



Resolução do pronome nulo em Português Brasileiro: Evidência de movimentação ocular

Elisângela Nogueira Teixeira (UFC)
Maria Cristina Micelli Fonseca (UFC)
Maria Elias Soares (UFC)

RESUMO: Ainda são controversas as razões pelas quais certas línguas possuem um sistema pronominal duplo, formado de um nulo e um pleno, que aparentemente exercem a mesma função. Tomando por base a Hipótese da Posição do Antecedente (CARMINATI, 2002) e a Teoria da Acessibilidade (ARIEL, 1991), este artigo examina o sistema pronominal do Português Brasileiro. Procuramos evidências de custo de processamento na movimentação ocular de participantes durante um teste de compreensão leitora de frases ambíguas, em que foram manipulados o tipo e a posição da correferência anafórica. A análise dos resultados sugere que o pronome nulo desempenha o papel de desfazer ambiguidades, estabelecendo correferência com o antecedente em posição de proeminência sintática.

Palavras-chave: correferência; processamento anafórico; pronome nulo; pronome pleno; movimentação ocular.

Introdução

Este artigo apresenta os resultados de uma pesquisa¹ sobre a correferência de pronomes, nulos e plenos, em posição anafórica e catafórica, em períodos complexos compostos por subordinação. Trata-se de pesquisa não sobre restrições gramaticais às formas pronominais, mas sobre o processamento dos pronomes nulo e pleno em função de fatores de ordem sintática. Para realizar a pesquisa, conduzimos um experimento de rastreamento ocular durante a leitura de frases ambíguas, formadas por uma oração principal e por uma oração subordinada temporal, contendo esta última um sujeito representado por um pronome nulo (a categoria vazia [*pro*]) ou por um pronome pleno.

O objetivo de nosso estudo de compreensão leitora é testar a hipótese de que há uma função específica para cada uma das formas pronominais do Português Brasileiro (PB). Uma conjectura adicional deste estudo é a de que, qualquer que seja a expressão anafórica pronominal, em línguas com sujeito nulo e pleno, os pronomes nulos tendem a estabelecer correferência com a entidade mais saliente no discurso que, em geral: (i) ocupa a posição mais alta da oração; (ii)

¹ A pesquisa apresentada neste artigo foi desenvolvida com o auxílio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) ao projeto nº 481899/2011-1.

possui proeminência sintática de posição de sujeito e (iii) satisfaz a noção de tópico discursivo.

Nosso estudo se insere particularmente no âmbito dos esforços que buscam compreender as razões da especialização pronominal (de sua falta ou presença) em PB, por meio de evidências de processamento *on-line*. Estudos desta natureza, que investigam o sistema do PB, ainda são contribuições necessárias por aportar evidências a construções teóricas. Não iremos fazer referência explícita aqui à controvérsia sobre o *status* do pronome nulo em PB, apesar de esta pesquisa tangenciar o assunto, sobretudo aportando dados empíricos sobre o processamento de tais pronomes durante a recuperação do seu correferente.

Se o sistema pronominal do PB ainda possui especialização funcional, poderíamos hipotetizar que tal sistema estaria relacionado com outros aspectos cognitivos, como a capacidade de armazenamento da memória de trabalho, no que diz respeito ao grau de acessibilidade dos antecedentes em relação a posição estrutural da retomada (ARIEL, 1991). Se esta hipótese não pode ser testada e se a teoria linguística não oferece ainda uma explicação plausível para este sistema pronominal, é preciso continuar a investigar as razões que levam línguas a manterem em seu sistema, na mesma sincronia, a presença concomitante de uma categoria vazia e uma plena.

A anáfora correferencial é um dos temas que fazem parte da lista de problemas controversos e ainda não totalmente resolvidos por teorias linguísticas. O problema consiste na identificação da entidade anteriormente mencionada no discurso como referência para o termo anafórico. Anáfora², do grego *αναφορα*, (*ανα*, acima ou atrás e *φορα*, ação de levar), quer dizer “levar para cima ou para trás”, ou seja, retomar uma entidade, evento ou termo anteriormente apresentado no discurso. A entidade mencionada ou referida é chamada de antecedente. Em um sentido lato, a anáfora se refere tanto às entidades já mencionadas quanto às que irão ser mencionadas no discurso. No sentido estrito, se diferencia do conceito de catáfora, que precede a entidade que ainda será mencionada ou introduzida.

A anáfora pronominal representa o problema de modo prototípico, já que é a expressão anafórica, segundo Ariel (1991), que ocupa a base da escala de acessibilidade. Ariel (1991) entende que as expressões anafóricas de uma língua marcariam uma gradação hierárquica de acessibilidade uma vez que as representações mentais dos referentes discursivos (entidades) não são igualmente acessíveis em qualquer estágio do discurso. O pronome, portanto, estaria na base dessa escala já que não traz informação semântica nenhuma que permita identificar um único antecedente quando vários possíveis co-existem, além do fato de que as informações carregadas por um pronome em língua portuguesa são basicamente morfossintáticas, restritas à concordância de gênero e número. Diferentemente de anáforas sob a forma de descrições definidas que fornecem informação sintática e semântica, a identificação do antecedente de uma anáfora pronominal requer que este seja altamente acessível e recente na memória de trabalho, o que configura um problema longe de ter uma solução trivial.

A escolha entre um pronome nulo e um pronome pleno em línguas *pro-drop* tem sido investigada nos últimos anos em diferentes línguas, relacionadas e não-relacionadas com o PB³ (ALONSO-OVALLE *et al.*, 2002; BARBOSA; DUARTE; KATO, 2005; CARMINATI, 2002; CORREA, 1998; COSTA; FARIA; MATOS, 1998; MAIA, 1997; FOLEY; VAN VALIN, 1984; FONSECA; GUERREIRO, 2012; HUANG, 1989; LUEGI, 2012; MERIDOR, 2006; MORGADO, 2011; MELO; MAIA, 2005).

² O termo *anáfora* é usado aqui em sentido tradicional.

³ Línguas relacionadas são línguas provenientes do mesmo ramo.

Para comentar alguns desses trabalhos, optamos por mencionar os resultados de Carminati (2002), que reuniu uma série de evidências experimentais para sustentar fortemente a hipótese de que nas línguas com sistema pronominal composto por dois pronomes, um nulo e um pleno, os pronomes estão em distribuição complementar, apresentando cada um deles uma função especializada. Seu trabalho testou tal hipótese no escopo da oração. A hipótese de Carminati (2002), que examinou a língua italiana, ficou conhecida como Hipótese da Posição do Antecedente, que seria uma estratégia do *parser* de atribuir para os pronomes nulos os antecedentes em posição estrutural de sujeito e para os pronomes plenos os antecedentes em posição estrutural relativamente mais baixa à posição do sujeito.

Para Carminati (2002), durante a resolução anafórica, a categoria vazia *pro* se liga a antecedentes proeminentes. Seu trabalho também sugere que a atribuição inicial do antecedente é estruturalmente baseada e que este ocupa a posição de Spec IP da oração. Sorace e Filiaci (2006) observaram que, nos casos de ambiguidade, em que é preciso reanalisar o predicado para desfazê-la, há aumento de custo de processamento, que refletiria, segundo as autoras, uma estratégia contrária à Hipótese da Posição do Antecedente. Tal constatação quer dizer que, embora a Hipótese da Posição do Antecedente seja um princípio altamente eficiente para resolver as dependências entre pronome e antecedente em línguas de sujeito nulo, é improvável que tal hipótese esteja no nível do processador sintático⁴, uma vez que a violação deste princípio não leva à agramaticalidade, mas somente a inapropriações.

