

## **PROCESSAMENTO E INTERPRETAÇÃO DE SUJEITOS NULOS E PLENOS EM PORTUGUÊS EUROPEU E EM PORTUGUÊS DO BRASIL**

*Paula Luegi  
Armanda Costa  
Marcus Maia*

### RESUMO

Neste estudo, analisámos o processamento e a interpretação de sujeitos nulos e plenos em PE e em PB na retoma de antecedentes que ocupam diferentes posições estruturais. Resultados: função sintática e posição estrutural influenciam a resolução da correferência em PE; em PB, o nulo é mais facilmente interpretado como correferente com o Sujeito ou com o antecedente mais próximo que o c-comanda.

**PALAVRAS-CHAVE:** processamento de frases; interpretação e processamento de sujeitos nulos e plenos; Português Europeu e Português do Brasil.

### Introdução

**A**s relações correferenciais, ou, mais especificamente, o modo como se estabelece a relação entre uma expressão anafórica e o seu antecedente tem sido analisado de diferentes perspetivas ligadas ao estudo da linguagem, entre elas, a da linguística teórica, a da linguística computacional e a da psicolinguística. No âmbito da psicolinguística, a área de estudo em que se insere este trabalho, tem sido central tentar compreender como as expressões anafóricas são processadas e interpretadas, sendo para isso necessário identificar que fontes de informação linguística estão envolvidas no seu processamento.

O primeiro objetivo deste trabalho<sup>1</sup> é o de tentar justamente identificar e distinguir o efeito de diferentes fontes de informação linguística, nomeadamente, o efeito da função sintática e da posição estrutural dos antecedentes no processamento e interpretação de sujeitos nulos e pronominais em condições intrafrásicas em PE e em PB. O segundo objetivo é o de contrastar as estratégias de processamento utilizadas por falantes nativos de PE e de PB na interpretação de sujeitos nulos e plenos.

Este artigo tem a seguinte estrutura: na primeira secção, apresentamos as principais propostas psicolinguísticas desenvolvidas para dar conta do processamento e interpretação de expressões anafóricas e descrevemos as razões por que considerámos pertinente contrastar PE e PB, identificando alguns dos aspetos que distinguem as duas variedades. Na segunda secção, apresentamos as duas experiências desenvolvidas em que contrastamos PE e PB, apresentando no final as conclusões que consideramos poder extrair dos resultados obtidos.

## Propostas psicolinguísticas

Uma das propostas mais comumente aceite, e que tem servido de base para outras propostas mais específicas, é a Teoria da Acessibilidade, desenvolvida por Mira Ariel (1985, 1990 e seguintes; cf. ARIEL, 2001). Esta teoria descreve como a linguagem humana, especificamente, como os mecanismos cognitivos envolvidos na resolução da correferência são reativos a fatores da memória. O que Ariel propõe é que cada expressão anafórica codifica um grau específico e diferente de acessibilidade mental da entidade que retoma e que as expressões anafóricas são na verdade marcadores de acessibilidade: quanto mais acessível está a informação (o antecedente) menor a probabilidade de ser retomada por uma expressão altamente informativa. Assim, as expressões anafóricas mais informativas recuperam antecedentes pouco acessíveis e, ao contrário, as formas menos informativas e mais reduzidas retomam antecedentes muito acessíveis no discurso. Prevê-se que, em línguas em que tanto formas nulas como formas plenas sejam possíveis, a forma nula, mais reduzida e menos informativa, seja usada para retomar entidades muito salientes e a forma plena para retomar entidades menos salientes do que as retomadas pela forma nula.

<sup>1</sup> O texto aqui apresentado é parte da tese de doutoramento da primeira autora, desenvolvida sob orientação dos coautores deste estudo.

Com base nesta ideia da relação inversa entre a informatividade da expressão anafórica e a saliência do antecedente, várias propostas têm surgido tentando dar conta do que define a proeminência de uma entidade ao nível da frase. É nesse contexto que surgem propostas como a Teoria da Centralidade, de Grosz, Joshi e Weinstein (1995), ou a Hipótese da Posição do Antecedente, de Carminati (2002), que preveem que, ao nível da frase, a relação entre a expressão anafórica e o antecedente se estabeleça, sobretudo, se não exclusivamente, com base em informações sintáticas (ou estruturais), especificamente, na função sintática do antecedente.

A Hipótese da Posição do Antecedente, desenvolvida por Carminati (2002), dá conta do processamento de pronomes em Italiano<sup>2</sup>, especificamente ao nível intrafrásico e destaca o papel da função sintática do antecedente. Esta proposta prediz que, em condições intrafrásicas, em Italiano, os pronomes nulos retomam preferencialmente antecedentes em Spec IP, enquanto os pronomes plenos retomam antecedentes que não se encontrem em Spec IP.

Carminati (2002) aproxima-se da Teoria da Acessibilidade assumindo que existe uma relação entre a forma da expressão anafórica e a acessibilidade do antecedente, mas considera que, em contextos mais locais, e sobretudo ao nível da frase, a proeminência é estabelecida com base em fatores exclusivamente sintáticos. Carminati (2002) defende que, como os autores da Teoria da Centralidade, se assumirmos a interpretação reducionista que às vezes é feita desta teoria, que em situações intrafrásicas, na resolução de pronomes, o processador é sensível meramente a fatores sintáticos e que, conseqüentemente, a entidade mais proeminente é o sujeito em Spec IP. Assim, o processador, ao encontrar um pronome nulo, inicia uma procura pelo antecedente em Spec IP, fazendo o inverso quando encontra um pronome pleno.

