecia uma pergunta a respeito da frase lida e o participante tinha que responder apertando o botão verde para SIM e o botão vermelho para NÃO. Com essa pergunta objetivamos controlar a atenção e a compreensão dos participantes. Os tempos de todos os 10 segmentos foram gravados e também a opção de resposta (SIM ou NÃO) referente à pergunta feita ao final da frase. Se os participantes respondessem equivocadamente às perguntas os tempos da frase correspondente seriam eliminados.

Todos os participantes reportaram ter achado a tarefa simples e demoraram em torno de 12 minutos cada para executar a tarefa experimental. É importante ainda frisar que cada participante, antes do início da tarefa experimental, participou de uma prática contendo apenas frases com estruturas semelhantes às distratoras, divididas em 10 segmentos, assim como as frases utilizadas no experimento. Essa prática visa deixar os participantes familiarizados com a tarefa experimental.

*Resultados e discussão*⁵:

Os resultados do Experimento 1 estão explicitados no Gráfico 1, que contém as médias dos tempos de resposta em milissegundos encontradas para as duas condições experimentais (PR e NR). Atestando positivamente a hipótese prevista, a média dos tempos de leitura das retomadas anafóricas com pronomes lexicais (PR) foi significativamente menor que a média dos tempos de leitura das retomadas com nomes repetidos (NR), como podemos observar pela análise estatística aplicada, um teste-T: $t(49) = 2,32$; $p < 0,03$.

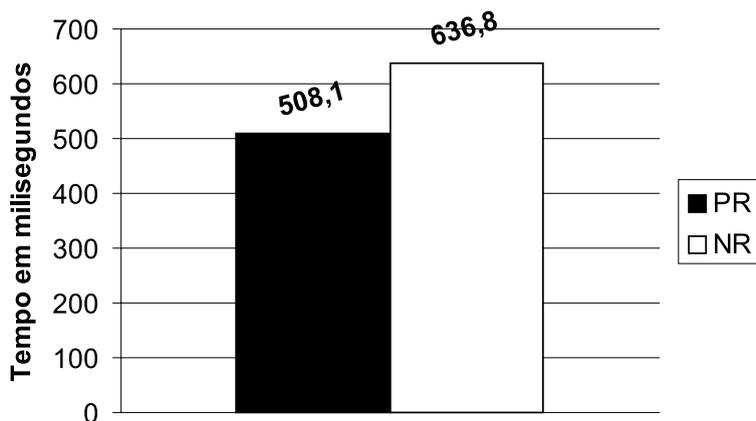


Gráfico 1 - Médias dos tempos de leitura de pronomes (PR) e nomes repetidos (NR)

Esse resultado mostra que ocorre Penalidade do Nome-Repetido assim como previsto pela teoria da centralização, que tem como um de seus princípios a eficiência dos pronomes no estabelecimento da co-referencialidade, sendo estes os veículos naturais para essa função na linguagem. Entretanto, tanto o princípio quanto a penalidade explicitados pela teoria da centralização dizem respeito a antecedentes proeminentes (basicamente na posição de sujeito) e a retomadas também em posição proeminente de sujeito (GORDON, HENDRICK, 1997, 1998).

Em nenhum dos estudos que seguem a teoria da centralização parece haver a menção ou a cogitação de que a retomada anafórica na posição de objeto siga esse mesmo princípio e essa mesma Penalidade do Nome-Repetido. Os resultados obtidos no Experimento 1 corroboram, dessa maneira, o que foi encontrado por CHAMBERS, SMYTH (1998) no inglês e por STREB, RÖSLER, HENNIGHAUSEN (1999) no alemão. Ambos mostraram que nomes repetidos são processados mais lentamente do que pronomes (Penalidade do Nome-

Repetido, nos termos da teoria da centralização) também em retomadas anafóricas em posição de objeto e estabeleceram, a partir desses resultados, uma crítica à exclusividade de tal penalidade para retomadas em posição de sujeito proposta pela teoria da centralização.