Nos dados analisados por Carminati (2002) também foi observado que há uma diferença de força entre pronome nulo e pleno. Dito de outro modo, significa que, enquanto a preferência do pronome nulo pelo antecedente em posição de Spec IP é muito robusta, o pronome pleno apresenta certa flexibilidade para estabelecer correferência com a posição estrutural de objeto. Esta flexibilidade também pode ser observada nos resultados de Sorace e Filiaci (2006), reproduzidos na Tabela 1, abaixo, e em nossos resultados, analisados na seção 2 deste artigo.

Almor (1999), no entanto, já defende que a multiplicidade de formas referenciais em várias línguas não é um problema para ser explicado via processador sintático, mas que seria, antes, uma espécie de solução, usada por nosso sistema cognitivo, para contornar as limitações da nossa memória de trabalho.

Diante deste quadro controverso, o experimento que construímos procura testar se ainda há especialização sintática da categoria vazia *pro* para falantes do PB, em um esforço que se soma aos trabalhos recentemente concluídos sobre o Português Europeu e o Português Brasileiro (MORGADO, 2011; LUEGI, 2012). Luegi (2012) investigou em vários experimentos o estatuto do pronome nulo comparando as variedades brasileira e portuguesa. Controlou primordialmente a função sintática e a posição estrutural, manipulando a ordem canônica da língua portuguesa. Seu objetivo era encontrar interação entre os dois fatores (restrição sintática e posição estrutural). Suas evidências *off-line* sugerem que sujeitos nulos estabelecem correferência com antecedentes em função de sujeito e em posição estrutural mais saliente (ou na posição mais alta da oração). Já suas evidências *on-line* sugerem que nas primeiras leituras a correferência se liga mais à posição estrutural do que à função sintática, sugestão esta que se modifica nas segundas leituras e reanálises que se seguem.

Numa linha de estudos semelhantes, mas investigando o estatuto do nulo em posição de objeto, Maia (1997) estudou em PB a realidade psicológica da categoria vazia na função de

⁴ O experimento apresentado neste artigo não discute a questão do *locus* da resolução anafórica.

objeto, ocorrência particularmente bem mais frequente em PB do que em Português Europeu, em que raramente ocorre. Com base na restrição do pronome pleno de Montalbetti (1984), que diz que em várias línguas românicas os pronomes plenos, diferentemente dos pronomes nulos, não podem ser ligados localmente, os resultados de Maia (1997) apontam na direção de uma forte correlação para a correferência entre antecedente proeminente (tópico e posição estrutural) e categoria vazia em posição de objeto.

O estudo de Duarte (1995) sobre a perda do parâmetro de sujeito nulo para o PB advoga que está havendo em PB o preenchimento da posição de sujeito nulo em primeira e segunda pessoas verbais, já que nestas pessoas haveria perda da característica morfossintática número pessoal dos verbos. Em relação à terceira pessoa, o pronome *ele*, Duarte (1995) sugere, no entanto, que o preenchimento desta pessoa parece ser apenas uma tendência, porque a frequência de atualização da terceira pessoa verbal observada foi relativamente menor do que a encontrada para as primeiras e segundas pessoas verbais. Esta constatação é em si mesma uma justificativa para continuar a compreender o papel de um sistema pronominal duplo em PB.

Como já dissemos, o objeto do experimento aqui reportado é a resolução anafórica em período complexo, formado por uma oração principal e uma oração subordinada temporal. Na oração principal há dois argumentos do verbo, um DP na posição de sujeito e um outro DP na posição de objeto, cujo núcleo é um nome próprio, frequente em PB, como “a Maria”, “a Lúcia”, “o David”, “o André”. As características morfológicas dos nomes próprios foram equiparadas, porque ambos podem ser do gênero masculino ou feminino. A oração subordinada temporal é formada, por sua vez, de uma conjunção temporal “quando” ou “enquanto” e de uma oração cujo sujeito pode ser um pronome pleno “ele” ou “ela” ou um pronome nulo (a categoria vazia *pro*). A subordinada temporal ainda poderia anteceder ou suceder a oração principal, criando, com o antecedente, uma relação catafórica (quando antecede a oração principal) ou anafórica (quando sucede a oração principal), como nos mostram os exemplos (1) e (2).

- (1) [A Maria]_i; conversava com [a Joana]_j enquanto [pro]_{i/j}/[ela]_{i/j} cozinhava.
- (2) Enquanto [pro]_{i/j}/[ela]_{i/j} cozinhava, [a Maria]_i conversava com [a Joana]_j.

O que objetivamos examinar foi o processamento, em um experimento *on-line*, e a decisão do falante de PB diante de uma ambiguidade, em um experimento *off-line*, que consistia em uma competição entre potenciais antecedentes para a correferência de pronomes nulos e plenos. Com o registro da movimentação ocular na tarefa de compreensão, procuramos indícios de reanálise ou de elevação de custos de processamento relacionáveis à resolução da correferência. Estes indícios relativos foram comparados entre as condições experimentais.

Além do exame das funções do pronome nulo e do pleno, observamos também o efeito que a posição catafórica ou anafórica do pronome poderia provocar, baseando-nos nos estudos de Sorace e Filiaci (2006) e de Fonseca e Guerreiro (2012), que procuraram comprovação das conjecturas de Carminati (2002), fazendo variar a posição da expressão anafórica em relação ao antecedente. A variação da posição entre catáfora e anáfora tem como objetivo observar efeitos de antecipação na ligação entre as expressões correferentes. Diferentemente do estudo aqui reportado, a pesquisa de Sorace e Filiaci (2006), investigou o processamento anafórico em segunda língua.

Percentual de respostas comportamentais de atribuição do antecedente				
	Anáfora		Catáfora	
	Pronome nulo	Pronome Pleno	Pronome nulo	Pronome Pleno
Sujeito	51%	8%	85%	12%
Objeto	44%	81%	11%	24%
Outro	5%	11%	4%	64%

Tabela 1: Dados do experimento de Sorace e Filiaci (2006).

Na Tabela 1, acima, os percentuais mais altos encontrados por Sorace e Filiaci (2006) estão marcados em negrito. Note-se que a catáfora parece causar um efeito na atribuição da categoria vazia ao primeiro competidor apresentado. Seria este um efeito devido às limitações da memória de trabalho, como sugere Almor (1999), ou um efeito de restrição do processador sintático? Observando-se o percentual de correferência entre o sujeito em posição anafórica e o pronome nulo, constata-se que os resultados de Sorace e Filiaci (2006) sugerem um comportamento aleatório, ou seja, em torno dos 50%, o que sugere a não capacidade do participante de língua italiana de desfazer a ambiguidade sugerida durante a leitura do período complexo. O que se percebe também na leitura da Tabela 1 é que, embora Sorace e Filiaci (2006) tenham apresentado três opções (sujeito, objeto e outro), as decisões dos participantes se concentraram entre sujeito e objeto.

O estudo de Fonseca e Guerreiro (2012) apresenta resultados bastante semelhantes aos de Sorace e Filiaci (2006). Algumas diferenças foram percebidas, no entanto. Uma delas diz respeito à correferência entre antecedente e pronome pleno em posição catafórica, como se pode perceber ao analisar a Tabela 2 abaixo, que resume os resultados encontrados. Os participantes ficaram em dúvida e escolheram praticamente o mesmo percentual tanto para sujeito quanto para objeto, o que não se revelou no estudo de Sorace e Filiaci (2006) em que os participantes atribuíram a correferência a um terceiro, não mencionado na oração.