A Hipótese da Posição do Antecedente acaba por ser uma formalização como proposta teórica de resultados que já haviam sido encontrados em outros estudos, inclusivamente para o Português, tais como os trabalhos de Costa, Faria e Matos (1998) ou Costa (2003/2005), para o PE, ou de Corrêa (1998), para o PB. No entanto, depois do trabalho de Carminati (2002), proliferaram os estudos com o objetivo de testar especificamente a Hipótese

---

<sup>2</sup> A proposta de Carminati (2002) é desenvolvida para o Italiano e, apesar de prever um comportamento semelhante para as restantes línguas pro-Drop, assume que possa haver diferenças entre os comportamentos das diferentes expressões anafóricas em cada língua.

da Posição do Antecedente (para uma revisão de alguns destes trabalhos, veja-se Luegi (2012)). De um modo geral, estes trabalhos assumem que a forma nula é preferencialmente interpretada como correferente com o Sujeito e que a forma plena é mais variável.

No entanto, tanto no trabalho de Carminati (2002) como em muitos outros estudos que demonstraram a estreita relação entre o Sujeito e a forma nula, o Sujeito era sempre a primeira entidade a ser referida e, consequentemente, a entidade estruturalmente mais saliente. Assim, o Sujeito combina em si várias informações que o promovem a entidade mais saliente. Contudo, existem trabalhos que demonstram que, quando outras variáveis são manipuladas, tais como a informação de género (ARNOLD, EISENBAND, BROWN-SCHMIDT e TRUESWELL, 2000), as propriedades semânticas do verbo (COSTA, 2003/2005; COSTA, FARIA e KAIL, 2004), a função sintática e a posição estrutural dos constituintes (JÄRVIKIVI, VAN-GOMPEL, HYÖNÄ e BERTRAM, 2005; MAYOL, 2010), o estatuto discursivo dos antecedentes (KAISER, 2006; FRANA, 2007), a ordem dos constituintes e a relação de coerência entre as orações (FILIACI, 2010; MAYOL e CLARK, 2010) ou o papel temático dos argumentos verbais (MORGADO, 2011), a relação entre formas nulas (ou as formas mais reduzidas possíveis na língua em análise) e o sujeito e formas menos reduzidas e os complementos do verbo não é tão estável como proposto pela Hipótese da Posição do Antecedente. Na verdade, estes estudos demonstram que a informação relativa à função sintática não é a única informação que intervém no estabelecimento de relações de correferência e que a escolha de um antecedente para uma expressão anafórica resulta da combinação de diferentes fatores.

Estes resultados, e, sobretudo, os resultados de Kaiser e colaboradores (2006, 2008, entre outros) levam ao desenvolvimento de propostas alternativas que assumem perspectivas multifatoriais, como a proposta de Kaiser e Trueswell (2008), que prevê uma análise mais restritiva, que os autores definem como *form-specific multiple-constraints approach*, considerando que diferentes expressões anafóricas são sensíveis a diferentes fatores (informações linguísticas) e que a proeminência resulta da combinação de diferentes fontes ou níveis de informação.

Kaiser (2006) assume que a saliência não é um conceito monolítico, mas antes que resulta da combinação de diferentes informações. Kaiser e

Trueswell (2008) propõem mesmo um modelo que designam de *form-specific multiple-constraints approach*, em que preveem não só que a saliência seja o resultado da competição entre os diferentes fatores em jogo como também que as diferentes formas anafóricas sejam sensíveis a diferentes níveis de informação<sup>3</sup>.

### **Omissão de sujeitos em dados de *corpora*: contraste entre pe e pb**

Com base em evidências de que o PB estaria a passar por mudanças profundas com uma redução da omissão de Sujeitos, Duarte (1995) desenvolve um trabalho em que analisa e compara dados de *corpora* orais (do PE e do PB) e escritos (do PB). Conclui que, na oralidade, em PE, o Sujeito nulo é a opção preferida em todas as pessoas do discurso e que, em PB, se regista uma perda gradual da omissão de Sujeito, justificada pela redução dos paradigmas flexionais (a este respeito veja Figueiredo Silva (1996)). A maior resistência à mudança verifica-se na 3.<sup>a</sup> pessoa, na qual se regista ainda uma ocorrência de Sujeitos nulos acima dos 50%.

Da análise dos dados de escrita do PB, Duarte (1995) conclui que o PB se está a tornar uma língua não pro-drop em que a forma nula já não se encontra em distribuição complementar com a forma plena, mas antes que são intercambiáveis.

Barbosa, Duarte e Kato (2005), com base nos resultados de Duarte (1993, 1995) e de Barbosa (2000) (cf. BARBOSA, DUARTE e KATO, 2005), comparam a distribuição das formas pronominais Sujeito em PE e em PB, focando-se apenas na 3.<sup>a</sup> pessoa de modo a verificar se as alterações registadas na oralidade se confirmam na escrita.

Os resultados de Barbosa, Duarte e Kato (2005) indicam um maior uso de formas plenas em PB (56%) do que em PE (22%). Enquanto em PE o fator que mais influencia a escolha de uma ou de outra forma é a função sintática do antecedente, em PB a escolha é influenciada pela distância entre retoma e antecedente. Em ambas as variedades, no entanto, são importantes as relações de c-comando. A existência de relações de c-comando entre retoma e antecedente favorece a ocorrência de um Sujeito nulo.

---

<sup>3</sup> Veja-se Mayol (2010) e Morgado (2011) para uma proposta semelhante relativamente à sensibilidade das diferentes formas anafóricas.

Em termos concretos, este estudo mostra que, como em Duarte (1995), a forma nula é mais frequente em PE do que em PB e que há uma preferência por retomar o Sujeito da oração subordinante com a forma nula, tanto em PE (97%) como em PB (78%)<sup>4</sup>, e também que a forma plena é mais usada para retomar antecedentes não Sujeito em PB (57%) do que em PE (33%).