A Hipótese da Carga Informacional (*Informational Load Hypothesis – ILH*) proposta por ALMOR (1999, 2000) parece explicar consistentemente também os resultados referentes ao Experimento 1, já que a hipótese de Almor prevê também que retomadas anafóricas pronominais são processadas mais rapidamente do que nomes repetidos. Isso se deve a que as primeiras carregam menos traços semânticos (menor carga informacional) suficientes para identificar a representação discursiva de seu respectivo antecedente, caracterizando-se, assim, como retomadas mais gerais do que as retomadas anafóricas com nomes repetidos, que têm mais traços semânticos (maior carga informacional) para serem processados no estabelecimento da co-referência.

No Experimento 2, a seguir, observaremos a eficiência no estabelecimento de co-referência entre SNs mais gerais e SNs mais específicos, a partir de uma relação ora de hiperonímia, ora de hiponímia, com seus respectivos antecedentes. Com os resultados desse experimento, poderemos analisar e explicar mais abrangentemente a relação existente entre a forma do elemento anafórico e a eficiência desse elemento no estabelecimento da co-referência de objeto em PB. Além disso, poderemos testar efetivamente a adequação explicativa da hipótese da carga informacional de ALMOR (1999, 2000) em contraste com a explicação elaborada pela teoria da centralização (GORDON, HENDRICK, 1998).

2. Experimento 2

Este experimento tem como objetivo comparar, utilizando o paradigma on-line de leitura automonitorada (*self-paced reading*), o processamento de SNs superordenados, ou mais gerais, com o de SNs hipônimos, ou mais específicos, em posição de objeto em PB. Ambos retomam antecedentes com nível hierárquico intermediário dentro de um determinado subconjunto de hipônimos, e estão localizados em sentenças coordenadas anteriores também na posição de objeto direto.

A partir dos resultados encontrados no Experimento 1, corroboramos o que outros estudos *on-line* têm encontrado em relação à maior eficiência dos pronomes em termos de processamento da co-referência do que os nomes repetidos.

Esses mesmos resultados podem ser explicados tanto pelo conceito de penalidade do nome repetido proposta pela teoria da centralização, em que há uma penalidade a partir da quebra do princípio que prevê a preferência do uso de pronomes para o estabelecimento da co-referência, quanto pela hipótese da carga informacional proposta por Almor, em que o processamento mais lento por parte do nome repetido co-referencial, quando comparado ao pronome, ocorre por causa da sua carga informacional ser maior e por isso mais custosa para a memória de trabalho o que, conseqüentemente, dificulta o processamento da co-referência, como já explicitado no final da discussão do experimento anterior.

Para evidenciarmos qual das duas explicações é a mais adequada e abrangente, testaremos, no presente experimento, o processamento da co-referência estabelecida a partir de SNs superordenados, ou mais gerais, e de SNs hipônimos, ou mais específicos, em relação aos seus respectivos antecedentes. Verificaremos se os SNs superordenados são processados mais

rapidamente do que os SNs hipônimos como previsto pelos resultados de ALMOR (1999).

A variável independente manipulada neste segundo experimento é o tipo de SN utilizado na retomada anafórica: SN superordenado (S) ou SN hipônimo (H). Diferente do experimento anterior, temos duas variáveis dependentes: tempo de leitura aferido após a apresentação das retomadas anafóricas (S ou H) relacionadas aos seus respectivos antecedentes em estruturas sentenciais coordenadas e percentual de respostas Sim ou Não às perguntas apresentadas no fim da leitura das frases.