	Anáfora		Catáfora	
	Pronome nulo	Pronome Pleno	Pronome nulo	Pronome Pleno
Sujeito	81%	26%	96%	41%
Objeto	17%	74%	3%	40%
Outro	0%	0%	0%	14%
Não opina	2%	0%	1%	5%

Tabela 2: Dados do experimento de Fonseca e Guerreiro (2012).

Uma segunda diferença é percebida ao compararmos os resultados da posição anafórica. Houve em PB uma preferência clara pela correferência entre o pronome nulo em posição anafórica e o sujeito da oração principal, diferentemente do que encontraram Sorace e Filiaci (2006), cujos resultados não permitem afirmar qualquer preferência uma vez que parecem aleatórios (51% para sujeito contra 44% para objeto).

Na próxima seção, reportamos os materiais e métodos do nosso experimento que investigou como nosso processador sintático resolve as restrições obrigatórias, além de ter procurado investigar igualmente se há preferências de co-indexação para nulos e plenos. Procuramos, portanto, os indícios que nos levem a definir melhor quais são essas restrições obrigatórias e se existem preferências (facultativas) que podem nos sugerir a influência de traços semânticos e pragmáticos, ou de outros sistemas cognitivos, na resolução da correferência pronominal em PB.

1. Materiais e Métodos

Participantes

O grupo testado foi composto de vinte e dois (N=22) participantes saudáveis, adultos, com visão normal ou corrigida ao normal, sem histórico neurológico prévio e livres da ingestão de medicamentos que afetam o sistema nervoso central. Não foram escolhidos, para participarem dos experimentos, indivíduos que apresentavam astigmatismo, mesmo corrigido pelo uso de óculos. Esta decisão foi tomada após a verificação de que havia maior perda de dados dos participantes astigmatas. Os dados demográficos, na Tabela 3, abaixo, descrevem N (o número de participantes), sexo, idade, nível de educação e lateralidade. Os participantes foram convidados a colaborar voluntariamente com o estudo. Assinaram consentimento formal, aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Ceará sob o número 111/11, estavam previamente cientes de que suas identidades estariam preservadas e de que teriam sua movimentação ocular registrada durante a leitura dos estímulos.

Dados demográficos	
N	22
Sexo	20 Masculino / 2 Feminino
Idade	21.8 (\pm 1.28)
Nível educacional	14 anos de escolarização ⁵
Lateralidade	21 Destros / 1 Canhoto
Desvio padrão entre parênteses	

Tabela 3: Dados demográficos do estudo sobre processamento de pronomes nulos e plenos.

Estímulos e Procedimento experimental

O conjunto de estímulos experimentais foi distribuído em quatro diferentes listas. Cada lista foi composta por 16 frases experimentais entre 48 frases distratoras relacionadas⁶, totalizando quatro listas com 64 frases cada, sequenciadas aleatoriamente. As variáveis

⁵ Os participantes eram alunos matriculados no 4º semestre de graduação da Universidade Federal do Ceará.

⁶ Seguindo as tendências de produção de distratoras em Psicolinguística, este experimento usou orações distratoras com estrutura sintática semelhante a dos estímulos experimentais, com o objetivo de diminuir a chance de haver um possível efeito de consciência do teste.

independentes do estudo são *tipo de pronome* (nulo e pleno) e *posição* (anafórica e catafórica), distribuídas em quatro condições experimentais entre as 16 frases de cada lista. Cada participante leu um conjunto formado por quatro frases na condição de pronome nulo em posição anafórica (PNAN), quatro na condição de pronome pleno em posição anafórica (PPAN), quatro na condição de pronome nulo em posição catafórica (PNCA) e quatro na condição de pronome pleno em posição catafórica (PPCA), configurando um *design* em quadrado latino do tipo 2 x 2.

Somando-se todas as quatro listas, registramos a movimentação ocular de 352 frases (16*22). O número de 22 participantes é o resultado de decisão metodológica por excluir dois participantes cuja movimentação ocular não chegou ao limiar⁷ mínimo estabelecido para considerar os dados válidos para análise.

Cada lista iniciava por uma fase de treinamento. Os procedimentos de coleta consistiam em convidar os participantes para passarem por um teste que demoraria em torno de 20 minutos, compreendidas as fases de treinamento e experimental. Os testes foram conduzidos individualmente. Os participantes tiveram seus movimentos oculares gravados a uma distância de aproximadamente 60 cm do monitor do rastreador ocular Tobii T120. O programa usado para apresentação e registro dos estímulos foi o Tobii Studio 2.3.2. Todos os testes ocorreram em uma sala do Laboratório de Sistemas Complexos do Departamento de Física da Universidade Federal do Ceará.

Os procedimentos padronizados de calibração foram aplicados a todos os participantes. Participantes que não atingiram o nível desejado na fase de calibração não chegaram a ser testados.

As frases experimentais foram exibidas individualmente no centro da tela do rastreador ocular, posição em que percebemos maior acurácia do ponto de fixação do olhar. As pequenas variações de captura, como se pode ver na Figura 1, são compensadas pelo desenho de regiões de interesse que se ajustam às variações no ângulo de captura dos olhos. A Figura 1 é apenas ilustrativa dos tempos de primeira leitura, não representa a forma como os estímulos eram apresentados aos participantes.

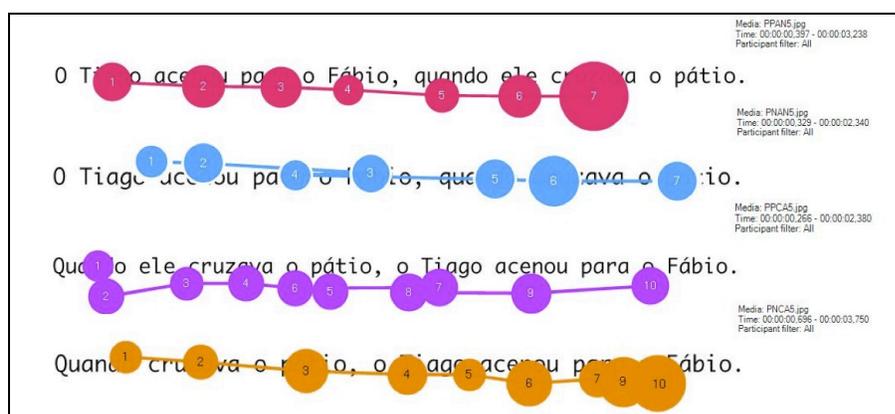


Figura 1: Exemplos de variação na posição de fixações oculares.

⁷ Consideramos como limiar mínimo o índice de 85% de registro ocular.

Como dito acima, os participantes foram distribuídos por listas experimentais (seis participantes por lista). As frases foram exibidas em uma linha simples, em fonte Monaco, monoespçada, de 25 pontos. Uma vez lidas as frases, o participante clicava no *mouse*, ação que levava à exibição de uma nova tela, em que era projetada a pergunta de controle. O participante respondia na própria tela, escolhendo com o *mouse* uma das opções de resposta (no caso, os competidores – sujeito e objeto – da oração principal, tanto para as frases experimentais quanto para as distratoras).

Os materiais experimentais foram baseados nos estímulos produzidos por Carminati (2002), no entanto foram inteiramente reconstruídos e adaptados para a língua portuguesa. Com a intenção de neutralizar, ao máximo, o viés que pode ocasionar os aspectos de frequência, aplicamos vários controles para a construção dos estímulos experimentais. Em relação aos verbos da oração principal, privilegiamos verbos bastante frequentes, tais como os verbos “conversar”, “falar”, “morar”, etc.

