

Referenciação e técnicas experimentais: aspectos metodológicos na investigação do processamento correferencial em português brasileiro

Reference and Experimental Techniques:
Methodological Aspects in the Investigation of
Co-Referential Processing in Brazilian Portuguese

Jefferson de Carvalho Maia*
Universidade Federal de Minas Gerais
Maria Luiza Cunha Lima
Universidade Federal de Minas Gerais

Resumo

Neste artigo, discutimos algumas questões metodológicas envolvidas na realização de experimentos psicolinguísticos, com ênfase em métodos *on-line* de experimentação, no paradigma de leitura autocadenciada e em estudos sobre referenciação. Apresentamos, também, resultados de dois experimentos de leitura autocadenciada sobre o processamento correferencial, em português brasileiro, de nomes repetidos, pronomes plenos e pronomes nulos, tanto em retomada de antecedentes em posição de sujeito quanto de objeto. Como subproduto de tal pesquisa, obtivemos evidências de que o processamento de pronomes tônicos e clíticos em posição de objeto direto anafórico é sensível à modalidade específica de língua (fala ou escrita) em questão, fato que aponta para a necessidade de controle rigoroso das diferenças entre fala e escrita durante a elaboração de experimentos de leitura.

Palavras-chave

Metodologia experimental, leitura autocadenciada, processamento correferencial, pronomes tônicos e clíticos, fala e escrita.

Abstract

In this paper we discuss some methodological issues concerning the realization of psycholinguistic experiments, with emphasis on on-line experimental methods, on the self-paced reading paradigm and on studies about reference. We also present the results of two self-paced reading experiments on the co-referential processing of repeated names, overt pronouns and null pronouns referring to antecedents both in subject and in object positions in Brazilian Portuguese. As a by-product of this research, we have found evidence that the processing of strong and weak pronouns in the position of anaphoric direct object is sensitive to the specific language mode (oral or written) on focus. This suggests the need of a strict control of differences between oral and writing production in the construction of reading experiments.

Keywords

Experimental methodology, self-paced reading, co-referential processing, strong and weak pronouns, oral production and written production.

0. Introdução

As ciências cognitivas, em suas muitas subdisciplinas, têm se dedicado a entender os processos através dos quais opera a mente humana. No tocante à linguagem, um dos grandes desafios é compreender os aspectos cognitivos do texto, isto é, os processos linguísticos mentais que, independentes ou relacionados a outros sistemas cognitivos (como memória e atenção), permitem aos ouvintes/leitores, por exemplo, acessar itens lexicais, decodificar estruturas sintáticas e construir os sentidos de textos orais/escritos.

A investigação de tais processos linguísticos de produção, percepção e representação tem se dado, do ponto de vista metodológico, principalmente por meio da realização, em laboratório, de experimentos psicolinguísticos, que precisam ser rigorosamente controlados, de modo que se possa isolar a influência das variáveis de interesse sobre o fenômeno em observação, minimizando ou eliminando efeitos indesejáveis de variáveis outras que possam interferir nos resultados da pesquisa. Esse processo de eliminação de variáveis incontroláveis integra o que é comumente chamado de planejamento experimental, procedimento crucial na determinação da confiabilidade dos resultados. Assim sendo, o objetivo deste artigo é somar à literatura brasileira sobre planejamento experimental em psicolinguística, especificamente no que tange ao controle das diferenças entre as modalidades falada e escrita da língua quando da elaboração de experimentos, fim que será perseguido através da discussão dos resultados de dois estudos sobre processamento correferencial em português brasileiro, avaliando estruturas anafóricas típicas de situações de fala e escrita. Antes, contudo, de descrever os experimentos mencionados, faz-se necessário discorrer sobre os paradigmas experimentais atualmente disponíveis aos pesquisadores em psicolinguística.

Tais paradigmas podem ser dispostos em dois grandes grupos, a saber, métodos que envolvem medidas fisiológicas e métodos que envolvem medidas comportamentais. Ao passo que os primeiros se baseiam na observação de reações fisiológicas involuntárias, sejam elas eletromagnéticas (ERP, MEG) ou

hemodinâmicas (PET scan, fMRI), os segundos medem a atividade mental através de reações voluntárias dos sujeitos do experimento diante de alguma tarefa linguística.¹ Os métodos de medidas comportamentais, que são os mais disponíveis no Brasil e aqueles de relevância para este estudo, podem, por sua vez, ser agrupados em métodos *off-line* e métodos *on-line* de experimentação.

Experimentos *off-line* (e.g., julgamentos de gramaticalidade, testes de associação e completação) captam interpretações ou respostas que os sujeitos dão *a posteriori*, depois que o processamento de algum estímulo linguístico já teve lugar em suas mentes e já foi completamente finalizado, com a integração de todos os níveis linguísticos (fonético/fonológico, morfológico, sintático, semântico-pragmático). Desse modo, tais métodos, apesar de legítimos e correntes na pesquisa em psicolinguística experimental, são pouco informativos, dado que não são sensíveis aos processos mentais que ocorrem no curso do processamento linguístico, não sendo possível aferir nem o que aconteceu durante o processamento, nem o que foi mais fácil ou mais difícil de processar.

Além dessas limitações, a literatura psicolinguística fornece evidências (e.g., WHEELER, 1970) de que os processos mentais ocorrem, em geral, muito rapidamente para que possam ser reportados conscientemente pelos sujeitos, e muitos desses processos escapam completamente do palco da consciência. Por conseguinte, métodos que não são sensíveis ao processamento linguístico em curso não são capazes de responder a questões centrais a respeito da arquitetura dos processos mentais relacionados à linguagem, por exemplo: como se dá a interação entre os níveis linguísticos? Uma etapa termina antes da outra, havendo uma série fixa de eventos no processamento? Ou processos de diferentes níveis ocorrem simultaneamente?

Para ilustrar esse ponto, faremos referência a um debate clássico dos estudos sobre acesso lexical. Ele diz respeito à influência de informações contextuais no processamento de palavras ambíguas (polissêmicas ou homônimas). Quando um sujeito lê ou escuta uma palavra ambígua, qual é o papel do contexto linguístico e quando ele começa a desempenhar esse papel? Todos os sentidos da palavra ambígua seriam inicialmente ativados e aqueles inapropriados ao contexto seriam descartados depois de determinado tempo (SWINNEY, 1979), ou apenas o sentido favorecido pelo contexto seria acessado de imediato (TABOSSI, ZARDON, 1993)? Para que se possa defender um ou outro posicionamento, faz-se naturalmente necessário lançar mão de métodos experimentais que sejam capazes de rastrear o processamento linguístico passo

a passo, enquanto os sujeitos ainda estão engajados no processo de compreensão (ou produção) linguística. Esses métodos são chamados de métodos *on-line* de experimentação.

Métodos *on-line* englobam uma série de paradigmas (*naming, shadowing, decisão lexical, priming, leitura autocadenciada, rastreamento ocular, etc.*), a maioria dos quais se baseia em medidas de tempo de reação (também *reaction time*, ou RT).

Em termos simples, o tempo de reação, que é medido com grande precisão (na ordem de milésimos de segundo) pode ser definido como o intervalo entre a apresentação de um estímulo linguístico qualquer e a resposta subsequente do sujeito do experimento, havendo variações no tipo de estímulo (palavras isoladas, fragmentos de oração ou sintagmas, sentenças inteiras, textos completos) e no tipo de resposta dos sujeitos (leitura em voz alta, acionamento de uma tecla, etc.). Ainda mais importante, assume-se que tempos de reação baixos são indicativos de facilidade ou menor carga de processamento, enquanto tempos de reação elevados indicam dificuldade ou maior demanda cognitiva.