## Trabalho experimental

Neste estudo, com o objetivo de identificar o efeito da função sintática e da posição estrutural dos antecedentes no processamento e interpretação de Sujeitos pronominais nulos e plenos, analisamos estruturas complexas com ordem subordinante-subordinada. Na primeira oração são sempre introduzidas duas entidades com diferentes funções sintáticas, um Sujeito e um Oblíquo, que podem estar na sua ordem canónica,  $SVO_{bi}$ , em que o Sujeito está na posição estrutural mais destacada (é primeira referência), ou não canónica,  $O_{bi}VS$ , em que o Sujeito está numa posição estrutural mais baixa (é segunda referência). A subordinada é sempre uma oração adverbial temporal, introduzida pelo conector ‘quando’, em que o Sujeito pode ser nulo ou pleno (pronomes pessoais na 3.<sup>a</sup> pessoa do singular), à exceção da primeira experiência, em que apenas testámos condições com Sujeito nulo.

O objetivo deste trabalho é o de contribuir para a identificação dos fatores linguísticos envolvidos no processamento e interpretação de expressões anafóricas, em particular, e, conseqüentemente, encontrar indícios sobre os fatores envolvidos no processamento da linguagem, em geral. Pretendemos ainda, com este trabalho, identificar se as estratégias de processamento e as preferências de interpretação de formas nulas e plenas são semelhantes em PE e em PB.

Como o contraste entre PE e PB foi sempre realizado com base em estudos de dados de *corpora*, considerámos necessário comparar PE e PB ao nível do processamento e da interpretação, ou seja, da compreensão (por oposição aos dados de produção obtidos com *corpora*). A razão para essa comparação é a existência de propostas aparentemente contraditórias. Se por um lado existem estudos de *corpora*, tais como os de Duarte (1995) ou de Barbosa, Duarte e

<sup>4</sup> Note-se, no entanto, que há uma diferença de quase 20% entre as duas variedades.

Kato (2005), que verificam que em PB tem havido um decréscimo do uso de formas nulas e um aumento do uso de formas plenas comparativamente com PE, por outro existem vários estudos de processamento de frases que assumem uma distribuição complementar das duas formas, prevendo que formas nulas e plenas se comportam do mesmo modo em PE e em PB.

Apesar de haver alguns estudos sobre a resolução e interpretação de expressões referenciais tanto em PE como em PB, nenhum desses trabalhos foi construído com o propósito de comparar as duas variedades. Assim, apesar de os resultados encontrados indicarem que existem semelhanças entre o PE e o PB, os dados não podem efetivamente ser alvo de comparação.

## **Experiência 1**

### Método

#### Desenho experimental

Na Experiência 1 manipularam-se duas variáveis a dois níveis (2X2): função do antecedente retomado, Sujeito (S) ou Oblíquo (O), e ordem de referência, primeira referência (Ref1), em que a entidade retomada é o constituinte estruturalmente mais destacado, e segunda referência (Ref2), em que a entidade retomada é o constituinte estruturalmente mais encaixado. A combinação dos diferentes níveis de cada uma das variáveis independentes gerou quatro diferentes condições experimentais: S\_Ref1, O\_Ref1, S\_Ref2 e O\_Ref2. Foram criadas quatro listas experimentais, contendo cada uma delas 6 frases de cada condição, três com Sujeito do género masculino e com Oblíquo do género feminino e três com a relação inversa (Sujeito-feminino e Oblíquo-masculino) num total de 24 frases experimentais.

A retoma de Sujeito ou de Oblíquo foi forçada pela flexão em género do participio da oração subordinada ('foi convidado/a'). Nesta experiência, a condição pronome pleno não foi manipulada, tendo sido apresentadas apenas frases com a forma nula.

Como variáveis dependentes, mediram-se os tempos de leitura da região crítica, a sublinhado no exemplo (1), e o número de respostas certas e erradas à tarefa secundária que consistia na avaliação de veracidade de uma afirmação sobre a frase lida (exemplificada em (2)).

(1) A Sara escreveu para o Emanuel em Maio quando foi jubilada pela faculdade no ano passado.

(2) A Sara foi jubilada pela faculdade.

### Participantes

Participaram nesta experiência 40 falantes nativos de PE, estudantes de graduação e pós-graduação da Faculdade de Letras da Universidade de Lisboa, e 40 falantes nativos de PB, estudantes de graduação e pós-graduação da Faculdade de Letras da Universidade Federal do Rio de Janeiro.

### Materiais e aparato experimental

Cada frase foi construída de modo a que pudesse ser apresentada nas quatro condições descritas anteriormente. Exemplificamos de (3) a (6) cada uma das condições resultante do desdobramento de uma mesma frase.

(3) A Sara escreveu para o Emanuel em Maio quando foi jubilada pela faculdade no ano passado. (S\_Ref1)

(4) Para o Emanuel escreveu a Sara em Maio quando foi jubilado pela faculdade no ano passado. (O\_Ref1)

(5) Para o Emanuel escreveu a Sara em Maio quando foi jubilada pela faculdade no ano passado. (S\_Ref2)

(6) A Sara escreveu para o Emanuel em Maio quando foi jubilado pela faculdade no ano passado. (O\_Ref2)

Uma condição de cada frase foi associada a uma das quatro listas criadas, numa distribuição em Quadrado Latino, e a cada participante foi atribuída uma lista, lendo este todas as condições testadas, mas apenas um exemplar de cada frase em cada condição. As 24 frases experimentais de cada lista foram misturadas com 48 frases distratoras e apresentadas de modo aleatório a cada participante.

Para a apresentação dos estímulos foi usado um computador Macintosh PowerBook G4 de 17 polegadas. A apresentação dos estímulos foi feita no programa PsyScope (COHEN, MACWHINNEY, FLATT e PROVOST, 1993) que registou não só o tempo de leitura dos segmentos críticos como também a resposta dada pelos participantes na tarefa secundária.