Para controlar possíveis influências do fator extensão do constituinte, assim como fizemos no Experimento 1, tentamos equilibrar o tamanho dos SNs superordenados em termos de sílaba (por exemplo: animal, vegetal), adequando-o ao tamanho dos SNs hipônimos (por exemplo: jacaré, pepino). Porém, em cinco estruturas experimentais das dez utilizadas, tivemos uma sílaba a mais para os SNs superordenados (por exemplo: veículo, alimento), o que não interfere na confiabilidade do experimento, já que essa variação de tamanho jogaria um papel a favor da hipótese nula, de que os SNs mais gerais são processados mais lentamente do que os SNs mais específicos. Com isso, apostamos que os resultados realmente reflitam o custo de processamento de SNs mais gerais e SNs mais específicos, independentemente do tamanho dos SNs, focalizando a diferença da carga informacional entre eles, que varia de acordo com os seus traços semânticos.

O Quadro 2, na página seguinte, fornece um exemplo de cada uma das condições experimentais que foram testadas no experimento.

Quadro 2⁶ – Exemplos das condições do experimento 2

(3) Retomada anafórica de antecedente nominal com SN superordenado (S)																		
Seg. 1	/	Seg. 2	/	Seg. 3	/	Seg. 4	/	Seg. 5	/	Seg. 6	/	Seg. 7	/	Seg. 8	/	Seg. 9	/	Seg. 10
Os vizinhos	/	adquiriram	/	um carro	/	na	/	loja	/	mas depois	/	venderam	/	o veículo	/	no	/	feirão.
Pergunta: Os vizinhos adquiriram o veículo?																		
(4) Retomada anafórica de antecedente nominal com SN hipônimo (H)																		
Seg. 1	/	Seg. 2	/	Seg. 3	/	Seg. 4	/	Seg. 5	/	Seg. 6	/	Seg. 7	/	Seg. 8	/	Seg. 9	/	Seg. 10
Os vizinhos	/	adquiriram	/	um carro	/	na	/	loja	/	mas depois	/	venderam	/	o chevette	/	no	/	feirão.
Pergunta: Os vizinhos adquiriram o chevette?																		

Em ambas as condições, o antecedente é um termo hierarquicamente intermediário dentro de uma estrutura de hiponímia, em relação aos elementos da retomada anafórica, isto é, em relação ao SN superordenado, o antecedente é um hipônimo e, em relação ao SN hipônimo, o antecedente é um hiperônimo. Por exemplo, o antecedente **carro** é hipônimo de **veículo**, mas é hiperônimo de **chevette**, já que chevette é um tipo de carro, e carro é um tipo de veículo.

Esse tipo de relação semântica estabelecida entre o antecedente e a retomada anafórica permite analisarmos se há diferença, em PB, entre o processamento da co-referência de termos mais gerais (superordenados) que os seus antecedentes, e de termos mais específicos (hipônimos) que os seus antecedentes, corroborando ou não a hipótese da carga informacional e os resultados de ALMOR (1999) para o inglês.

Além disso, estamos testando, assim como no Experimento 1, a hipótese de que elementos anafóricos, sejam eles pronomes ou expressões nominais, em posição não proeminente como a de objeto direto, estabelecem, de maneira natural e eficiente, a co-referência com um antecedente também em posição não proeminente (objeto direto).

2.1 Método

Participantes:

Foram voluntários deste experimento 26 participantes, falantes nativos de português, idade média de 19 anos, alunos de graduação da UFRJ, sendo 17 do sexo feminino e apenas 9 do sexo masculino.

Material:

O material consistiu de 2 conjuntos de 10 frases experimentais. Cada informante foi exposto a um desses conjuntos

experimentais, embutidos em um conjunto extra de 22 frases distratoras. Cada conjunto experimental é composto de 2 condições com 5 frases por condição. Em uma condição, a retomada anafórica do SN antecedente é feita por um SN superordenado (S), enquanto que, na outra condição, a retomada anafórica do antecedente é feita por um SN hipônimo (H). As frases experimentais são formadas por duas sentenças coordenadas e foram divididas em 10 segmentos, sendo que o segmento crítico é o segmento 8, em que se localiza a retomada com o SN superordenado ou com o SN hipônimo (ver quadro 2). Apesar das frases serem muito parecidas, diferindo apenas quanto à variável testada, a sua distribuição em 2 conjuntos permitiu que todas as frases fossem comparadas e que cada participante fosse exposto a apenas uma frase de cada tipo.