Um desses paradigmas experimentais baseados em tempo de reação mais largamente empregado na literatura é o de leitura autocadenciada (também *self-paced reading*, ou ainda leitura automonitorada). Esse paradigma consiste na medida dos tempos de reação de sujeitos na leitura de sentenças de um dado conjunto de textos, sendo essas sentenças segmentadas seja em palavras, seja em fragmentos maiores ou sintagmas, ou ainda, como foi o caso dos experimentos a serem reportados neste artigo, sem segmentação (i.e., sentença inteira).² Nas tarefas de leitura de sentenças sem segmentação, o tempo de reação corresponde, então, à medida do intervalo entre a apresentação de uma sentença na tela de um computador, sua leitura e o acionamento posterior de uma tecla. Quando essa tecla é pressionada, uma nova sentença substitui a anterior no mesmo local da tela, e o procedimento de leitura se repete, com o computador registrando os tempos de reação, até que os sujeitos completem a leitura de todas as sentenças de uma determinada passagem. Além disso, conforme indica o próprio nome do paradigma, o tempo de exposição de cada sentença é sempre controlado pelo próprio sujeito.

Cumpre observar que, na elaboração desse e de outros tipos de experimento, é sempre necessário definir, como já mencionado, quais são as variáveis dependentes, isto é, as variáveis mensuráveis, e quais são as variáveis independentes, ou seja, aquelas cuja manipulação gera impacto nas dependentes. Além disso, é também necessário

realizar diversos controles (e.g., tamanho das sentenças em termos de sílabas; frequência e concretude das palavras; classe gramatical; tempo e aspecto verbal; estrutura argumental; estrutura sintática; tipo de relação semântica, etc.), de modo a eliminar possíveis ruídos ou vieses, que podem influenciar negativamente o comportamento das variáveis dependentes e ameaçar a validade das generalizações feitas com base nos dados coletados.³ Em outras palavras, o pesquisador precisa se assegurar de que (i) o experimento capte as operações normais envolvidas no processamento linguístico e de que (ii) a única fonte de alterações nas variáveis dependentes sejam as variáveis independentes, e não fatores de outra natureza que não foram identificados de antemão e devidamente controlados.

Refeita essa importante consideração quanto à necessidade de um rigoroso controle dos estímulos utilizados em experimentos psicolinguísticos, é também digno de nota que eles, em especial os realizados no paradigma de leitura autocadenciada, têm sido amplamente empregados nos estudos sobre referenciação, tanto em português brasileiro (e.g., CORRÊA, 1998; CUNHA LIMA In: KOCH, MORATO, BENTES, 2005; LEITÃO, 2005), quanto em outras línguas, como o inglês (e.g., GORDON *et al.*, 1993; ALMOR, 1999), o chinês (e.g., YANG *et al.*, 1999) e o espanhol (e.g., GELORMINI-LEZAMA, 2008). A realização de experimentos nesse campo se vincula à tentativa de elucidação dos aspectos psicológicos relacionados à questão da referência, isto é, à investigação de como as pessoas identificam o referente de uma expressão nominal/pronominal qualquer e de como esse referente se mantém ou se desativa na memória dos falantes.

Para que fique mais claro o que é um experimento *on-line* e como esse tipo de experimento pode ser utilizado na investigação de aspectos cognitivos do texto, com ênfase nos processos mentais de resolução da referência, descreveremos, após uma sucinta revisão da literatura, nosso estudo em português brasileiro sobre processamento correferencial, que incorpora dois experimentos psicolinguísticos de leitura autocadenciada. Após essa descrição, voltaremos a ressaltar a necessidade de zelar por um rigoroso controle dos estímulos linguísticos utilizados nesses estudos experimentais, em especial no que concerne às diferenças entre as duas modalidades de uso da língua (fala e escrita),⁴ utilizando as recentes mudanças no sistema pronominal do português brasileiro como exemplo de tais diferenças. Com isso, esperamos contribuir para o avanço da discussão, em português brasileiro, sobre as questões metodológicas envolvidas na realização de experimentos psicolinguísticos.

1. O processamento correferencial de nomes repetidos, pronomes plenos e nulos

Os estudos sobre processamento correferencial abrangem uma ampla gama de problemas de diferentes naturezas, que têm sido pesquisados por diferentes métodos e abordagens. Uma dessas abordagens prioriza a investigação de projeções e princípios sintáticos que possam restringir a recuperação de antecedentes por expressões referenciais, linha à qual se vinculam diversos estudos, por exemplo, Corrêa (1993), Corrêa (1998), Costa (2010) e Lessa, Maia (2010), realizados tanto em português brasileiro (doravante PB), quanto em português europeu.

Outras abordagens voltam sua atenção para aspectos semânticos e discursivos, como a questão da escolha das formas referenciais (e.g., expressões nominais e pronominais). Esses problemas, contudo, têm sido, comparativamente, menos investigados em PB através do uso de técnicas experimentais, sendo Leitão (2005) um dos trabalhos pioneiros. E é justamente sobre esse problema da escolha de expressões referenciais que nosso estudo se debruçará.

Como se sabe, em um discurso, é possível fazer referência de diversas maneiras a entidades previamente mencionadas. É possível introduzir um referente utilizando, por exemplo, seu nome próprio, como em (1) e, posteriormente, retomar esse referente repetindo seu nome próprio, como em (2a), utilizando alguma outra expressão nominal, como em (2b), ou ainda fazendo uso de pronomes, tanto plenos, como em (2c), quanto nulos, como em (2d).

- (1) Obama anunciou o fim da guerra do Iraque.
- (2a) Obama declarou que os soldados serão retirados até o fim do ano.
- (2b) O presidente dos EUA declarou que os soldados serão retirados até o fim do ano.
- (2c) Ele declarou que os soldados serão retirados até o fim do ano.
- (2d) (pro) Declarou que os soldados serão retirados até o fim do ano.

Diante do fato de que os sistemas linguísticos permitem a existência de diferentes formas referenciais anafóricas, perguntas que naturalmente surgem são: o que leva os falantes a escolherem uma ou outra forma? Quais são os fatores que comandam essa escolha? Em determinados contextos linguísticos, quais

dessas formas são mais facilmente processadas pelos leitores/ouvintes? Essas questões têm sido objeto de intenso debate na literatura psicolinguística.

À primeira vista, conforme defendia Gernsbacher (1989), a mera repetição do nome seria, em qualquer contexto, a forma mais fácil de ser processada, já que contém maior carga de informações semânticas, capazes de ativar representações mentais previamente geradas, e refere-se sem ambiguidade ao seu antecedente, sem introduzir informação nova nem exigir processos inferenciais de interpretação.

Entretanto, Gordon *et al.* (1993) apresentaram fortes evidências experimentais de que, ao menos em inglês, nomes repetidos são mais difíceis de serem processados do que pronomes plenos, mas apenas em um contexto linguístico específico, a saber, quando o antecedente ocupa a posição de sujeito de sua sentença; essa preferência pelo pronome pleno desaparece quando o antecedente ocupa a posição de objeto. Essa maior dificuldade de processamento de nomes repetidos foi chamada pelos autores de *repeated-name penalty* (RNP, ou penalidade do nome repetido). Como é facilmente perceptível, tal penalidade relaciona forma referencial a saliência sintática do antecedente: é a posição de sujeito que coloca o referente em foco,⁵ tornando-o altamente saliente do ponto de vista cognitivo e penalizando sua retomada através de nomes repetidos, de modo que se pode postular que, quanto mais saliente for o antecedente, menos informativo precisa ser o termo anafórico (GUNDEL *et al.*, 1993).