As frases foram apresentadas por segmentos (assinalados pelas barras oblíquas no exemplo (7)), no centro do ecrã, a Chicago (a fonte definida por defeito pelo programa), tamanho 24, de modo não cumulativo, ou seja, cada segmento substituiu o anterior exatamente na mesma posição do ecrã.

(7) /A Sara / escreveu / para o Emanuel / em Maio / quando / foi jubilada / pela faculdade / no ano passado.

No final da frase, aparecia uma afirmação, como exemplificado em (8), que permanecia no ecrã durante dois segundos.

(8) A Sara foi jubilada pela faculdade.

A afirmação era então substituída por três pontos de interrogação que permaneciam no ecrã até que o participante avaliasse a afirmação como verdadeira, carregando na tecla 'S', assinalada a verde, ou falsa, carregando na tecla 'L', assinalada a vermelho.

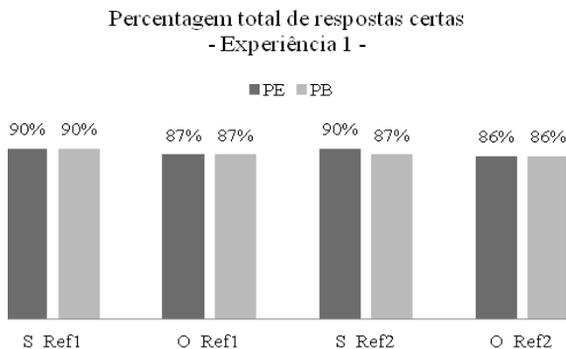
### Procedimento

No início de cada sessão individual, e durante uma conversa informal, foi explicado a cada participante o que teria de fazer durante a experiência. Era explicado que seriam apresentadas, no ecrã do computador, frases, divididas em pequenos fragmentos, e que para passarem de um excerto para outro teriam de carregar na barra de espaço. Antes do final de cada frase, aparecia no centro do ecrã um '+' que permanecia no ecrã até os participantes carregarem na barra de espaço e, no final da leitura de cada frase, aparecia uma afirmação sobre a frase lida que tinham de avaliar como verdadeira ou falsa.

Antes do início da tarefa, realizava-se um pequeno bloco de treino com seis itens (frase, pergunta, hipóteses de resposta).

### Resultados

Apresentamos no Gráfico 1 as percentagens de respostas certas em cada uma das condições analisadas. Os resultados são semelhantes não só entre PE e PB, mas também entre as diferentes condições. Os testes estatísticos realizados não revelaram diferenças estatisticamente significativas nem entre condições nem entre grupos.



**Gráfico 1 - Percentagem total de respostas certas da Experiência 1. Tarefa de julgamento de veracidade.**

Relativamente aos tempos de reação, no contraste entre as duas variedades, apesar de existir uma diferença entre os valores de PE e de PB, como se pode ver no , essa diferença não é estatisticamente significativa. De resto, nenhum dos efeitos se revelou significativo.

Em PE, como no contraste entre as duas variedades, também não se registaram diferenças estatisticamente significativas de nenhuma das variáveis manipuladas.

Em PB, os testes estatísticos revelaram um efeito da função sintática do antecedente ( $b= 160.4$ ;  $t$ -valor= $2.02$ )<sup>5</sup>, com tempos mais altos na retoma de Oblíquo do que na retoma de Sujeito, mas nenhum efeito da ordem de referência ou da interação entre os dois fatores.

**Tabela 1 - Tempos de reação das diferentes condições nas duas variedades.**

	PE	PB
<b>S_Ref1</b>	987 ms	1123 ms
<b>S_Ref2</b>	1014 ms	1201 ms
<b>O_Ref1</b>	1011 ms	1281 ms
<b>O_Ref2</b>	1073 ms	1364 ms

<sup>5</sup> Utilizamos, neste estudo, Mixed Effect Models, no software R (<http://www.r-project.org/>) e com o pacote lme4 (Bates, 2010). Douglas M. Bates. *lme4: Mixed-effects modeling with R*. June 25, 2010. A versão mais recente do modelo não disponibiliza, para variáveis não categoriais, valores de p, pelo que assumimos aqui, de acordo com a literatura, como significativo, valores de t superiores a 2.00.

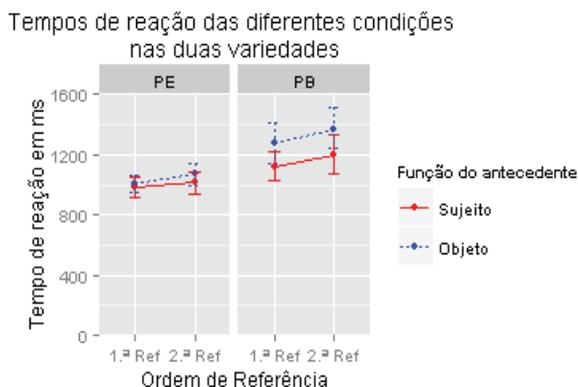


Gráfico 2 – Tempos de reação das diferentes condições nas duas variedades.

## Discussão dos resultados

Nesta experiência, em que analisámos unicamente condições com a forma nula, registámos diferenças estatisticamente significativas apenas em PB.

Em PB, os tempos são sempre mais altos na retoma de Oblíquo, independentemente de este ocupar a posição estrutural mais alta (primeira referência) ou mais baixa (segunda referência), uma vez que não se registou um efeito da ordem de referência. Este resultado pode revelar um efeito claro da função sintática, ou seja, uma preferência por interpretar as formas nulas como retomando antecedentes com a função de Sujeito, mas também pode refletir um efeito das relações de c-comando.