Em relação aos competidores, escolhemos nomes próprios com mais de um milhão de ocorrências no buscador *Google*. Todos os nomes próprios das frases experimentais possuem cinco letras. As frases experimentais possuem em média 57 caracteres com espaço⁸.

Todos os controles aplicados nos oferecem mais segurança para que a ambiguidade construída não pudesse ser resolvida em nenhuma fase da leitura, ou seja, para minimizar o efeito de qualquer pista, quer sintática, semântica ou pragmática, que viesasse os dados. Os exemplos (3), (4), (5) e (6), abaixo, representam um conjunto experimental. Cada uma destas frases era disposta em uma das listas do experimento.

- (3) Lista 1/PPAN6 [O André]_i reconheceu [o David]_j, quando [ele]_{i/j} olhava pela janela.
- (4) Lista 2/PPCA6 Quando [ele]_{i/j} olhava pela janela, [o André]_i reconheceu [o David]_j.
- (5) Lista 3/PNAN6 [O André]_i reconheceu [o David]_j, quando [pro]_{i/j} olhava pela janela.
- (6) Lista 4/PNCA6 Quando [pro]_{i/j} olhava pela janela, [o André]_i reconheceu [o David]_j.

As perguntas de compreensão, feitas após a leitura das frases, seguiam a seguinte forma:

(q.1) Quem olhava pela janela?

O André - O David

Equipamento

Os movimentos oculares foram gravados no rastreador ocular de 120 Hz, da marca Tobii. O sistema faz uma amostragem e gravação dos movimentos e fixações das pupilas dos participantes enquanto escaneiam e leem as sentenças, as questões e as alternativas de respostas. O rastreador ocular é embutido em uma tela de 17”. Duas câmeras de alta velocidade fotografam os olhos em uma amostragem de 120 imagens por segundo para cada olho. O equipamento usado possui um *software* que compensa os movimentos leves de cabeça, o que elimina a necessidade de um repouso para queixo e testa. Neste estudo, consideramos apenas os dados válidos para

⁸ A maior frase experimental tinha 63 caracteres com espaço. A menor frase continha 52 caracteres com espaço.

ambos os olhos⁹, que foram posteriormente processados usando o software do Tobii, o programa *Excel* e o programa de estatística SPSS. O filtro utilizado para identificar uma fixação ocular foi configurado com o limiar máximo de 35 *pixels* de velocidade por 35 *pixels* de distância. Foram examinados todos os segmentos das frases, mas as regiões relativas aos competidores foram preferencialmente analisadas.

1.2. Variáveis dependentes

1.2.1. Performance comportamental

A resposta comportamental avaliada é uma medida *off-line*. Trata-se de um julgamento, feito após a leitura da frase, que procura verificar a compreensão do leitor. No caso, as respostas indicavam se o leitor havia estabelecido correferência com a entidade proeminente da oração principal ou com a entidade menos proeminente, o objeto da oração. Reportamos três diferentes visualizações dos resultados. Na primeira visualização, pode-se observar a soma dos resultados e o cálculo do percentual de atribuições para todas as condições na combinação de 2x2. Na segunda e na terceira visualização, pode-se ver os resultados agrupados para as condições de anáfora e catáfora e para as condições de pronome pleno e nulo.

1.2.2. Tempo total de leitura

Esta é a medida que representa o tempo total de exibição do estímulo ao participante, que tem a liberdade de, ao clicar no *mouse*, passar de tela. A medida vai do tempo 0 (zero), correspondente ao *onset*, até o tempo F (final), correspondente ao último instante (*offset*) de exibição da imagem.

1.2.3. Tempo total de fixação nas regiões de interesse

Esta medida representa a soma das durações de todas as fixações dentro das duas regiões de interesse correspondentes à região do sujeito e à região do objeto da oração principal, entendidas como a área dos competidores. A medida do tempo total de fixação contempla todas as leituras possíveis feitas naquela região. Os participantes costumam realizar mais de uma leitura, às vezes, até cinco leituras, sobretudo, se as orações são ambíguas.

1.2.4. Duração média de fixação nos segmentos do período complexo

A duração média de fixação é calculada fazendo-se a soma de todas as fixações e dividindo-a pelo número de fixações. Foram feitos dois cálculos de duração média da fixação.

⁹ Apenas foram considerados os registros de rastreamento ocular que capturaram o movimento dos dois olhos, o que configura este estudo como binocular.

Em primeiro lugar, a medida foi calculada para as regiões dos competidores. Em segundo lugar, foi calculada a medida para as demais regiões do período, com o objetivo de observar se havia algum efeito tardio de processamento, possível de ser inferido a partir da medida de duração média da fixação.

2. Resultados

2.1. *Performance comportamental*

A correferência dos pronomes nulos e plenos foi preferencialmente estabelecida com o sujeito em posição canônica em todas as condições, como nos mostra o Gráfico 1, com exceção da condição de pronome pleno em posição anafórica, em que se verificam valores em torno dos 50%, um resultado curioso, pois revela uma escolha aleatória, o que merece atenção porque sugere que o papel do pronome “ele” na anáfora não está claro para o falante do PB. Encontrou-se efeito principal para a variável *tipo de pronome vs. posição* ($F [1,42]=11,318, p=0,002$).

Os resultados comportamentais encontrados nos surpreenderam porque não estavam dentro da predição que havíamos feito, a da especialização funcional dos pronomes nulo e pleno. Interessante é notar que, em posição catafórica, os participantes fizeram a correferência entre o pronome pleno e a posição de objeto em 30% dos casos (Gráfico 1). Este índice é bem mais alto do que os valores encontrados para o pronome nulo em relação à correferência com o objeto. O pronome nulo seleciona em 90,34% das vezes o sujeito como antecedente (Gráfico 2). O resultado do pronome nulo não surpreende e confirma as predições feitas a partir da Hipótese da Posição do Antecedente. Já o papel do pronome pleno, embora observemos um percentual de 54% de correferência com o antecedente em posição de objeto na anáfora (Gráfico 1), o maior entre todas as condições, não corresponde às predições feitas, alinhando-se antes à flexibilidade do pronome pleno observada por Carminati (2002).

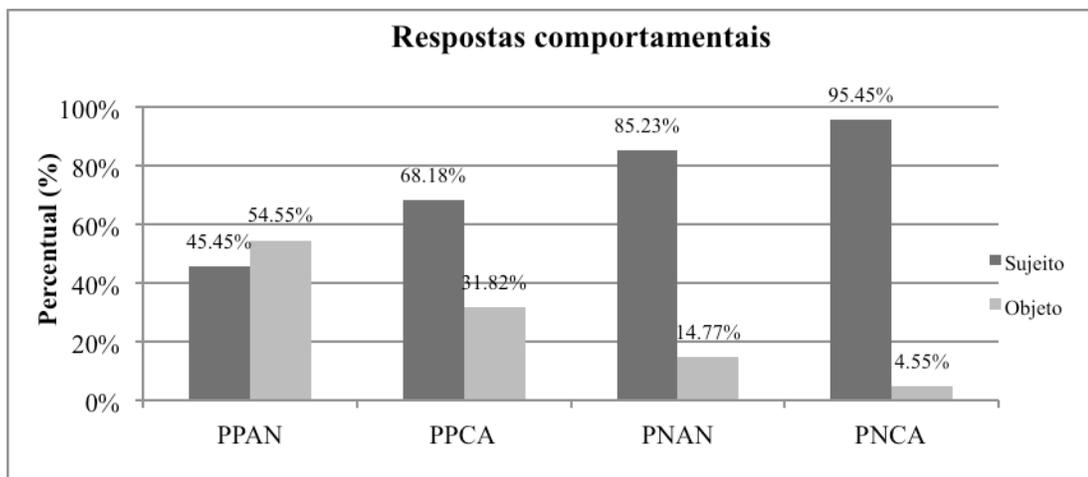


Gráfico 1: Percentual dos tipos de resposta por condição. Pronome nulo em posição anafórica (PNAN), Pronome pleno em posição anafórica (PPAN), Pronome nulo em posição catafórica (PNCA) e Pronome pleno em posição catafórica (PPCA).