Ainda a respeito da RNP, uma questão de pesquisa interessante é se esse efeito de processamento seria uma peculiaridade apenas do inglês ou se também se manifestaria em outras línguas. Yang *et al.* (1999), em uma adaptação dos experimentos de leitura autocadenciada conduzidos por Gordon *et al.* (1993), investigaram, em chinês, a carga de processamento de nomes repetidos, pronomes plenos e também de pronomes nulos em retomadas anafóricas. Curiosamente, encontraram o mesmo efeito de penalidade do nome repetido nessa língua. Mas, ao compararem os tempos de reação das sentenças contendo pronomes plenos e nulos, não encontraram diferenças estatisticamente significativas, o que os levou a postular que a escolha entre ambas as formas é estilisticamente motivada em chinês.

Tomados em conjunto, os resultados de Gordon *et al.* (1993) e Yang *et al.* (1999) indicaram que a RNP poderia ser um princípio de processamento universal, independentemente do parâmetro *pro-drop*, uma vez que foi verificada tanto em

línguas totalmente *pro-drop*, como o chinês, quanto em línguas não *pro-drop*, como o inglês. Por sua vez, o estudo das línguas românicas se mostra potencialmente esclarecedor para tais fenômenos sob investigação, devido à sua maior riqueza flexional, riqueza esta que pode ser mobilizada durante o processo de resolução da referência.

Gelormini-Lezama (2008) conduziu experimentos de leitura autocadenciada semelhantes aos dos trabalhos mencionados anteriormente para estudar, em espanhol, a eficiência de nomes repetidos, pronomes plenos e nulos em retomadas anafóricas com antecedentes sintaticamente salientes (i.e., em posição de sujeito) e não salientes (i.e., em posição de objeto). Os resultados, segundo análise do autor, apontaram a ocorrência de RNP nessa língua. Todavia mostraram, ao contrário dos dados provenientes do chinês, uma diferença significativa entre pronomes plenos e nulos: na retomada de antecedentes sintaticamente salientes, há, em espanhol, uma forte preferência pelo pronome nulo, cujo antecedente pode facilmente ser inferido pelo contexto e pelas marcas de concordância verbal, ao passo que o pronome pleno é processado mais lentamente. Em outras palavras, isso significa dizer que, em espanhol, passagens como (3) são processadas mais facilmente em relação a passagens como (4), que, por sua vez, são mais facilmente processadas em relação a (5), conforme exemplos extraídos de Gelormini-Lezama (2008, p. 17):

- (3) Juan se encontró con María. La vio triste.
- (4) Juan se encontró con María. Juan la vio triste.
- (5) Juan se encontró con María. Él la vio triste.

O autor chamou essa diferença entre pronomes plenos e nulos, com a penalização dos primeiros nos casos de retomada de antecedentes sintaticamente salientes, de *overt pronoun penalty* (OPP, ou penalidade do pronome pleno). Já na retomada de antecedentes não salientes, a diferença de comportamento entre nomes repetidos e pronomes plenos não foi estatisticamente significativa em espanhol, tendo sido os pronomes nulos as formas anafóricas mais penalizadas em comparação com as outras duas, elicitando os tempos de reação mais elevados entre todas as condições testadas.

Diante do que foi exposto, o objetivo geral dos dois experimentos que conduzimos e que serão reportados nas próximas seções foi investigar, através de uma replicação fiel do primeiro experimento de leitura autocadenciada reportado

em Gelormini-Lezama (2008), como nomes repetidos, pronomes plenos e nulos são processados em retomadas anafóricas com antecedentes salientes e não salientes em PB. E os objetivos específicos foram (i) verificar a existência ou não de RNP e OPP em PB e (ii) oferecer uma análise do fenômeno da correferência em PB que pudesse ser diretamente comparada aos trabalhos anteriormente mencionados sobre o inglês, o chinês e o espanhol.⁶

De modo a cumprir com os objetivos apresentados, replicamos, em colaboração com o autor, o primeiro experimento de leitura autocadenciada descrito em sua dissertação (GELORMINI-LEZAMA, 2008). Tal replicação se deu mediante tradução de todos os itens experimentais (do espanhol para o PB) empregados no seu teste de leitura autocadenciada, dos quais um exemplo pode ser conferido no quadro 1. Como se vê, esses estímulos experimentais consistiram em passagens, cada uma composta por duas sentenças que, por sua vez, podiam se realizar em seis diferentes condições experimentais, geradas pela associação de diferentes formas referenciais (nomes repetidos, pronomes plenos e nulos) a diferentes tipos de antecedente (sintaticamente salientes e não salientes).

QUADRO 1
Exemplo de estímulo experimental

Condições		Passagem Experimental	
Saliência do Antecedente	Forma Referencial	Sentença 1	Sentença 2
Sujeito	<i>Nome repetido</i>	Maria se encontrou com João.	Maria achou ele triste.
	<i>Pronome pleno</i>	Maria se encontrou com João.	Ela achou ele triste.
	<i>Pronome nulo</i>	Maria se encontrou com João.	Achou ele triste.
Objeto	<i>Nome repetido</i>	João se encontrou com Maria.	Maria achou ele triste.
	<i>Pronome pleno</i>	João se encontrou com Maria.	Ela achou ele triste.
	<i>Pronome nulo</i>	João se encontrou com Maria.	Achou ele triste.

Com relação às hipóteses de trabalho, é sabido que o PB tem passado por uma rápida simplificação em sua morfologia verbal, conforme tem sido vastamente atestado por diversos autores, entre eles Duarte (In: ROBERTS, KATO, 1996, p. 107):

De fato, a mudança que se observa no português do Brasil, que parece estar evoluindo de uma marcação positiva para uma marcação negativa dentro do parâmetro ‘pro-drop’, coincide com significativa redução ou simplificação nos paradigmas flexionais.

Tal simplificação estaria conduzindo a um uso cada vez mais frequente do sujeito explícito e ao virtual desaparecimento do sujeito nulo pessoal. Em decorrência disso, seria de se esperar que, em PB, diferentemente do que foi atestado em espanhol, sujeitos nulos em correferência com antecedentes salientes fossem as formas anafóricas mais penalizadas (i.e, as mais difíceis de serem processadas), dado que o antecedente não pode mais ser inferido pela concordância verbal. Também seria de se esperar que o processamento de nomes repetidos, supostamente regido por princípios universais, fosse penalizado no contexto em que seus antecedentes são sintaticamente salientes. Nesse mesmo contexto, se pensarmos em uma escala decrescente de facilidade de processamento, a hipótese seria a de que pronomes nulos seriam seguidos por nomes repetidos e, por fim, pronomes plenos, que seriam, então, a forma referencial mais favorecida do ponto de vista cognitivo. Em suma, esperávamos verificar, em PB, a existência de RNP, mas não de OPP.

Já para os casos de correferência com antecedentes não salientes, a expectativa seria de que não houvesse diferenças significativas entre nomes repetidos e pronomes plenos, já que essas diferenças não foram encontradas nem em inglês, nem em chinês, nem em espanhol. Esperava-se, porém, que, na retomada desse tipo de antecedente, à semelhança dos resultados encontrados em espanhol, pronomes nulos fossem altamente penalizados. Essa previsão encontra sustentação no quadro teórico proposto por Gundel *et al.* (1993): quanto menos saliente for o antecedente, mais informativo precisa ser o termo anafórico; se ele não for informativo o suficiente, como é o caso dos pronomes nulos, é possível prever que sua retomada seja penalizada.