O aparato experimental consistiu de um iMAC, G3, 233Mhz e uma caixa de botoeira conectada ao computador, específica para utilização com o programa *Psyscope*.

Procedimento:

Assim como no Experimento 1, o Experimento 2, elaborado por meio do programa *Psyscope*, utilizou uma técnica *on-line* de leitura automonitorada (*self-paced reading*), em que os participantes monitoram sua própria leitura em frente da tela do computador e em frente de uma caixa de botoeira, em uma sala isolada. A tarefa consistia em ler, em velocidade natural, frases divididas em 10 segmentos, como pudemos observar nos exemplos do Quadro 2.

Os participantes foram testados individualmente. Todos foram primeiramente orientados oralmente pelo experimentador e depois por instruções que apareciam na tela do computador. O início da tarefa consistiu em ler o primeiro segmento

e, ao apertar o botão amarelo da caixa de botoeira à sua frente, o participante fazia com que esse segmento sumisse. Automaticamente, o segundo segmento aparecia e assim seguidamente, até o término do último segmento (final da frase), sinalizado com um ponto final. Logo em seguida, aparecia uma pergunta a respeito da frase lida e o participante tinha que responder apertando o botão verde para SIM e o botão vermelho para NÃO. Com essa pergunta, diferentemente do Experimento 1, aferíamos se na compreensão da frase os participantes estabeleciam a co-referência entre antecedente e os dois tipos de SN, verificando se houve uma dificuldade maior ou menor de acordo com o tipo de SN. Os tempos de todos os 10 segmentos foram gravados, e também a opção e o tempo de resposta (SIM ou NÃO) referente à pergunta feita ao final da frase.

Todos os participantes reportaram ter achado a tarefa simples e demoraram executando a tarefa experimental em torno de 15 minutos cada. Como no Experimento 1, cada participante, antes do início da tarefa experimental, participou de uma prática contendo apenas frases com estrutura semelhante às distratoras, divididas em 10 segmentos, como as frases experimentais, visando a familiarização dos participantes com a tarefa.

Resultados e discussão⁷

Os resultados do presente experimento mostram, primeiro, que, independentemente do SN co-referencial ser superordenado (S) ou hipônimo (H), os sujeitos, ao ler as frases experimentais, estabelecem a co-referência com o antecedente disponível. Isso pode ser confirmado no Gráfico 2, a seguir, em que temos a distribuição do percentual de respostas SIM e NÃO por tipo de SN. Notamos que, seja nas frases com SNs mais gerais (S), seja nas frases com SNs mais

específicos (H), o percentual de respostas SIM é significativamente maior do que o percentual de respostas NÃO, confirmando o estabelecimento da co-referência de uma maneira geral como aponta o Teste-T: $t(258)=2,70$; $p < 0,01$.

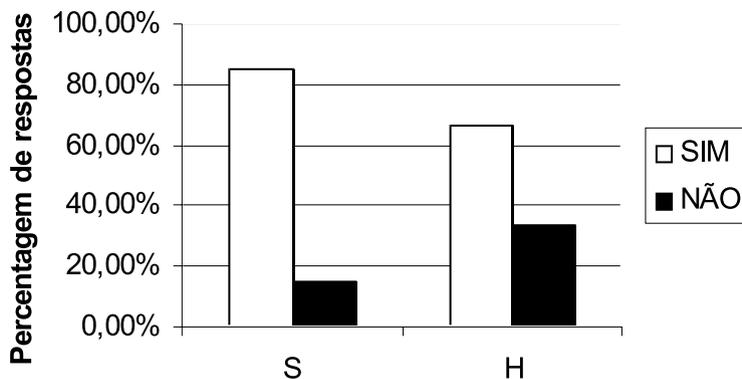


Gráfico 2 - Distribuição das percentagens de respostas SIM e NÃO entre o grupo de frases com SNs superordenado (S) e o grupo de frases com SNs hipônimos (H)