O que está bastante evidente é que sob a condição de nulo os participantes claramente identificam o sujeito como antecedente. Nossos resultados sustentam a interpretação de que o pronome pleno não está apto para resolver a ambiguidade imposta nas orações por competidores do mesmo gênero, do mesmo número e nas posições sintáticas estudadas. É importante salientar que, se observadas as preferências entre todas as condições, o sujeito é preferido numa proporção de 3:1 em relação ao objeto, como pode-se observar no Gráfico 3, abaixo.

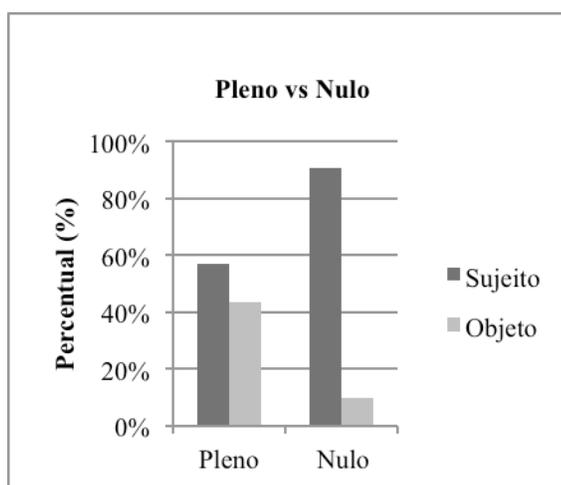


Gráfico 2: Percentual de resposta por tipo de forma pronominal.



Gráfico 3: Preferência pela posição de sujeito em todas as condições experimentais.

Como anteriormente dito, o controle durante a produção dos estímulos sob a forma de períodos complexos, observando-se o tamanho e a frequência dos competidores, procurou afastar qualquer tendência para a correferência com a posição de sujeito. No entanto, o paralelismo das orações principal e subordinada, pode ainda desempenhar um papel fundamental na atribuição da correferência, como discutiremos na seção final deste artigo.

2.2. *Movimentação ocular*

Passemos agora a examinar os resultados encontrados no estudo da movimentação ocular dos participantes. Como apresentado na seção 1.4, iremos reportar os dados de três diferentes medidas de movimentação ocular.

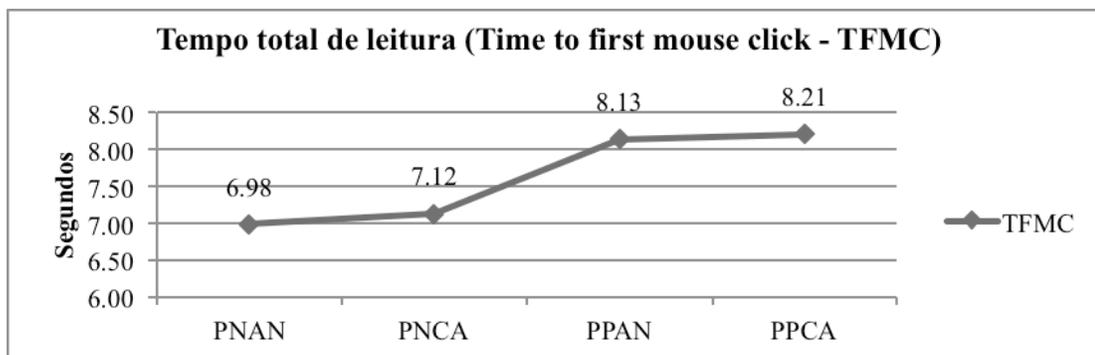


Gráfico 4: Tempo total de leitura em cada condição experimental¹⁰. Os tempos estão distribuídos entre as condições: Pronome nulo em posição anafórica (PNAN), Pronome pleno em posição anafórica (PPAN), Pronome nulo em posição catafórica (PNCA) e Pronome pleno em posição catafórica (PPCA).

A primeira medida, que apresentamos no Gráfico 4, é relativa ao tempo total de leitura dos períodos complexos. O Gráfico 4 apresenta um quadro geral que revela tempos de leitura maiores para as condições com o pronome pleno em relação ao tempo de leitura das frases com pronome nulo. Comparando-se a posição do pronome pleno, percebe-se que há mais custo para a catáfora do que para a anáfora, assim como ocorre na condição do pronome nulo, já que registramos tempos maiores para a catáfora do que para a anáfora. Não é um resultado estranho pois a catáfora parece ser uma estratégia mais raramente utilizada, mesmo na língua escrita. Apesar de haver uma tendência, detectável pelo crescimento da curva da esquerda para a direita, a análise da variância não demonstra haver significância na interação das condições experimentais ($F[3,19]=1,805$, $p=0,180$). Talvez o aumento da amostra pudesse encontrar efeito significativo. Caso fosse encontrado tal efeito, seria preciso ainda investigar se seria decorrente de um aumento de custo para processar a correferência na presença do pronome pleno ou se decorreria apenas do aumento no número de caracteres das orações experimentais.

¹⁰ TMFC é uma sigla em inglês correspondente a “time to first mouse click”. Este tempo é calculado subtraindo-se do tempo final (t_f), momento que o participante clica no *mouse*, o tempo inicial (t_i), relativo ao início da exibição do estímulo.

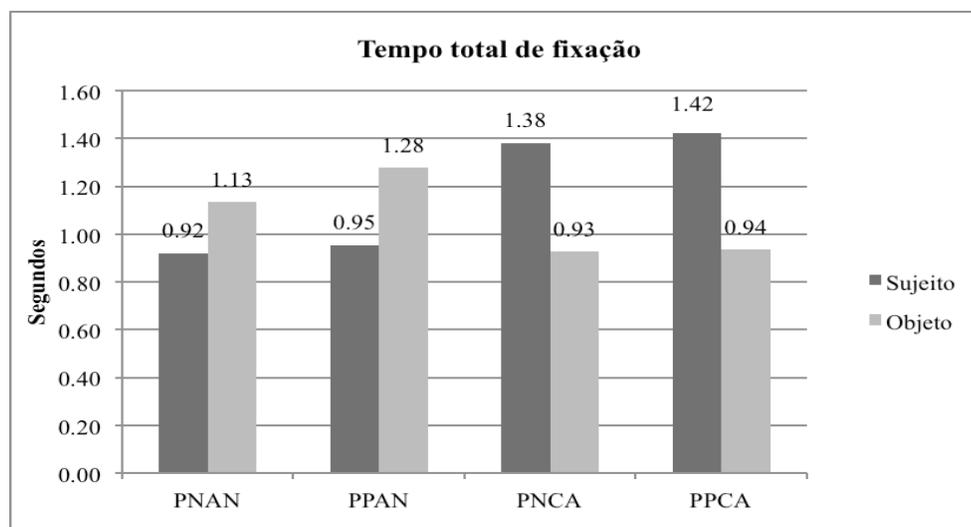


Gráfico 5: Tempo total de fixação nas regiões de interesse. Os tempos estão distribuídos entre as condições: Pronome nulo em posição anafórica (PNAN), Pronome pleno em posição anafórica (PPAN), Pronome nulo em posição catafórica (PNCA) e Pronome pleno em posição catafórica (PPCA).