2. Experimento 1

Participantes

Os participantes da pesquisa foram 45 voluntários falantes nativos de PB, entre os quais 30 do sexo feminino e 15 do sexo masculino. A idade média dos sujeitos foi de 25 anos, variando de 18 a 54 anos.

Estímulos

Procedeu-se à tradução das 36 passagens experimentais de Gelormini-Lezama (2008), como a que está exemplificada no quadro 1. Conforme mencionado anteriormente, cada passagem continha duas sentenças, que se submeteram aos seguintes controles:

- A primeira sentença introduzia, através de nomes próprios, dois indivíduos de gêneros distintos, dispostos em uma ordem SVO, de modo que uma entidade se realizava na posição de sujeito, enquanto a outra, na de objeto. A escolha de nomes de indivíduos de gêneros diferentes e da ordem SVO se deu pautada, respectivamente, na necessidade de eliminar qualquer influência de ambigüidade de gênero sobre o processo de resolução anafórica e no fato de a ordem SVO ser, atualmente, a ordem sintática não marcada em PB. Além disso, metade das passagens experimentais continha nomes próprios masculinos seguidos por femininos, enquanto, na outra metade, essa seqüência se invertia, ou seja, nomes femininos eram seguidos por masculinos.
- As funções sintáticas exercidas pelos referentes na sentença inicial eram ou mantidas ou alteradas na segunda sentença, a sentença crítica, da qual foram construídas seis diferentes versões, por meio de associações entre os níveis das duas variáveis independentes do experimento, a saber, forma referencial (nomes repetidos, pronomes plenos e nulos) e saliência do antecedente (sujeito ou objeto), o que gerou um desenho experimental 3 x 2.

Foi também traduzido o conjunto de 36 passagens distratoras (ou *fillers*) do primeiro experimento de Gelormini-Lezama (2008), com vistas a mascarar as passagens experimentais, disfarçando, para os sujeitos, qual o objetivo do teste. Assim como as passagens experimentais, elas também consistiam de duas sentenças contendo referentes introduzidos por nomes próprios ou outros tipos de descrições definidas e indefinidas, possuindo estruturas diversas, não diretamente relacionadas às manipulações experimentais.

Procedimento

O experimento foi rodado em um computador pessoal, em uma única sessão de aproximadamente 20 minutos de duração. Os sujeitos foram instruídos a ler cada sentença no ritmo mais natural possível, nem apressada, nem vagarosamente. Após lerem as instruções na tela, eles iniciavam uma sessão de treinamento com cinco passagens distratoras, para que pudessem se familiarizar com a tarefa de leitura autocadenciada, apreender a parte mecânica da tarefa antes do início do experimento propriamente dito e também tirar suas dúvidas.

Ao término da sessão de treinamento, os sujeitos eram avisados de que, ao pressionarem a barra de espaço, o experimento teria início. Antes da apresentação de cada passagem (tanto experimental quanto distratora), aparecia a seguinte sentença no centro da tela do computador: “Pressione a barra de espaço para ler a oração seguinte”. Ao pressionarem a barra de espaço, os sujeitos viam então a primeira sentença da passagem. Ao pressionarem novamente a barra de espaço quando do término da leitura, a sentença anterior desaparecia da tela e era substituída, no mesmo local, pela sentença crítica (sendo o tempo de reação dessa segunda sentença a variável dependente do experimento).

Ao final de cada passagem, isto é, após a leitura da sentença crítica, uma pergunta de compreensão aparecia no centro da tela. Os sujeitos respondiam a ela pressionando ou a barra de espaço, para dizerem “sim”, ou a tecla *shift* (direita ou esquerda, a depender de o sujeito ser destro ou canhoto), para responderem “não”. Essas perguntas de compreensão, incluídas também nos estímulos de treinamento, não exigiam conhecimentos dos nomes específicos dos indivíduos mencionados, mas sim das estruturas dos eventos descritos nas passagens. Além disso, elas foram elaboradas de modo que metade das respostas corretas fosse “sim” e a outra metade fosse “não”. Suas funções eram, por um lado, contribuir para disfarçar o objetivo do experimento e, por outro, garantir a atenção dos sujeitos durante toda a tarefa de leitura.

Foi utilizado o programa *E-prime*⁷ para (i) aleatorizar, diferentemente para cada sujeito, a ordem dos estímulos experimentais e distratores a ser apresentada na tela do computador, de modo a eliminar possíveis efeitos da ordem de apresentação dos estímulos nos resultados e (ii) gerar um delineamento experimental em quadrado latino, de modo que, na leitura de cada passagem experimental, fosse apresentada aos sujeitos apenas uma entre as seis possíveis versões da sentença crítica, com a ocorrência, no entanto, de cada uma das seis versões homoganeamente, considerando-se o conjunto de todos os sujeitos.

Resultados

Estão reproduzidas, na tabela 1, as médias dos tempos de reação (RTs) das sentenças críticas em todas as seis condições testadas.

TABELA 1
RTs médios (em ms) das sentenças críticas

Saliência do Antecedente	Forma Referencial	N	Média	Desvio Padrão	Erro Padrão
Sujeito	<i>Nome repetido</i>	276	2166	1007	61
	<i>Pronome pleno</i>	275	2448	1237	75
	<i>Pronome nulo</i>	276	2027	1132	68
Objeto	<i>Nome repetido</i>	276	2229	1203	72
	<i>Pronome pleno</i>	275	2701	1511	91
	<i>Pronome nulo</i>	275	2840	1612	97

Os dados foram submetidos a uma análise estatística de variância (ANOVA 3 x 2), na qual foi verificado se os fatores saliência do antecedente e forma referencial tiveram algum impacto sobre a variável dependente (RTs das sentenças críticas). A análise foi conduzida tanto por sujeitos (F1) quanto por itens experimentais (F2).

Verificou-se um efeito principal de saliência do antecedente tanto por sujeitos (F1 (1,45) = 25.4, $p < 0.001$) quanto por itens (F2 (1,35) = 25.8, $p < 0.001$), de modo que todas as expressões referenciais foram processadas mais lentamente quando associadas a um antecedente sintaticamente não saliente do que quando em situação de retomada de um antecedente saliente. Foi também encontrado um efeito significativo de forma referencial (F1 (2,90) = 11.4, $p < 0.001$; F2 (2,70) = 12.8, $p < 0.001$), com nomes repetidos, independentemente do *status* do antecedente, sendo processados mais rapidamente em relação a nulos, que, por sua vez, mostraram-se mais fáceis de serem processados do que pronomes plenos. A interação entre os dois fatores também se mostrou significativa (F1 (2,90) = 12.6, $p < 0.001$; F2 (2,70) = 12.7, $p < 0.001$).

Em seguida, conduzimos testes *post hoc* de comparações pareadas seguindo o método de normalização Bonferroni, com vistas a explorar mais detalhadamente as interações entre os níveis de saliência do antecedente (sujeito e objeto) e de forma referencial (nomes repetidos, pronomes plenos e nulos). Foram feitas três novas análises, comparando: (i) pronomes plenos e nulos; (ii) nomes repetidos e pronomes plenos; e (iii) nomes repetidos e nulos.