Barbosa, Duarte e Kato (2005) referem que os nulos retomam, preferencialmente, consituientes que os c-comandem ou mais especificamente, de acordo com Guessier (2008), os constituintes que imediatamente os c-comandem. Estes resultados, em que o Oblíquo não é aceite como um bom antecedente do constituinte nulo, podem demonstrar que em PB há uma preferência por selecionar como antecedente o constituinte que c-comande o Sujeito nulo e, quando mais do que um constituinte preenche este requisito, há uma preferência por selecionar o constituinte que o c-comande imediatamente. Nas estruturas por nós testadas, quando o Oblíquo é segunda referência, ou seja, quando está na posição mais baixa da estrutura, não pode c-comandar o cons-

tituinte nulo e, quando está na posição mais alta, apesar de c-comandar o Sujeito nulo, não é o constituinte mais próximo que o c-comanda.

Assim, em PB, a preferência de retoma de Sujeito com forma nula pode não ser apenas um efeito da função sintática de Sujeito, mas das relações estruturais de c-comando. Estes resultados estão de acordo tanto com as propostas da proeminência da função sintática, como da influência das relações de c-comando, sendo este efeito apenas evidente em PB.

De resto, no contraste entre as duas variedades, não se registam resultados estatisticamente significativos, apesar de, aparentemente, haver uma diferença de tempos entre as duas variedades.

A ausência de resultados em PE (e também no contraste entre as duas variedades) pode ter sido resultado da falta de efeito das manipulações introduzidas ou consequência do tipo de estrutura selecionada. Não nos parece que a falta de resultados estatisticamente significativos se possa justificar pelo fraco efeito das manipulações introduzidas até porque outros trabalhos, anteriores ao nosso, como por exemplo os trabalhos de Carminati (2002) ou Filiaci (2010), registaram também uma ausência de efeitos estatísticos em experiências em que foram usadas estruturas sintáticas semelhantes (para uma discussão mais extensa sobre este assunto, leia-se Filiaci (2010) e Luegi (2012)), mas em que se manipularam outras variáveis independentes. Assim, assumimos que a ausência de resultados se justifica pelo tipo de estrutura testada e não pela ausência de efeito das manipulações introduzidas.

## Experiência 2

### Método

#### Desenho experimental

Na Experiência 2 manipularam-se duas variáveis a dois níveis (2X2): tipo de pronome, Nulo (N) ou Pleno (P), e ordem de referência, Sujeito primeira referência (Sref1) – estruturas SVO – e Sujeito segunda referência (Sref2) – estruturas OVS. A combinação dos diferentes níveis de cada uma das variáveis independentes gerou quatro diferentes condições experimentais: Sref1\_N, Sref1\_P, Sref2\_N e Sref2\_P. Foram criadas quatro listas experimentais, contendo cada uma delas 5 frases de cada condição, num total de 20 frases experimentais.

A variável dependente medida foi a percentagem de retomas de Sujeito ou de Oblíquo em cada condição.

### Participantes

Participaram nesta experiência 24 falantes nativos de PE, estudantes de graduação e pós-graduação da Faculdade de Letras da Universidade de Lisboa, e 24 falantes nativos de PB, estudantes de graduação e pós-graduação da Faculdade de Letras da Universidade Federal do Rio de Janeiro.

### Materiais e aparato

Cada frase foi construída de modo a que pudesse ser apresentada nas quatro condições descritas anteriormente. Exemplificamos, de (9) a (12), cada uma das condições resultante do desdobramento de uma mesma frase.

(9) O bombeiro perguntou pelo militar no quartel quando recebeu a medalha de condecoração. (Sref1\_N)

(10) O bombeiro perguntou pelo militar no quartel quando ele recebeu a medalha de condecoração. (Sref1\_P)

(11) Pelo militar perguntou o bombeiro no quartel quando recebeu a medalha de condecoração. (Sref2\_N)

(12) Pelo militar perguntou o bombeiro no quartel quando ele recebeu a medalha de condecoração. (Sref2\_P)

As frases foram distribuídas por diferentes listas em Quadrado Latino, tendo, no total, resultado quatro diferentes listas experimentais, contendo, cada uma, 20 frases experimentais e 40 frases distratoras.

Foram construídos cadernos com as frases, as perguntas e as hipóteses de resposta, apresentadas sempre nesta sequência – frase, pergunta e respostas (como apresentado no exemplo (13)). Cada página continha quatro frases (mais a pergunta e as hipóteses de resposta): uma ou duas experimentais, no máximo, e duas ou três distratoras. As frases foram distribuídas aleatoriamente pelas páginas, mas intercalando-se sempre experimentais e distratoras de modo a que as experimentais nunca aparecessem seguidas, nem na mesma página nem na passagem de uma página para a seguinte.

(13) O bombeiro perguntou pelo militar no quartel quando recebeu a medalha de condecoração.

Quem recebeu a medalha de condecoração?

(a) o bombeiro                      (b) o militar

#### Procedimento

As instruções foram primeiro explicadas oralmente, indicando aos participantes que deveriam selecionar a resposta que achassem mais adequada, mesmo que mais do que uma hipótese fosse possível, e posteriormente lidas por cada participante no caderno que lhes foi distribuído.

#### Resultados

Na análise estatística realizada, no contraste entre as duas variedades, regista-se um efeito marginal de ordem de referência ( $b=-0.556$ ;  $z\text{-value}=-1.94$ ;  $p=0.053$ ), um efeito que não é particularmente relevante<sup>6</sup>, um efeito significativo de tipo de pronome ( $b=-2.037$ ;  $z\text{-value}=-6.78$ ;  $p<0.001$ ), com mais retomas de Sujeito com nulo do que com pleno, e um efeito da interação entre os dois fatores ( $b=1.212$ ;  $z\text{-value}=4.09$ ;  $p<0.001$ ), com mais retomas de Sujeito com nulo na condição em que o Sujeito é a primeira referência, ou seja, em que está numa posição estrutural destacada, e mais retomas de Oblíquo com a forma plena na condição em que o Oblíquo é segunda referência (posição pós-verbal). Não há no entanto nenhum efeito da variedade, o que fica claro pela observação do Gráfico 3.