Quando examinamos o tempo total de fixação no Gráfico 5, encontramos efeitos significativos na interação entre o fator *posição* e o fator *tipo de pronome* ($F[1,20]=10,344$, $p=0,004$)¹¹. O que se constata com esse resultado é que, na condição de pronome pleno anafórico, os participantes fixaram por mais tempo a região do objeto, o que talvez possa servir de indício para a explicação de maior atribuição de correferência ao objeto nesta mesma condição (PPAN). No entanto, apenas este resultado não pode ser tomado como indicativo sem se considerar a análise conjunta dos demais tempos totais de fixação. Em linhas gerais, pode-se dizer que os participantes fixaram, por mais tempo, as regiões do sujeito na condição de catáfora, sendo o inverso constatado para a condição de anáfora.

O resultado acima descrito pode ser analisado de muitas maneiras. Em primeiro lugar, se constatamos mais tempo nas regiões do sujeito para a condição de catáfora, isto pode querer dizer que a catáfora impõe dificuldade na atribuição do antecedente tanto para um pronome nulo (1.38s) quanto para um pronome pleno (1.42s). Após a leitura das frases, como já ficou claro, os participantes precisavam optar entre o sujeito e o objeto. O que observamos na decisão posterior, *off-line*, é que o sujeito foi majoritariamente preferido (> 80%) nesta condição de catáfora (ver Gráfico 6). Na condição de anáfora, percebemos que o tempo total fixado na região do objeto foi relativamente maior do que na região do sujeito, embora os dados comportamentais não atestem para a condição anafórica a mesma correspondência da condição catafórica. Esta correspondência nem poderia ser atestada, uma vez que o nulo estabelece correferência com o sujeito. O que, sim,

¹¹ Para a análise destes resultados, precisamos excluir os dados de um participante porque não foram encontrados registros de suas fixações nas regiões dos competidores. Portanto, a análise foi feita considerando-se 21 participantes, no lugar de 22 como foi descrito na seção 1.

se verifica, é que o índice de correferência com o objeto é maior, considerada a média, para a condição de anáfora (34.66%) do que de catáfora (18.18%), na razão de 1.9:1 (ver Gráfico 6).

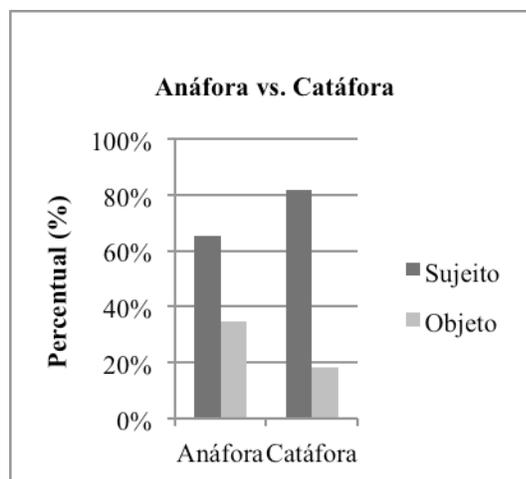


Gráfico 6: Contraste entre anáfora e catáfora (medida *off-line*).

O terceiro dado de movimentação ocular que iremos analisar são os dados relativos à duração média de fixação (Gráfico 7). Rodamos a análise da variância (ANOVA) para observar a interação entre o fator *tipo de pronome* e o fator *posição*, mas não obtivemos resultados significativos. No entanto, é importante perceber que, uma vez retirado o fator posição, encontramos diferença significativa no tempo de duração média da fixação para o fator *tipo de pronome*, ou seja, para as condições PNAN e PPAN ($F[1,19]=13,529$, $p=0,002$)¹².

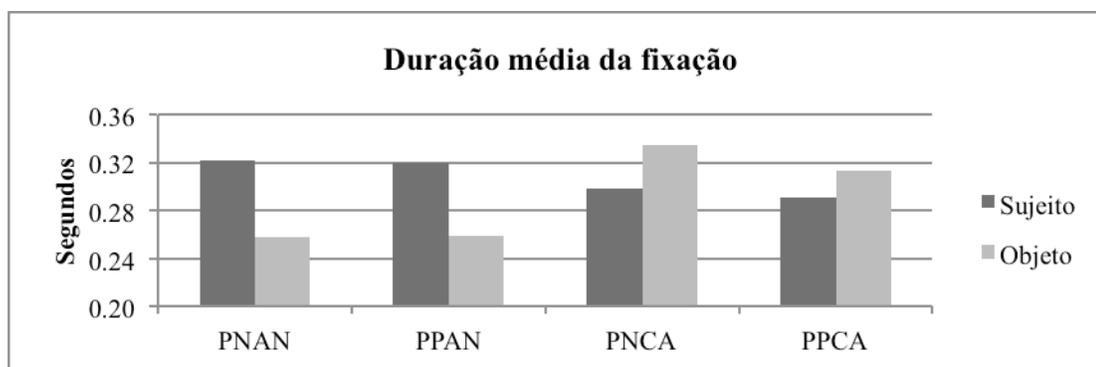


Gráfico 7: Duração média das fixações nas regiões de interesse. Os tempos estão distribuídos entre as condições: Pronome nulo em posição anafórica (PNAN), Pronome pleno em posição anafórica (PPAN), Pronome nulo em posição catafórica (PNCA) e Pronome pleno em posição catafórica (PPCA).

¹² Para este cálculo foram excluídos os dados de dois participantes, cujas médias estavam fora da curva da normalidade.

Além de observar as medidas isoladamente, é ainda interessante comparar os gráficos 5 e 7, porque nos chama atenção o fato de que a duração média de fixação nas condições PNAN e PPAN (Gráfico 7) para a região do objeto é bem menor (0.261s e 0.262s) do que para a região do sujeito 0.331s (PNAN) e 0.325s(PPAN). Já na medida do tempo total de fixação (Gráfico 5) observamos o contrário. Na região do objeto, o tempo total de fixação (1.13s e 1.28s) é maior do que na região do sujeito (0.92s e 0.95s) para as mesmas condições experimentais PNAN e PPAN. O raciocínio é similar quando comparamos os dados na condição de catáfora (PNCA e PPCA). Uma primeira interpretação para essa aparente discrepância seria a constatação de que os participantes inspecionaram mais vezes a região do objeto, ou seja, fizeram mais visitas¹³ à região do objeto, mas se detiveram menos tempo nesta região, ou seja, as visitas eram breves, mas em maior quantidade. Já a região correspondente ao sujeito da oração foi visitada em menor quantidade, mas essas visitas foram mais longas do que as visitas à região do objeto. Isso explicaria porque, apesar de haver mais tempo total dispensado na região do objeto, a correferência foi estabelecida com o sujeito, que está na região em que as visitas permaneceram por mais tempo, em que houve, portanto, maior custo de processamento.

Nos Gráficos 8 e 9, apresentamos a duração média das fixações em toda a extensão do período complexo. Este gráfico foi elaborado com a intenção de se observar custos de processamento em outras regiões do período, que servissem como índice de processamento tardio ou antecipatório e também para comparar a leitura nas condições de pronome nulo e pleno.