Em (i), encontrou-se um efeito principal de saliência do antecedente, com tempos de reação mais rápidos na retomada de sujeitos do que de objetos (F1 (1,45) = 3.2, $p < 0.001$; F2 (1,35) = 40.5, $p < 0.001$), mas não houve nenhum

efeito de forma referencial ($F1(1,45) = 3.2, p < 0.1$; $F2(1,35) = 2.7, p > 0.05$). Contudo, a interação entre os fatores se mostrou significativa ($F1(1,45) = 11.9, p < 0.01$; $F2(1,35) = 8.8, p < 0.01$).

Já em (ii), foi verificado, por um lado, um efeito marginal de saliência do antecedente ($F1(1,45) = 3, p < 0.1$; $F2(1,35) = 3.1, p < 0.1$), mas, por outro, um efeito significativo principal de forma referencial ($F1(1,45) = 22.3, p < 0.001$; $F2(1,35) = 35.5, p < 0.001$). Não houve interação entre os fatores, com $F_s < 2$.

Por fim, em (iii), foi encontrado um efeito principal de saliência do antecedente ($F1(1,45) = 30.9, p < 0.001$; $F2(1,35) = 26.8, p < 0.001$) e de forma referencial ($F1(1,45) = 8.3, p < 0.01$; $F2(1,35) = 10, p < 0.01$). A interação foi também significativa ($F1(1,45) = 19.3, p < 0.001$; $F2(1,35) = 39.3, p < 0.001$).

Discussão

Nas condições com antecedentes salientes, pronomes nulos foram as formas referenciais mais fáceis de serem processadas, gerando os menores tempos de reação entre todas as condições testadas. Nomes repetidos e pronomes plenos apresentaram comportamento semelhante entre si, assim como nomes repetidos e pronomes nulos, não tendo sido estatisticamente significativas as diferenças entre os tempos de reação das sentenças que os continham. Por sua vez, pronomes plenos foram penalizados em relação a nulos na retomada de antecedentes em posição de sujeito. Portanto, foi verificada a existência de OPP, mas não de RNP.

Já na retomada de antecedentes não salientes, os pronomes pleno e nulo mostraram-se, dessa vez, as formas anafóricas menos eficientes, gerando os tempos de reação mais elevados da condição; em termos simples, pronomes plenos e nulos foram igualmente penalizados, isto é, as diferenças entre seus tempos de reação não atingiram significância. Em contrapartida, nomes repetidos foram as formas referenciais mais fáceis de serem processadas quando em situação de retomada de antecedentes em posição de objeto, tanto em relação a pronomes nulos, quanto em relação a pronomes plenos.

Logo, os resultados obtidos não corroboraram inteiramente as hipóteses iniciais. Curiosamente, na retomada de antecedentes não salientes, foram verificadas diferenças significativas entre nomes repetidos e pronomes plenos, até o presente momento não encontradas em nenhuma das outras línguas que têm sido objetos de estudo em psicolinguística comparada. Ainda mais surpreendentemente, na retomada de antecedentes salientes, verificou-se que (i) a RNP não se estende ao PB,

e que (ii) sujeitos nulos foram as formas mais favorecidas processualmente, com pronomes plenos sendo severamente penalizados, dado que, por um lado, é altamente contraintuitivo e, por outro, constitui evidência contra uma mudança de marcação dentro do parâmetro *pro-drop* em PB (ao menos para os tipos de passagens estudadas neste experimento), a despeito do que vem sendo vastamente afirmado na literatura linguística.

Refletindo sobre os resultados, levantamos o seguinte questionamento: teria a utilização de pronomes tônicos em posição de objeto (rever quadro 1) sido capaz de enviesar os resultados, isto é, de interferir negativamente na aferição da variável dependente?

Inicialmente, a opção pela tradução dos itens experimentais do espanhol para o PB com o uso de pronomes tônicos (em vez de clíticos) em posição de objeto direto foi motivada pela sua maior frequência e naturalidade no vernáculo do PB atual, conforme vem sendo indicado por diversos autores, entre eles Tarallo (In: ROBERTS, KATO, 1996, p. 85):

Com referência aos pronomes clíticos, o sistema brasileiro já não os emprega há algum tempo. Na variedade portuguesa, entretanto, pronomes clíticos são frequentes no uso e ocorrem, conforme o esperado, em posição enclítica.

Uma vez que experimentos pretendem captar as operações naturais dos sujeitos durante o processamento linguístico, nada mais natural do que a recusa à elaboração de estímulos experimentais contendo clíticos. Em razão desse argumento, tem sido prática corrente na pesquisa em psicolinguística a confecção de estímulos experimentais contendo pronomes tônicos em posição de objeto (e.g., LEITÃO, 2005).

No entanto, diante dos surpreendentes resultados que obtivemos, procedemos a um levantamento de estudos sociolinguísticos sobre mudanças recentes no sistema pronominal do PB, através dos quais constatamos que elas não parecem ter atingido homoganeamente as duas modalidades de uso linguístico, isto é, fala e escrita.

Um desses estudos variacionistas é o de Corrêa (1991), que, em uma investigação dos fatores condicionantes do preenchimento/não preenchimento da posição de objeto direto em PB, acabou também por fornecer dados esclarecedores quanto ao uso de pronomes tônicos e clíticos de terceira pessoa

na função de objeto direto anafórico - que correspondem, coincidentemente, às características estruturais das sentenças críticas utilizadas em nosso experimento. A autora apresentou evidências de que a frequência de uso desses pronomes é afetada, por um lado, pelo nível de escolaridade dos falantes e, por outro, pelo tipo de modalidade de uso da língua (fala ou escrita) em questão, conforme mostram os dados, provenientes de um *corpus* de narrativas espontâneas orais e escritas, plotados nas tabelas 2 e 3, elaboradas por nós mediante um recorte e adaptação do material presente em Corrêa (1991, p. 56).

TABELA 2
Pronomes tônicos e clíticos anafóricos de 3ª pessoa em narrativas orais

Tipo de Objeto	Adultos analfab. (%)	Série (%)					Total (%)
		1ª/2ª	3ª/4ª	5ª/6ª	7ª/8ª	Univers.	
<i>Pron. Tônico</i>	100	100	100	86,7	95,5	40	92,5
<i>Clítico</i>	0	0	0	13,3	4,5	60	7,5

TABELA 3
Pronomes tônicos e clíticos anafóricos de 3ª pessoa em narrativas escritas

Tipo de Objeto	Série (%)					Total (%)
	1ª/2ª	3ª/4ª	5ª/6ª	7ª/8ª	Univers.	
<i>Pron. Tônico</i>	100	40	45,5	26,1	0	29,6
<i>Clítico</i>	0	60	54,5	73,9	100	70,4

A tabela 2 mostra que, nos textos orais analisados, os clíticos acusativos de terceira pessoa foram os menos utilizados como objetos diretos anafóricos (7,5% = 6/80 ocorrências), e sua ocorrência só se mostrou expressiva na fala de estudantes universitários. Note-se também que esses clíticos só aparecem a partir da quinta série na fala dos estudantes, passando, segundo Corrêa, por uma fase de hipercorreção antes de se firmarem; além disso, não se encontraram clíticos na fala de analfabetos. Em conjunto, esses dados são interessantes porque indicam que, em PB atual, os pronomes tônicos são adquiridos naturalmente pelas crianças, ao passo que o uso de clíticos só se dá a partir da ação normativa dos meios educacionais.

Já a tabela 3 revela que o uso do pronome tônico vai gradualmente desaparecendo nos textos escritos (29,6% = 21/71 ocorrências), à medida que os falantes tornam-se mais escolarizados. Em escala inversamente proporcional estão os clíticos, cuja frequência vai aumentando ao longo dos anos escolares, sendo máxima na escrita dos mais escolarizados.