Em PE registou-se um efeito de ordem de referência<sup>7</sup> ( $b=-0.996$ ;  $z\text{-value}=-2.03$ ;  $p=0.04271$ ), um efeito do tipo de pronome ( $b=-3.218$ ;  $z\text{-value}=-5.87$ ;  $p<0.001$ ), com mais retomas de Sujeito com a forma nula, e um efeito da interação entre estes dois fatores ( $b=2.173$ ;  $z\text{-value}=4.60$ ;  $p<0.001$ ), com mais retomas de Sujeito com forma nula quando o Sujeito é a primeira referência e mais retomas de

Oblíquo segunda referência com pronome pleno.

<sup>6</sup> Não faz sentido considerar um efeito de ordem porque as construções Sujeito primeira referência tanto têm formas nulas como formas plenas. A preferência por Sujeito é maior nas condições Sref2 porque nas condições Sref1 há um grande contraste entre a condição com nulo (em que há uma forte preferência por Sujeito) e a condição com pleno (em que há uma forte preferência por Oblíquo). Interessa-nos, sobretudo, o efeito da função do antecedente e o efeito de interação entre função e ordem de referência.

<sup>7</sup> Não relevante (ver nota anterior).

Tabela 2 - Percentagem de retomas de Sujeito nas diferentes condições nas duas variedades.

	PE	PB
<b>Sref1_N</b>	75%	68%
<b>Sref1_P</b>	29%	41%
<b>Sref2_N</b>	65%	62%
<b>Sref2_P</b>	46%	49%

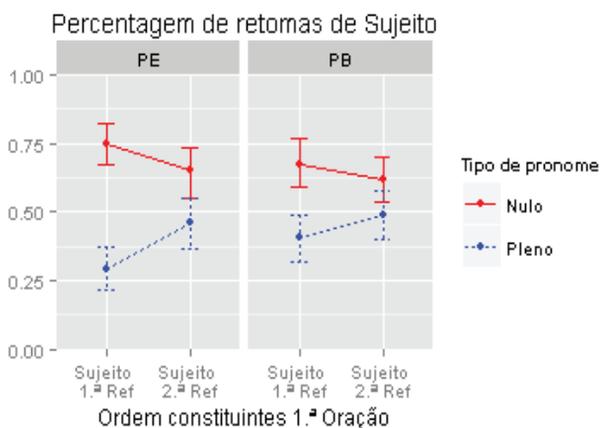


Gráfico 3 – Percentagem de retomas de Sujeito nas diferentes condições nas duas variedades.

Em PB, há um efeito do tipo de pronome, ou seja, uma preferência por retomar antecedentes Sujeito com o constituinte nulo, o que se mostra estatisticamente significativo ( $b=-0.954$ ;  $z\text{-value}=-3.72$ ;  $p=0.0002$ ).

#### Discussão dos resultados

Os nossos resultados revelam efeitos da ordem de referência, do tipo de pronome e da interação entre os dois fatores tanto no contraste entre as variedades como em PE. No entanto, os resultados do contraste entre as variedades parecem ser influenciados, sobretudo, pelos dados do PE, já que em PB não se registam efeitos nem de ordem nem de interação. Assim, os resultados do PE

parecem evidenciar que a posição estrutural dos constituintes também é um fator a ter em conta na atribuição de saliência a uma entidade discursiva, para além da função sintática, e que este fator atenua o efeito da função sintática. O Sujeito perde saliência quando não é a entidade estruturalmente mais destacada e, por esse motivo, passa não só a ser um antecedente menos acessível para o pronome nulo (com um decréscimo de 10% da condição Sref1\_N para a Sref2\_N) como também a ser um antecedente mais disponível para o pronome pleno (com uma subida de 17% da condição Sref1\_P para a condição Sref2\_P). A relação nulo-Sujeito é de facto mais forte quando o Sujeito é também a entidade estruturalmente mais saliente, ou seja, quando estes dois níveis de informação se combinam. Contudo, quando há uma competição entre eles, ou seja, quando o Sujeito não é a entidade estruturalmente mais destacada, a proeminência do Sujeito atenua-se. Este resultado parece estar de acordo com as propostas de Kaiser e colaboradores (2006, 2008, entre outros) e contra as propostas que preveem a função sintática como o fator único na atribuição da saliência a uma entidade.

O efeito da ordem de referência, ou da posição estrutural dos constituintes, parece não se fazer sentir em PB, onde o único resultado que se encontra é um efeito do tipo de pronome. Os resultados indicam que há uma preferência por retomar Sujeito com a forma nula e Oblíquo com a forma plena, como em PE, mas que, ao contrário de PE, essa preferência não se altera em função da ordem de referência dos constituintes. Estes resultados são, de resto, semelhantes aos resultados da experiência anterior e parecem indicar que o PB é mais sensível ao efeito da função sintática ou a relações de c-comando, e que, por essa razão, o Sujeito nulo tende a retomar um antecedente que o c-comande (BARBOSA, DUARTE e KATO, 2005), de preferência o antecedente mais próximo (GUESSER, 2008), ou seja, nas estruturas testadas neste estudo, o Sujeito.