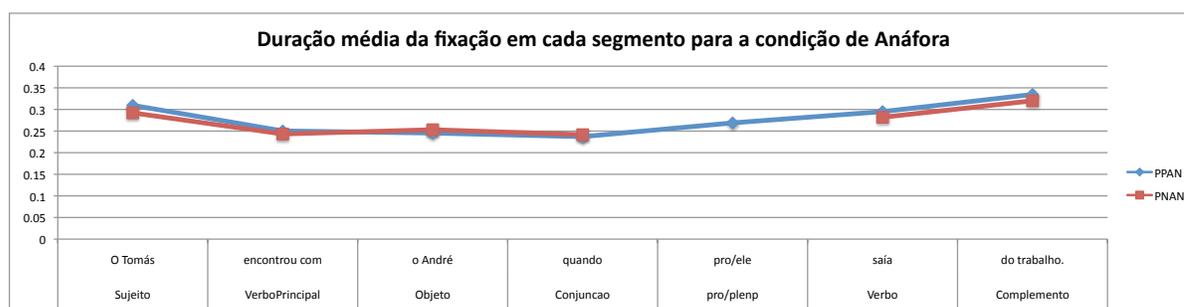


Gráfico 8: Duração média da fixação em cada segmento para a condição de Anáfora.

¹³ O termo visita está sendo usado aqui como cada fixação ocular em uma região específica. Dizer que uma região recebeu mais de uma visita implica em dizer que após a primeira leitura houve um número $x + 1$ de refixações na mesma região.

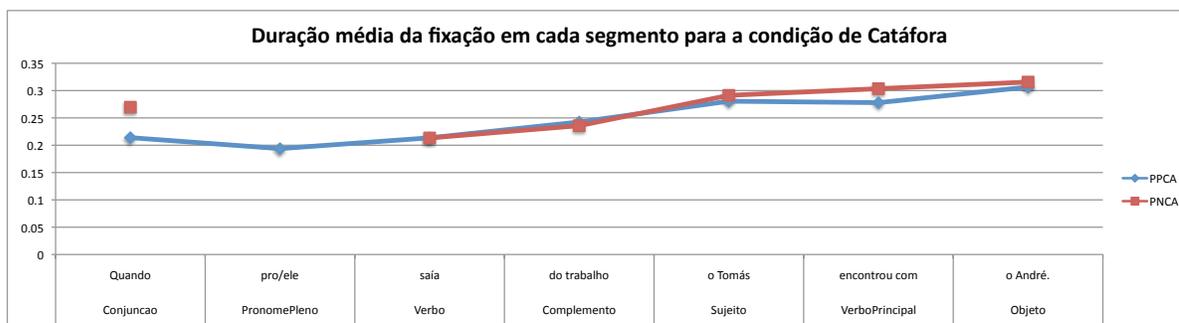


Gráfico 9: Duração média da fixação em cada segmento para a condição Catáfora.

Observamos que não há praticamente diferenças nos tempos médios de duração entre pronomes nulos e plenos quando se observa separadamente a condição de anáfora e catáfora. Estes dados, no entanto, não nos permitem afirmar que o processador se comporta igualmente em ambas as condições porque representam apenas a duração média de fixação, que é um índice interessante para se analisar a diferença entre o tempo de primeira leitura e primeira fixação e os tempos de releitura da frase. Para a Psicolinguística apenas o tempo de duração média, sem contraste com demais medidas, não tem se mostrado relevante para a análise do processamento da linguagem.

Como dito, Duarte (1995) e outros pesquisadores, em estudos de *corpora* sobre o Português Brasileiro, já vêm observando a tendência para o preenchimento da posição de sujeito nulo com o uso do pronome pessoal “ele”. Esta diferença entre produção, escrita e oral, e compreensão, auditiva e leitora, pode ser uma chave que nos ajude a entender a distribuição do pronome “ele” e sua função na língua portuguesa, em todas as suas variantes, já que os estudos de compreensão leitora, isoladamente, não têm ainda aportado contribuições definitivas para esta questão.

Conclusão

O estudo, reportado neste artigo, procurou observar, em períodos complexos compostos por subordinação, efeitos de posição (catáfora e anáfora) e de tipo de expressão anafórica (pronome nulo e pleno), em termos de custos de processamento e em termos de preferências no estabelecimento da correferência.

Assim como dito na Introdução, a hipótese de especialização das funções de um sistema pronominal composto por uma categoria vazia e um pronome pleno está ainda a ser confirmada para o PB. O que podemos dizer é que nossos resultados comportamentais e de movimentação ocular corroboram achados de pesquisas anteriores (CORREA, 1998; COSTA; FARIA; MATOS, 1998; FONSECA; GUERREIRO, 2012; LUEGI, 2012; MORGADO, 2011; MELO; MAIA, 2005), que atribuem à categoria vazia, a preferência no estabelecimento da correferência com o antecedente mais saliente, em posição de sujeito da oração. A análise dos dados comportamentais nos sugeriu ser possível distinguir no sistema pronominal do PB um grau de acessibilidade

distinto para os pronomes nulos e plenos. O pleno tenderia a se ligar a antecedentes menos acessíveis, enquanto o pronome nulo se ligaria àqueles mais salientes e acessíveis.

Nossa afirmação se baseia no fato de que os participantes claramente preferiram estabelecer correferência entre o pronome nulo e a posição do sujeito em todas as condições experimentais, com subordinadas em posição anafórica ou catafórica. Na resposta ao estudo comportamental, decidimos por não oferecer ao leitor a possibilidade de escolher um referente outro para o pronome pleno ou nulo, porque observamos índices ínfimos para a opção “outro” nos resultados de Sorace e Filiaci (2006) e de Fonseca e Guerreiro (2012), comentados na introdução.

Diante dessa escolha forçada, por assim dizer, a análise dos nossos resultados comportamentais nos sugere que a ambiguidade é resolvida através da presença da categoria vazia na oração, que aponta claramente para o antecedente em posição de Spec IP, em uma relação cuja robustez nos exige classificar o fenômeno como restrição obrigatória do sistema de regras do falante do português do Brasil. Uma vez constatada a regularidade, é de se supor que não se deveria falar de mudança de parâmetro de sujeito nulo em PB, já que a resposta comportamental é por demais nítida, indicando a coerência que faz com que a categoria vazia preencha suas características semânticas e sintáticas com o antecedente em posição de sujeito da oração principal.

O pronome pleno não parece resolver a ambiguidade na condição de anáfora, já que os resultados comportamentais encontrados apresentam índices aleatórios (em torno dos 50%), o que contradiz achados para outras línguas relacionadas ao português, como o italiano (CARMINATI, 2002) e o espanhol (ALONSO-OVALLE *et al.*, 2002).

A análise das medidas de duração média de fixação ocular sugeriu que os competidores escolhidos como correferentes são fixados por mais tempo ainda que em um menor número de vezes. Observamos ainda que a posição produz um efeito nos custos de processamento, que são inversamente proporcionais para sujeito e objeto. Encontramos que cada fixação ocular do participante na região do sujeito foi mais longa, em torno de 0.32s, mas a soma de todas as fixações revela que o olhar permaneceu menos tempo nesta região, em torno de 0.92s. Dizemos, portanto, que os participantes realizaram em média 3 visitas à região do sujeito durante a leitura das frases. Já na região do objeto, cada fixação ocular foi mais curta, em torno de 0.26s. Acontece que os participantes fizeram mais visitas a esta região, numa média um pouco maior do que 4 vezes, já que a soma do tempo de visitas foi de 1.13s. O número de fixações é obtido dividindo-se a soma pela média.