Nos termos de Corrêa (1991, p. 80), os resultados por ela obtidos autorizam as seguintes conclusões:

Quanto ao clítico, seu contexto favorito é a língua culta escrita. Não sendo 'natural' entre os falantes mais jovens, tem de ser aprendido. (...) A relutância em se usar o clítico na fala é comparável a de se usar o pronome pleno na escrita culta. Entre os menos escolarizados, o clítico não ocorre e, na escrita culta, o pronome pleno não ocorre. A fala dos mais escolarizados admite, embora com restrição, o pronome lexical (...)

Considerando-se que experimentos de leitura autocadenciada lidam sempre com a modalidade escrita da língua e que a população de sujeitos de nosso experimento foi constituída por falantes universitários, há fortes razões para suspeitar de que o uso de pronomes tônicos em posição de objeto, dada sua pouca naturalidade na escrita, tenha afetado o processo de leitura das sentenças críticas, constituindo um viés para os nossos dados e, portanto, afetando negativamente o comportamento da variável dependente, fato que justificaria os resultados obtidos.

Diante disso, optamos por fazer uma nova rodada do mesmo experimento de leitura autocadenciada previamente reportado, com uma nova população de sujeitos, mas com a utilização de clíticos em lugar de pronomes tônicos nas sentenças críticas.

3. Experimento 2

Participantes

Os participantes da pesquisa foram 48 voluntários falantes nativos de PB, entre os quais 35 do sexo feminino e 13 do sexo masculino. A idade média dos sujeitos foi de 24 anos, variando de 16 a 50 anos.

Estímulos

Em comparação com a primeira versão do experimento, foram feitas alterações nos itens experimentais apenas quanto ao tipo de pronome utilizado em posição de objeto direto nas sentenças críticas. Essas alterações podem ser conferidas no quadro 2.

QUADRO 2
Novo exemplo de estímulo experimental

Condições		Passagem Experimental	
Saliência do Antecedente	Forma Referencial	Sentença 1	Sentença 2
Sujeito	<i>Nome repetido</i>	Maria se encontrou com João.	Maria o achou triste.
	<i>Pronome pleno</i>	Maria se encontrou com João.	Ela o achou triste.
	<i>Pronome nulo</i>	Maria se encontrou com João.	Achou-o triste.
Objeto	<i>Nome repetido</i>	João se encontrou com Maria.	Maria o achou triste.
	<i>Pronome pleno</i>	João se encontrou com Maria.	Ela o achou triste.
	<i>Pronome nulo</i>	João se encontrou com Maria.	Achou-o triste.

Procedimento

O experimento, com os 36 itens experimentais modificados (rever quadro 2) e os mesmos 36 distratores, sempre seguidos por perguntas de compreensão, foi rodado em um computador pessoal, em uma única sessão de aproximadamente 20 minutos de duração.

As instruções se mantiveram idênticas, tendo sido novamente incluída uma sessão de treinamento com cinco passagens distratoras. Assim como na primeira versão, o experimento, rodado com a utilização do programa *E-prime*, foi elaborado no mesmo paradigma experimental de leitura autocalendariada, com delineamento em quadrado latino e com o tempo de reação da segunda sentença como variável dependente.

Resultados

Na tabela 4, encontram-se reproduzidas as novas médias dos RTs das sentenças críticas em todas as seis condições testadas.

TABELA 4

RTs médios (em ms) das sentenças críticas com pronomes clíticos

Saliência do Antecedente	Forma Referencial	N	Média	Desvio Padrão	Erro Padrão
Sujeito	<i>Nome repetido</i>	288	1864	852	50
	<i>Pronome pleno</i>	288	2014	945	56
	<i>Pronome nulo</i>	288	1697	859	51
Objeto	<i>Nome repetido</i>	288	1977	1142	67
	<i>Pronome pleno</i>	288	2211	1140	67
	<i>Pronome nulo</i>	287	2269	1313	78

Uma análise estatística de variância (ANOVA 3 x 2), com os fatores saliência do antecedente e forma referencial, foi conduzida tanto por sujeitos (F1) quanto por itens experimentais (F2). Foi verificado um efeito principal de saliência do antecedente (F1 (1,35) = 38, $p < 0.001$; F2 (1,47) = 28, $p < 0.001$), de modo que todas as expressões referenciais associadas a antecedentes salientes foram processadas mais rapidamente em comparação com as expressões que se encontravam em correferência com antecedentes não salientes. Do mesmo modo, houve um efeito significativo de forma referencial (F1 (2,70) = 6.3, $p < 0.01$; F2 (2,94) = 5, $p < 0.01$), com nomes repetidos, independentemente do *status* do antecedente, sendo processados mais rapidamente do que pronomes nulos, que, por sua vez, foram mais rápidos em relação a pronomes plenos. A interação entre os dois fatores também se mostrou significativa (F1 (2,70) = 13.7, $p < 0.001$; F2 (2,94) = 11.5, $p < 0.001$).

Conduzimos também testes *post hoc* de comparações pareadas com a adoção do método de normalização Bonferroni, visando a constatar com maior grau de profundidade as interações entre os níveis de saliência do antecedente (sujeito e objeto) e forma referencial (nomes repetidos, pronomes plenos e nulos). Foram feitas três novas análises, comparando: (i) pronomes plenos e nulos; (ii) nomes repetidos e pronomes plenos; e (iii) nomes repetidos e pronomes nulos.

Na primeira nova análise, foi constatado um efeito principal de saliência do antecedente (F1 (1,35) = 43.4, $p < 0.001$; F2 (1,47) = 33.4, $p < 0.001$) e de forma referencial (F1 (1,35) = 6.1, $p < 0.05$; F2 (1,47) = 4.9, $p < 0.05$). A interação entre os dois fatores mostrou-se significativa (F1 (1,35) = 21, $p < 0.001$; F2 (1,47) = 14.9, $p < 0.001$).

Na segunda análise, à semelhança da primeira, foi também descoberto um efeito principal de saliência do antecedente ($F_1(1,35) = 8.3, p < 0.01$; $F_2(1,47) = 8.9, p < 0.01$) e de forma referencial ($F_1(1,35) = 14, p < 0.001$; $F_2(1,47) = 11.4, p < 0.01$). No entanto, nenhuma interação entre os fatores foi constatada, com todos os $F_s < 1$.

Por fim, a terceira análise revelou um efeito principal de saliência do antecedente ($F_1(1,35) = 43.1, p < 0.001$; $F_2(1,47) = 24.3, p < 0.001$), mas nenhum efeito de forma referencial ($F_1(1,35) = 1, p > 0.05$; $F_2(1,47) = 0.9, p > 0.05$). A interação entre os fatores foi, no entanto, significativa ($F_1(1,35) = 23.2, p < 0.001$; $F_2(1,47) = 17.2, p < 0.001$).

Discussão

Os resultados obtidos indicam que, nas condições com antecedentes salientes, não houve diferenças significativas entre os tempos de reação das sentenças que continham nomes repetidos e pronomes nulos, nem das sentenças contendo nomes repetidos e pronomes plenos; todavia, pronomes nulos foram processados de modo significativamente mais rápido em comparação com pronomes plenos. Logo, os dados novos continuaram revelando a ocorrência de OPP, mas nenhuma instância de RNP.

No entanto, na retomada de antecedentes não salientes, não houve diferenças significativas nos tempos com que pronomes plenos e nulos foram processados, mas estes foram significativamente mais lentos em relação a nomes repetidos, que, por sua vez, comportaram-se semelhantemente aos pronomes plenos, não tendo sido significativa a diferença entre seus tempos de reação.