Além dessa primeira constatação, percebemos também que há uma diferença entre as condições (S) e (H) no que diz respeito à percentagem de respostas SIM e NÃO. Podemos observar no gráfico que a maioria do total de respostas NÃO ocorreu nas frases em que o objeto direto anafórico é expresso por SNs hipônimos e um percentual bem menor nas frases em que o objeto direto é expresso por SNs superordenados, revelando, aparentemente, uma maior dificuldade na interpretação co-referencial nas frases em que o elemento anafórico estabelece uma relação de hiponímia com seu respectivo antecedente.

A partir desses resultados do percentual de respostas expressos no Gráfico 2, evidenciamos, ao menos no nível da

interpretação, apoio empírico à hipótese da carga informacional, pois os SNs hipônimos, que têm maior carga informacional, dificultam o estabelecimento da co-referência e, algumas vezes, parecem ser interpretados como uma entidade discursiva nova. Porém, para que a hipótese possa ser corroborada de maneira consistente, temos que analisar os resultados dos tempos de leitura tanto dos SNs hipônimos, quanto dos SNs superordenados, aferindo assim o estabelecimento da co-referência de forma *on-line*.

Os resultados dos tempos de leitura estão expressos no Gráfico 3, a seguir.

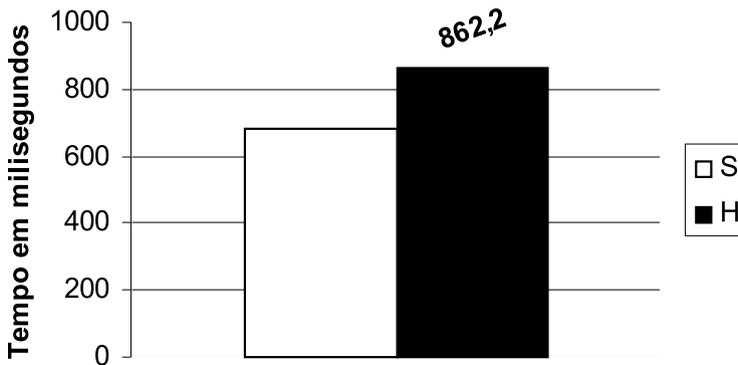


Gráfico 3 - Distribuição dos tempos de leitura das retomadas anafóricas expressas por SNs superordenados (S) e por SNs hipônimos (H)

Os tempos de leitura mostram que SNs superordenados são lidos mais rapidamente do que SNs hipônimos de forma significativa, como aponta o teste-T: $t(258) = 1,65$; $p < 0,01$. Sendo assim, a hipótese da carga informacional proposta por ALMOR (1999, 2000) é corroborada, já que SNs com mais traços semânticos a serem processados são lidos mais lentamente, enquanto SNs mais gerais, com menos traços semânticos,

acessados no estabelecimento da co-referência, são processados mais rapidamente.

Esses resultados não só corroboram a hipótese de Almor como põem em xeque a abrangência explicativa da teoria da centralização e o conceito de Penalidade do Nome-Repetido (GORDON, HENDRICK, 1998), pois se há um processamento distinto entre SNs co-referentes na posição de objeto, que penalidade explicaria isso, já que os dois tipos de SN violam o princípio do uso preferencial de pronomes para o estabelecimento de co-referência.

3. Considerações finais

Os resultados do Experimento 1 mostraram que, na posição de objeto direto em PB, pronomes são lidos mais rapidamente do que nomes repetidos, corroborando com resultados em outras línguas que utilizaram técnicas *on-line* de aferição (GORDON, CHAN, 1995; GORDON, HENDRICK, 1998; CHAMBERS, SMYTH, 1998).