As visitas mais curtas à região do objeto podem revelar que houve consciência da ambiguidade do estímulo, mas a ambiguidade não foi suficiente para aumentar a correferência entre o pronome nulo e o objeto. As médias de duração da fixação ocular também revelam a consciência do participante em relação à ambiguidade, uma vez que são muito superiores à média padrão na leitura de textos não manipulados em PB, que gira em torno de 0.212s (TEIXEIRA, 2013). Comparando a média padrão à média de 0.320s na leitura da região do sujeito da oração principal, região esta que continha 7 caracteres com espaço, por exemplo “O André”, somos levados a crer que houve consciência da ambiguidade da oração.

As medidas de duração média e tempo total de fixação observadas em conjunto sugerem uma relação entre posição e distância do antecedente para a forma pronominal. Uma vez que os dados comportamentais (medida *off-line*) demonstraram preferência em estabelecer a correferência entre catáforas e o nome próprio na região do sujeito, é de se pensar que o fator

distância esteja sendo considerado para a resolução anafórica. A condição catafóra revelou um aumento no tempo total de fixação na região do sujeito enquanto a condição anáfora revelou tal aumento na região do objeto. Como dito, se considerarmos que o alto número de visitas significa aumento de custo de processamento, pode-se inferir que a medida do número de fixações (tempo total dividido pela duração média) corrobora os resultados da medida do tempo total de fixação. A relação entre estas duas medidas *on-line*, portanto, nos faz observar que um fator de distância entre o termo correferente e seu antecedente estaria vinculado ao fator posição (catafórica e anafórica). A posição catafórica levaria ao aumento no custo de processamento do competidor mais próximo (o sujeito), percebido pelo tempo total e número de fixações, enquanto a posição anafórica levaria ao aumento do custo de processamento do competidor mais próximo (neste caso, o objeto). Se considerarmos a noção de acessibilidade (ARIEL, 1991), a saliência discursiva do sujeito representaria um papel fundamental estabelecendo a correferência com a forma nula ou plena na posição catafórica, embora os resultados para a forma plena apresentem menor robustez do que os da forma nula, uma vez que o pleno tende a buscar um correferente externo ao discurso ou em posição menos saliente.

Em conjunto, a análise dos dados sugere que percebemos dois extremos em que estariam a condição de pronome nulo em posição catafórica (PNCA) e a condição de pronome pleno em posição anafórica (PPAN), já que a condição PNCA registra a preferência pelo sujeito e a condição PPAN pelo objeto.

Acreditamos que novos estudos, com controle de outras variáveis, tais como recência e paralelismo, além da inclusão de sintagmas no início da frase, para colocar os competidores no centro da frase, podem nos fornecer mais dados para nos levar com segurança a uma conclusão a respeito do papel dos dois diferentes pronomes no sistema do PB e do comportamento do *parser* nesta tarefa.

Finalmente, queremos esclarecer que os resultados que encontramos, no experimento aqui reportado, e que o modo como os interpretamos nos permitem confirmar a hipótese deste artigo, segundo a qual, em períodos complexos, o DP que ocupa a função de sujeito da oração principal é o antecedente preferencialmente retomado para estabelecer correferência com a pronome nulo. Ainda é importante deixar claro que, como compreendemos a imensa variabilidade de situações de comunicação, acreditamos que a preferência ou a obrigatoriedade na resolução anafórica de pronomes nulos e plenos deva ser exaustivamente testada em busca de comprovação empírica definitiva.

Competition disambiguation through null pronouns in Brazilian Portuguese: Evidence from eye movements

ABSTRACT: It is still controversial why some languages have a dual pronominal system, consisting of a zero and full pronoun that apparently perform the same function. Based on the Position of Antecedent Hypothesis (CARMINATI, 2002) and the Theory of Accessibility (ARIEL, 1991), this paper examines the pronominal system of Brazilian Portuguese. We look for evidences of processing cost in eye movement of participants during a test of reading comprehension of ambiguous sentences where the type and position of the anaphoric coreference are manipulated. Our results suggest that the null pronoun establish coreference with the antecedent in a syntactic prominence position

Keywords: anaphoric processing; null pronoun; full pronoun; eye movement.

REFERÊNCIAS

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

ALONSO-OVALLE, L., CLIFTON, C., FRAZIER, L., SOLERA, S. F. Null vs. Overt Pronouns and The Topic-Focus Articulation in Spanish. *Journal of Italian Linguistics*, v. 14, n. 2, p. 151-169, 2002.

ARIEL, M. The function of accessibility in a theory of grammar. *Journal of Pragmatics*, v. 16, n. 5, p. 443-464, 1991.

BARBOSA, P., DUARTE, E., KATO, M. Null Subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics*, v. 4, n. 2, p.11-52, 2005.

CARMINATI, M.N. *The processing of Italian subject pronouns*. 2002. Dissertation (PhD in Linguistics) – Department of Linguistics, University of Massachusetts, Amherst, 2002.

CORRÊA, L. Acessibilidade e paralelismo na interpretação do pronome sujeito e o contraste pro/pronome em português. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, v. 14, n.2, p. 1-24, 1998.

COSTA, A., FARIA, I. H., MATOS, G. Ambiguidade referencial na identificação do Sujeito em estruturas coordenadas. In: ENCONTRO NACIONAL DA ASSOCIAÇÃO PORTUGUESA DE LINGÜÍSTICA, XIII, 1997, Lisboa, *Atas do XIII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Colibri, 1998. p. 173-188.

DUARTE, M. E. L. A perda do Princípio "Evite Pronome" no Português Brasileiro. 1995. Tese (Doutorado em Estudos da Linguagem) – Instituto de Estudos da Linguagem, Universidade de Campinas, Campinas, 1995.

FOLEY, W. A.; VAN VALIN, R. D. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press, 1984.

FONSECA, M. C. M.; GUERREIRO, E. Resolução de Correferência Pronominal no Português do Brasil. *Revista Linguística*, v. 8, n. 2, p. 112-133, 2012.

HUANG, Y. *Anaphora in Chinese: toward a pragmatic analysis*. 1989. Dissertation (PhD in Linguistics) – Trinity College, University of Cambridge, Cambridge, 1989.

LUEGI, P. *Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes*. 2012. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 2012.

MAIA, M. The processing of Object Anaphora in Brazilian Portuguese. *Recherches Linguistiques de Vincennes*, v. 26, p.151-172, 1997.

MELO, M. F. B.; MAIA, M. A. R. O processamento da correferência do sujeito pronominal em sentenças formadas por verbos de comunicação linguística no português do Brasil. *Revista Linguística*, v. 1, n. 2, p. 177-205, 2005.

MERIDOR, H. *An Experimental Investigation of the Antecedent Preferences of Hebrew Subject Pronouns*. 2006. Master of Science (Master on Cognitive Science and Natural Language Processing) – School of Informatics, Edinburg University, Edinburg, 2006.

MONTALBETTI, M. *After Binding: on the interpretation of pronouns*. 1984. Dissertation (PhD in Linguistics) – Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, 1984.

MORGADO, S. M. *Processamento da co-referência pronominal: informação sintáctica e semântica*. 2011. Dissertação (Mestrado em Linguística) - Faculdade de Letras, Universidade de Lisboa, Lisboa, 2012.

SORACE, A., FILIACI, F. Anaphora resolution in near-native speakers of Italian. *Second Language Research*, v. 22, n. 3, p. 339-368, 2006.

TEIXEIRA, E. N. *Preferências sintáticas e semânticas no processamento da correferência anafórica: evidências de movimentação ocular*. 2013. Tese (Doutorado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, Fortaleza, 2013.

Data de envio: 15/10/2013

Data de aceite: 19/02/2014

Data de publicação: 21/07/2014