Os resultados da segunda versão do experimento continuaram não corroborando inteiramente as hipóteses de trabalho. Em relação ao padrão de processamento verificado previamente, constatamos apenas uma diferença, todavia expressiva: na retomada de antecedentes sintaticamente não salientes, a diferença entre nomes repetidos e pronomes plenos encontrada na primeira versão do experimento desapareceu; ou seja, a segunda versão do experimento mostrou que, na retomada de antecedentes em posição de objeto, nomes repetidos e pronomes plenos não são processados de maneira significativamente distinta, dado que corrobora os achados e previsões de Gordon *et al.* (1993), Yang *et al.* (1999) e Gelormini-Lezama (2008).

Portanto, os dados novos revelaram que uma nova rodada do mesmo experimento, mas com a utilização de pronomes clíticos no lugar de tônicos, gerou, comparativamente com a versão anterior, um novo padrão de processamento para nomes repetidos e pronomes plenos em PB (nulos não foram aparentemente afetados).

Além disso, ao comparar as médias dos tempos de reação provenientes de ambas as versões do experimento, percebe-se facilmente que os tempos de reação do primeiro foram, de maneira bastante robusta, maiores que os do segundo (em média, 396,5 ms de diferença), em todas as seis condições testadas, conforme demonstrado na tabela 5.

TABELA 5
Comparação dos RTs médios (em ms) das sentenças críticas

Saliência do Antecedente	Experimento 1		Experimento 2	
	Forma Referencial	Média	Média	Diferença
Sujeito	<i>Nome repetido</i>	2166	1864	302
	<i>Pronome pleno</i>	2448	2014	434
	<i>Pronome nulo</i>	2027	1697	330
Objeto	<i>Nome repetido</i>	2229	1977	252
	<i>Pronome pleno</i>	2701	2211	490
	<i>Pronome nulo</i>	2840	2269	571

Como não houve nenhuma outra alteração nos estímulos entre as duas versões que não a presença de pronomes clíticos ou tônicos em posição de objeto, pode-se, com segurança, interpretar esse dado como uma forte evidência psicolinguística de que os pronomes tônicos, de fato, são naturais apenas na fala, sendo, na escrita, ainda processados com maior estranhamento ou dificuldade por falantes adultos escolarizados, conforme já evidenciado por Corrêa (1991) em estudo sociolinguístico.

4. Discussão geral

Os dois experimentos reportados são exemplos de experimentos *on-line* no paradigma de leitura autocadenciada. Testes psicolinguísticos como esses consistem em ferramentas legítimas e úteis para a investigação do processamento linguístico do texto em tempo “real”, incorporando aspectos como saliência

cognitiva, memória e capacidade de processamento aos modelos de cognição e processamento linguístico.⁸

Como se vê, nossos experimentos de leitura autocadenciada se vinculam, especificamente, à tentativa de investigação dos aspectos psicológicos do problema da referência, com foco na carga de processamento adquirida por algumas expressões referenciais (nomes repetidos, pronomes plenos e nulos) em função da saliência de seus antecedentes, de modo que foi possível a identificação de padrões de processamento distintos de tais expressões com base em nossos dois conjuntos de dados.

Conforme já apontado anteriormente, para que os dados provenientes dessa possibilidade metodológica de experimentação sejam ecologicamente válidos, é necessário que o pesquisador realize diversos tipos de controle, como uniformidade no tamanho dos itens experimentais, tipo de verbo, tempo verbal, coerência de sentido entre as sentenças, entre muitos outros, que variam a depender da natureza dos problemas investigados. Ou seja, de modo a isolar o efeito das variáveis independentes sobre as dependentes, o pesquisador precisa naturalmente eliminar possíveis efeitos indesejados de outros fatores sobre essas variáveis, mantendo-os constantes ao longo de todas as condições experimentais. E esse é o ponto mais relevante para o argumento que tentamos desenvolver aqui, porque é justamente para a discussão sobre assuntos de natureza metodológica como esse que esta pesquisa pode contribuir.

Como subproduto de nossos estudos sobre processamento correferencial em PB, foram encontradas corroborações experimentais para as evidências empíricas já levantadas na literatura linguística (CORRÊA, 1991) no que diz respeito às mudanças no sistema pronominal do PB.

Ao se compararem os dados provenientes do experimento 1 com os do experimento 2, que divergiram apenas quanto à presença de pronomes tônicos ou clíticos em posição de objeto direto, respectivamente, percebemos que a introdução dos clíticos nos itens experimentais teve ao menos duas consequências importantes: por um lado, as diferenças no modo como nomes repetidos e pronomes plenos foram processados quando em retomada de antecedentes sintaticamente não salientes desapareceram (rever tabela 4); e, por outro, os tempos de reação das sentenças contendo clíticos foram significativamente menores do que os das sentenças contendo pronomes tônicos em posição de objeto (rever tabela 5).

Por conseguinte, nossos resultados confirmam a naturalidade dos pronomes tônicos apenas no vernáculo; na língua culta escrita, clíticos ainda

parecem ser a forma pronominal processada mais naturalmente por falantes com alto grau de escolarização. Além disso, como o padrão de processamento de pronomes plenos, em relação a nomes repetidos, alterou-se em função da forma pronominal utilizada, encontramos evidência de que esse fator é robusto o suficiente para enviesar os resultados experimentais, necessitando, portanto, ser devidamente identificado e controlado de antemão pelo experimentador.

Essas descobertas têm implicações importantes para a comunidade psicolinguística brasileira, na qual tem sido prática corrente a elaboração de estímulos experimentais contendo pronomes tônicos em posição de objeto. Tal prática encontrava motivação no postulado (e.g., TARALLO In: ROBERTS, KATO, 1996) de que os pronomes clíticos não eram mais frequentes e naturais em PB; no entanto, evidências sociolinguísticas (e.g., CORRÊA, 1991) e as conclusões deste estudo psicolinguístico apontam para a necessidade de reformulação dessa visão.

Nossa pesquisa também lança luzes sobre o problema da generalização, para a modalidade de fala, dos dados obtidos em experimentos de leitura autocadenciada, que lidam, obviamente, com a modalidade escrita da língua. É claro que generalizações são frequentes e importantes para que seja possível construir modelos abstratos do processamento. Todavia, nossos resultados, apesar de preliminares, já indicam que estruturas de língua escrita e de língua falada não parecem ser processadas da mesma maneira pelos leitores, e que a presença de estruturas típicas da fala na escrita gera certo estranhamento ou dificuldade durante o processo de compreensão linguística, fato que se encontra refletido nas diferenças dos tempos de reação entre ambas as versões do experimento conduzido. Esse tema aguarda, contudo, novas pesquisas.

5. Considerações finais

Nossos resultados apontam, em suma, para a necessidade de controle também das diferenças entre fala e escrita durante a elaboração de experimentos de leitura autocadenciada, especialmente no tocante à escolha das formas pronominais quando da confecção de itens experimentais. Do contrário, a validade dos dados e das generalizações feitas com base neles pode ficar seriamente comprometida.

Assim, esperamos, com esta pesquisa e com as análises aqui empreendidas, ter contribuído para o avanço, em PB, das discussões sobre importantes questões metodológicas envolvidas nos estudos psicolinguísticos do processamento que se dão mediante experimentação.

Notas

* Graduado em Letras e mestre em Linguística Teórica e Descritiva pela Universidade Federal de Minas Gerais (UFMG). Doutorando em Linguística em University of South Carolina (USC).