Tanto a teoria da centralização com base no conceito de Penalidade do Nome-Repetido quanto a Hipótese da Carga Informacional têm uma explicação adequada para esses resultados do Experimento 1. Quando a co-referência é estabelecida por meio de um nome repetido há, segundo a teoria da centralização, violação de um princípio e, por conseguinte, ocorre uma penalidade em termos de processamento. Já a hipótese de ALMOR (1999) explica o processamento menos custoso de retomadas com pronome em contraste a retomadas com nomes repetidos com base na diferença de carga informacional entre essas duas formas lingüísticas.

O Experimento 2 mostrou que, na posição de objeto direto em PB, SNs mais gerais (superordenados) são lidos mais rapidamente do que SNs mais específicos (hipônimos). Nesse

caso, a teoria da centralização não prevê nenhum tipo de penalidade e, portanto, não tem uma explicação para esses resultados. Por outro lado, a hipótese da carga informacional dá conta também adequadamente desses resultados, por se basear na diferença de traços semânticos entre os dois tipos de retomada anafórica. Assim, como ocorre entre pronomes e nomes, respectivamente, SNs superordenados são menos custosos no estabelecimento da co-referência porque têm menos traços semânticos a serem processados na memória de trabalho, enquanto que SNs hipônimos são mais custosos porque têm mais traços semânticos a serem processados, o que também corrobora os resultados de ALMOR (1999) para o inglês.

Os resultados de ambos os experimentos mostraram também efeitos significativos em termos de processamento co-referencial entre as formas lingüísticas investigadas, relacionados à posição não proeminente de objeto direto, tanto das retomadas anafóricas quanto dos respectivos antecedentes também em posição de objeto, diferentemente dos experimentos relacionados à teoria da centralização, que se debruçam apenas sobre retomadas em posição de sujeito (GORDON, CHAN, 1998; GORDON, HENDRICK, 1998, entre outros). Com isso, não só a abrangência explicativa da teoria da centralização é colocada em xeque, como também o próprio efeito de proeminência referente ao processamento da co-referência proposto por essa teoria.

Sendo assim, os resultados aqui apresentados corroboram os resultados e as críticas em relação à teoria da centralização expressas tanto por ALMOR (1999, 2000), quanto por CHAMBERS, SMYTH (1998), que encontraram também para o inglês uma diferença no processamento entre pronomes e nomes repetidos em posição não proeminente de objeto.

A partir dos resultados apresentados, parece que cumprimos, ao menos em parte, a tarefa de expandir os

estudos da co-referência em posição de objeto no PB, em particular, na área de processamento lingüístico. Obviamente, ainda temos muito que avançar nessa área, para que, em conjunto com os estudos em outras línguas, possamos obter uma visão mais ampla e precisa do processamento da co-referência, fenômeno tão comum na linguagem humana.

Notas

- ¹ Os termos “anáfora” e/ou “retomada anafórica” estão sendo usados neste trabalho para descrever qualquer expressão que estabeleça co-referência com um antecedente, incluindo SNs definidos (por exemplo: o animal) e pronomes lexicais (por exemplo: ele ou ela). Esse uso do termo “anáfora” é diferente do uso empregado pela gramática gerativa em que esse termo faz referência, primordialmente, aos reflexivos e aos recíprocos (MIOTO et al., 2004; CHOMSKY, 1981, 1986).
- ² Cabe aqui o registro de que existem estudos (por exemplo: GERNSBACHER, 1989) que, utilizando a técnica de pré-ativação (*priming*) e reconhecimento de palavras (*probe word task*), têm encontrado resultados divergentes desses já citados, mas tais estudos são bastante criticados na literatura por não estarem capturando um efeito no estabelecimento da co-referência e sim efeitos apenas de reconhecimento lexical. Não entraremos aqui nessa discussão, pois apenas apresentaremos resultados de experimentos de leitura automonitorada.
- ³ Os exemplos em português mantêm relações semânticas semelhantes aos tipos de antecedentes e de retomadas anafóricas estudados por ALMOR (1999, 2000) em inglês. Um exemplo utilizado por Almor em inglês é composto pelo antecedente *bird*, retomado ora por *animal*, ora por *robin*.
- ⁴ As barras que dividem as frases mostram os dez segmentos que os sujeitos do experimento leram de maneira não cumulativa.
- ⁵ Os cálculos estatísticos foram elaborados com a colaboração do Professor Doutor Gastão Coelho do Departamento de Métodos e Estatística do Instituto de Matemática da UFRJ.
- ⁶ As barras que dividem as frases mostram os 10 segmentos aos quais os sujeitos do experimento leram de maneira não cumulativa.
- ⁷ Os cálculos estatísticos foram elaborados com a colaboração do Professor Doutor Gastão Coelho do Departamento de Métodos e Estatística do Instituto de Matemática da UFRJ.