De um modo geral, com base na observação dos dados, PE e PB têm resultados muito semelhantes. Em ambas as variedades, a maior diferença entre nulo e pleno regista-se nas condições com ordem canónica dos constituintes da primeira oração. O nulo é preferencialmente ligado ao Sujeito e o pleno ao Oblíquo. No entanto, a diferença entre nulo e pleno é menor em PB do que em PE. Em PB, não só a retoma de Sujeito com nulo tem sempre uma percentagem menor como também a retoma de Sujeito com pleno tem uma

percentagem maior do que em PE. No entanto, não se registou nenhuma diferença entre as duas variedades, o que não seria de esperar, tendo em consideração os trabalhos de Duarte (1995) e de Barbosa, Duarte e Kato (2005). De acordo com estes trabalhos, seria de esperar uma diferença entre PE e PB quanto ao tipo de pronome.

Duarte (1995) refere que o decréscimo do uso da forma nula em PB se atenua na 3.<sup>a</sup> pessoa (o foco da nossa análise), mas Barbosa, Duarte e Kato (2005) encontram diferenças entre o uso da forma nula na 3.<sup>a</sup> pessoa no contraste entre PE e PB. Assim, seria de esperar um efeito do tipo de pronome no contraste entre as duas variedades.

Na mesma linha, não seria de esperar um efeito do tipo de pronome em PB, uma vez que a proposta destes trabalhos é a de que as duas formas, nulas e plenas, se estejam a aproximar e não estejam mais em distribuição complementar. De facto, se analisarmos os resultados das condições com ordem canónica (Sujeito primeira referência, as estruturas mais frequentemente estudadas, e menos marcadas em ambas as variedades), verificamos que existe uma diferença entre PE e PB nas condições com forma plena, com maior percentagem de retomas de Sujeito com pleno em PB (41%) do que em PE (29%). No entanto, na análise estatística, estas condições não são contrastadas diretamente e este efeito pode ter ficado escondido pela manipulação da ordem dos constituintes. Assim, seria conveniente realizar um estudo comparativo entre as duas variedades em que se manipulasse apenas o tipo de pronome e a função do antecedente.

Em síntese, estes resultados parecem indicar que PE e PB são sensíveis a diferentes fatores no estabelecimento de relações de correferência. Em PE, tanto os efeitos da posição estrutural dos constituintes como os efeitos da função sintática são relevantes, enquanto em PB é relevante, sobretudo, a função sintática ou, aparentemente, as relações de c-comando.

## **Conclusões**

Das duas experiências realizadas, apenas a segunda nos permite extrair conclusões sobre as diferentes manipulações introduzidas, uma vez que na primeira experiência não encontramos resultados estatisticamente significativos, à exceção dos resultados de PB.

Relativamente à ausência de resultados da primeira experiência, e dado que esta ausência de resultados já se registou em diversos outros estudos, consideramos conveniente de futuro fazer uma análise detalhada dos aspetos que podem justificar tais resultados. Assim, julgamos necessário realizar novos estudos, focados apenas na questão do tipo e na ordem das orações, de modo a identificar as razões para os resultados encontrados. Para já, remetemos para a leitura dos trabalhos de Filiaci (2010) e de Luegi (2012), em que se tentam avançar algumas possíveis explicações.

Quanto aos resultados da segunda experiência, realçamos que, apesar de não se terem registado diferenças entre as duas variedades, na análise de cada variedade em separado, registamos efeitos diferentes para PE e para PB. Assim, parece que PE e PB são sensíveis a diferentes tipos de informação. Enquanto em PE se fizeram sentir tanto os efeitos da posição estrutural como os efeitos do tipo de pronome, em PB apenas se registaram diferenças quanto ao tipo de pronome.

Com base nos resultados, podemos concluir que, em PE, os dois fatores manipulados, ou seja, função sintática e posição estrutural dos antecedentes, assim como a forma da expressão anafórica que os retoma, são usados na resolução de expressões anafóricas, mais especificamente, de Sujeitos nulos e plenos. Não há, assim, com base nos nossos resultados, uma informação linguística única responsável pela atribuição de saliência a um antecedente: a saliência é o resultado da combinação de diferentes fatores que intervêm durante o processamento. Os resultados permitem-nos propor a seguinte hipótese, na linha dos trabalhos de Kaiser e colaboradores (2006, 2008, entre outros):

*Em condições intrafrásicas, em PE, com dois potenciais antecedentes para a expressão anafórica, a proeminência resulta da combinação de fatores como a função sintática e a posição estrutural de cada entidade.*

*Assim, estabelece-se uma hierarquia entre as entidades que combinam ambos os fatores de forma positiva, as que os opõem e as que os combinam de forma negativa, preferindo o pronome nulo retomar o antecedente Sujeito na posição mais alta da estrutura e o pleno o antecedente não Sujeito que esteja numa posição pós-verbal.*

Concordamos assim com Kaiser (2006) que considera que a saliência não é um conceito monolítico e que afirma que pelo menos na resolução de cadeias referenciais o processador se guia e faz uso das diferentes informações disponíveis.

Relativamente aos dados do PB, e tendo por base os resultados das duas experiências, este estudo, apesar de ser limitado, parece indicar que há evidências da relação estreita entre o Sujeito nulo e o antecedente (mais próximo) que o c-comanda. Assim, os resultados relativamente a PB parecem evidenciar que, em PB, o processamento e interpretação da correferência são mais sensíveis à função sintática e às relações de c-comando do que à posição estrutural do antecedente. Relativamente à perda de produtividade do Princípio Evitar Pronome em PB, os nossos dados, em consequência do desenho experimental desenvolvido, não nos permitem extrair conclusões, apesar de haver alguns indicadores da diferença entre PE e PB no que diz respeito às preferências de interpretação da forma plena.

Com base neste trabalho, que consideramos incipiente, consideramos que os estudos de processamento de frases devem ser desenvolvidos em paralelo com as análises teóricas que se têm até ao momento baseado, sobretudo, se não exclusivamente, em juízos intuitivos e em dados de análise de *corpora*. Consideramos, de resto, imprescindível a continuação deste estudo de modo a poder verificar alguns aspetos que não ficaram suficientemente claros. Nesse sentido, estamos no momento a desenvolver novos estudos com esse propósito.