¹ Para um panorama dos paradigmas experimentais disponíveis em psicolinguística, incluindo os que envolvem medidas fisiológicas, ver Starr, Rayner (In: NADEL, 2005) e Mitchell (In: CARREIRAS, CLIFTON, 2004).

² Existem, conforme indicado, diferentes versões da tarefa de leitura autocadenciada. Para uma revisão histórica, com discussões sobre as vantagens e limitações desse paradigma experimental, ver Mitchell (In: CARREIRAS, CLIFTON, 2004).

³ A especificação das variáveis dependentes e independentes de um experimento, bem como dos ruídos ou vieses, constitui o que na literatura se denomina pelo termo “delineamento (ou planejamento) experimental”. Um delineamento experimental também engloba um plano para se atribuírem sujeitos a condições experimentais e identifica, de antemão, a análise estatística associada a esse plano. Textos de referência sobre o tema podem ser buscados em Kirk (1995).

⁴ Para fins de simplificação, ao mencionarmos “fala” e “escrita”, estaremos lidando com os protótipos dessas modalidades. Mas que fique claro para o leitor que defendemos, na esteira de Marcuschi (2001), a perspectiva de que as diferenças entre ambas as modalidades se situam não em relação dicotômica, mas sim dentro de um contínuo de práticas de produção textual.

⁵ Cumpre observar que os fatores determinantes da saliência que um antecedente adquire constituem uma questão ainda não totalmente respondida na literatura. Há fortes evidências, porém, de que eles não se limitam apenas à posição sintática ocupada pelo antecedente. Fatores de natureza semântica (KEHLER *et al.*, 2007), por exemplo, também geram forte impacto no *status* cognitivo da informação dada, podendo ser até mais fortes que os fatores de natureza sintática.

⁶ Leitão (2005) e Queiroz e Leitão (2008) também apresentam estudos sobre o processamento correferencial de nomes repetidos e pronomes plenos em PB, mas acreditamos que seus experimentos não possam ser diretamente comparados aos de Gordon *et al.* (1993), Yang *et al.* (1999) e Gelormini-Lezama (2008), devido às características dos estímulos linguísticos utilizados, cuja explicitação ultrapassa, contudo, o escopo deste artigo.

⁷ Maiores informações sobre o programa estão disponíveis em: <<http://www.pstnet.com/eprime.cfm>> (Acesso em: 23 mar. 2012).

⁸ Além de permitirem a incorporação de uma perspectiva processual à investigação linguística, métodos experimentais como os aqui apresentados podem funcionar também como um exercício de reflexão sobre os modelos, possibilitando a identificação de suas limitações e podendo até mesmo indicar a necessidade de sua reformulação (CUNHA LIMA, In: KOCH, MORATO, BENTES, 2005, p. 214).

Referências

- ALMOR, A. Noun-phrase anaphora and focus: the informational load hypothesis. *Psychological Review*, v. 106, n. 4, p. 748-765, 1999.
- CORRÊA, L. M. S. Restrições ao pronome livre na linearização do discurso. *Revista paLavra*, v. 1, p. 75-95, 1993.
- CORRÊA, L. M. S. Acessibilidade, paralelismo na interpretação do pronome sujeito e o contraste pro/pronome em português. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, v. 24, n. 2, p. 295-392, 1998.
- CORRÊA, V. R. *O objeto direto nulo no português do Brasil*. 1991. 90f. Dissertação (Mestrado em Linguística) – Departamento de Linguística, Universidade Estadual de Campinas. Campinas, 1991.
- COSTA, M. A. Referential chains' processing. In: FRANÇA, A. I.; MAIA, M. A. R. (Eds.). *Papers in psycholinguistics: proceedings of the First International Psycholinguistics Congress*. Rio de Janeiro: Imprinta, 2010. p. 291-310.
- CUNHA LIMA, M. L. Referenciação e investigação do processamento cognitivo: o exemplo do indefinido anafórico. In: KOCH, I. V.; MORATO, E. M.; BENTES, A. C. (Orgs.). *Referenciação e Discurso*. São Paulo: Contexto, 2005. p. 197-218.
- DUARTE, M. E. L. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: ROBERTS, I.; KATO, M. A. (Orgs.). *Português Brasileiro: uma viagem diacrônica*. Campinas: Editora da UNICAMP, 1996. p. 107-125.
- GELORMINI-LEZAMA, C. *Processing repeated names, overt pronouns and null reference in Spanish*. 2008. 39f. Dissertação (Mestrado em Linguística) – College of Arts and Sciences, University of South Carolina. Columbia, 2008.
- GERNSBACHER, M. A. Mechanisms that improve referential access. *Cognition*, 32, p. 99-156, 1989.
- GORDON, P. C. *et al.* Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, p. 311-347, 1993.
- GUNDEL, J. K. *et al.* Cognitive status and the form of referring expressions in discourse. *Language*, v. 69, n. 2, p. 274-307, 1993.

KEHLER, A. *et al.* Coherence and correferrence revisited. *Journal of Semantics*, v. 25, p. 1-44, 2007.

KIRK, R. E. *Experimental design: procedures for the behavioral sciences*. 3. ed. Pacific Grove, CA: Brooks/Cole, 1995.

LEITÃO, M. M. Processamento correferencial de nomes e pronomes em português brasileiro. *Revista Linguística (PPGL/UFRJ)*, v. 1, n. 2, p. 235-258, 2005.

LESSA, P. C.; MAIA, M. A. R. The processing of R-expressions with lexical restrictions in Brazilian Portuguese. In: FRANÇA, A. I.; MAIA, M. A. R. (Eds.). *Papers in psycholinguistics: proceedings of the First International Psycholinguistics Congress*. Rio de Janeiro: Imprinta, 2010. p. 317-321.

MARCUSCHI, L. A. Oralidade e letramento. In: _____. *Da fala para a escrita: atividades de retextualização*. São Paulo: Cortez, 2001. p. 15-43.

MITCHELL, D. C. On-line methods in language processing: introduction and historical review. In: CARREIRAS, M.; CLIFTON, C. E. (Ed.). *The on-line study of sentence comprehension: eyetracking, ERPs and beyond*. Nova Iorque: Psychology Press, 2004. p. 15-32.

QUEIROZ, K. L.; LEITÃO, M. M. Processamento do sujeito anafórico em português brasileiro. *Veredas – Revista de Estudos Linguísticos Online*, v. 2, p. 163-166, 2008.

STARR, M. S.; RAYNER, K. Methodologies for studying language comprehension. In: NADEL, L. (Ed). *Encyclopedia of cognitive science*. Nova Iorque: Wiley, 2005. p. 1385-1391.

SWINNEY, D. A. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, p. 645-659, 1979.

TABOSSI, P.; ZARDON, F. Processing ambiguous words in context. *Journal of Memory and Language*, 32, p. 359-372, 1993.

TARALLO, Fernando. Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar ao final do século XIX. In: ROBERTS, Ian; KATO, Mary Aizawa. (Orgs.). *Português Brasileiro: uma viagem diacrônica*. 2. ed. Campinas: Editora da UNICAMP, 1996. p. 35-68.

WHEELER, D. D. Processes in word recognition. *Cognitive Psychology*, v. 1, n. 1, p. 59-85, 1970.

YANG, C. L. *et al.* Comprehension of referring expressions in Chinese. *Language and Cognitive Processes*, v. 14, n. 5/6, p. 715-743, 1999.

Submissão do artigo: 25/03/2011

Aprovação do artigo: 13/05/2012