Coreference processing of nouns and pronouns in Brazilian Portuguese

Abstract – This paper has as its main goal to investigate the roles played by nouns and pronouns, in terms of coreference processing, in anaphoric relations of direct objects in Brazilian Portuguese. Besides that, it intends to test and to contrast the Repeated-Name Penalty hypothesis (GORDON, HENDRICK, 1998) and the Informational Load Hypothesis (ALMOR, 1999), which are related respectively to efficiency of pronouns better than nouns and more general noun phrases (NP) better than specific NPs in what concerns to coreference processing.

Key words – Discourse processing. Coreference. Anaphoric direct object.

Referências bibliográficas

ALMOR, A. Noun-phrase anaphora and focus: the informational load hypothesis. *Psychological Review*, v. 106, n. 4, p. 748-765, 1999.

_____. Constraints and mechanisms in theories of anaphor processing. In: PICKERING et al. (Ed.). *Architectures and mechanisms for language processing*. Cambridge, UK: Cambridge University Press, 2000.

CHAMBERS, C.; SMYTH, R. Structural parallelism and discourse coherence: a test of centering theory. *Journal of Memory and Language*, v. 39, p. 593-608, 1998.

CHOMSKY, N. *Lectures on government and binding*. Dordrecht: Foris, 1981.

_____. *O conhecimento da língua*. Lisboa: Caminho, 1986.

GARROD, S.; FREUDENTHAL, D.; BOYLE, E. The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, v. 33, p. 39-68, 1994.

GERNSBACHER, M.A. Mechanisms that improve referential acces. *Cognition*, v. 32, p. 99-156, 1989.

Gordon, P.C.; GROSZ, B.J.; GILLION, L.A. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, v. 17, p. 311-347, 1993.

GORDON, P.C.; CHAN, D. Pronouns, passives and discourse coherence. *Journal of Memory and Language*, v. 34, p. 216-231, 1995.

GORDON, P.C.; HENDRICK, R. Intuitive knowledge of linguistic coreference. *Cognition*, v. 62, p. 325-370, 1997.

_____; _____. The representation and processing of coreference in discourse. *Cognitive Science*, v. 22, n. 4, p. 389-424, 1998.

KENNISON, S.; GORDON, P.C. Comprehending referencial expressions during reading: evidence from eye tracking. *Discourse Processes*, v. 24, p. 229-252, 1997.

LEITÃO, M.; MAIA, M. Processamento na interface sintaxe - semântica: objeto direto anafórico e traço de animacidade. In: FINGER, I.; MAIA, M. (Eds.). *Processamento da linguagem*. Pelotas: Educat, 2005.

MIOTO, C.; SILVA, M.C.F.; LOPES, R.E.V. *Novo manual de sintaxe*. Florianópolis: Insular, 2004.

STREB, J.; RÖSLER, F.; HENNIGHAUSEN, E. Event-related responses to pronoun and proper name anaphora in parallel and non-parallel discourse structures. *Brain and Language*, v. 70, n. 2, p. 273-286, 1999.

Apresentado para publicação em julho de 2005.

Aprovado para publicação em julho de 2005.