## Agradecimentos

Os autores deste artigo agradecem à Fundação para a Ciência e a Tecnologia o financiamento concedido através das bolsas de doutoramento (SFRH/BD/3574/2007) e pós-doutoramento (SFRH/BPD/84138/2012) concedidas à primeira autora, e através do projeto *Compreensão na Leitura. Processamento de palavras, frases e textos* (PTDC/LIN/67854/2006) e do projeto plurianual do Centro de Linguística da Universidade de Lisboa – CLUL (PEst-OE/LIN/UI0214/2011). Agradecem ainda ao CNPq e à CAPES o apoio disponibilizado para a apresentação deste trabalho no 3.º *Workshop em Processamento Anafórico* e a todos os voluntários que participaram nas experiências deste estudo.

## Referências

ARIEL, Mira. Accessibility theory: An overview. In: T. Sanders, J. Schliperoord, & W. Spooren, *Text representation*. John Benjamins (Human cognitive processing series), 2001, pp. 29-87.

ARNOLD, Jennifer, EISENBAND, Janet, BROWN-SCHMIDT, Sarah, & TRUESWELL, John. The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76, 2000, B13-B26.

BARBOSA, Pilar, DUARTE, Eugênia, & KATO, Mary. Null subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics*, 4:2, 2005, pp. 11-52.

CARMINATI, Maria. *The processing of Italian subject pronouns*. Electronic Doctoral Dissertations for UMass Amherst, 2002.

CORRÊA, Letícia. Acessibilidade, paralelismo na interpretação do pronome sujeito e o contraste pro/pronome em português. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 24, 1998, pp. 295-392.

COSTA, Armanda. *Processamento de Frases em Português Europeu: Aspectos Cognitivos e Linguísticos Implicados na Compreensão da Língua Escrita*. Lisboa: Fundação Calouste Gulbenkian, 2003.

\_\_\_\_\_, FARIA, Isabel Hub, & MATOS, Gabriela. Ambiguidade referencial na identificação do Sujeito em estruturas coordenadas. *Textos seleccionados do XIII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Colibri, 1998, pp. 173-188.

\_\_\_\_\_, \_\_\_\_\_, & KAIL, Michèle. Semantic and Syntactic Cues' Interaction on Pronoun Resolution in European Portuguese. In: BRANCO, MCENERY, & MITKOV (Ed.), *DAARC 2004, 5th Discourse Anaphora Resolution Colloquium*. Lisboa: Colibri, 2004, pp. 45-50.

DUARTE, Eugênia. *A perda do princípio "evite pronome" no português brasileiro*. Tese de Doutorado, Universidade Estadual de Campinas, 1995.

FIGUEIREDO SILVA, Maria Cristina. *A posição sujeito em português brasileiro: em frases finitas e infinitivas*. Campinas, São Paulo: Ed. da UNICAMP, 1996.

FILIACI, Francesca. *Anaphoric Preferences of Null and Overt Subjects in Italian and Spanish: a Cross-linguistic Comparison*. PhD dissertation, The University of Edinburgh, 2010.

FRANA, Ilaria. *The Role of discourse prominence in the resolution of referential ambiguities. Evidence from co-reference in Italian*. *UMOP* 37, 2007.

GROSZ, Barbara, JOSHI, Aravind, & WEINSTEIN, Scott. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21, 1995, pp. 203-225.

GUESSER, Simone. Embedded Null Subjects in Brazilian Portuguese. *Nanzan Linguistics: Special Issue*, 5, 2008, pp. 55-78.

JÄRVIKIVI, Juhani, VAN GOMPEL, Roger, HYÖNÄ, Jukka, & BERTRAM, Raymond. Ambiguous Pronoun Resolution: Contrasting the First-Mention and Subject-Preference Accounts. *Psychological Science*, 16, 2005, pp. 260-264.

KAISER, Elsi. Effects of topic and focus on salience. In: C. Ebert, & C. En-driss (Ed.), *Proceedings of Sinn und Bedeutung 10, ZAS Working Papers in*, Berlin, 2006, pp. 139-154.

\_\_\_\_\_, & TRUESWELL, John. Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, 23(5), 2008, pp. 709-748.

LUEGI, Paula. *Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes*. Lisboa: Tese de Doutorado apresentada à Universidade de Lisboa, 2012.

MAYOL, Laia. Refining salience and the Position of Antecedent Hypothesis: a study of Catalan pronouns. *University of Pennsylvania Working Papers in Linguistics*, 16, Artigo 15, 2010.

\_\_\_\_\_, & CLARK, R. Pronouns in Catalan: Games of Partial Information and the Use of Linguistic Resources. *Journal of Pragmatics*, 42, 2010, pp. 781-799.

MORGADO, Sara. *Processamento da co-referência pronominal: informação sintática e semântica*. Tese de Mestrado apresentada à Faculdade de Letras da Universidade de Lisboa, 2011.

## **PROCESSING AND INTERPRETATION OF NULL AND OVERT SUBJECTS IN EUROPEAN AND BRAZILIAN PORTUGUESE**

### **ABSTRACT**

In this study we analyzed the processing and interpretation of covert and overt subjects in European and Brazilian Portuguese (EP and BP) when referring to antecedents with different structural position and syntactic functions. Results: both structural position and syntactic function influence coreference resolution in EP; in BP null subjects are more easily interpreted when referring to the closest c-commanding antecedent.

**KEYWORDS:** sentence processing; anaphoric chain interpretation; null subjects; EP and BP.

Recebido em 25/03/14

Aprovado em 09/07/14