



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE HUMANIDADES**  
**DEPARTAMENTO DE LETRAS VERNÁCULAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

**ALISSON HUDSON VERAS LIMA**

**A DISTRIBUIÇÃO DOS NOMES REPETIDOS, DOS NOMES  
RECATEGORIZADOS, DOS PRONOMES PLENOS E DOS PRONOMES  
NULOS EM FUNÇÃO ANAFÓRICA EM PORTUGUÊS BRASILEIRO**

**FORTALEZA**

**2020**

ALISSON HUDSON VERAS LIMA

**A DISTRIBUIÇÃO DOS NOMES REPETIDOS, DOS NOMES  
RECATEGORIZADOS, DOS PRONOMES PLENOS E DOS PRONOMES  
NULOS EM FUNÇÃO ANAFÓRICA EM PORTUGUÊS BRASILEIRO**

Tese apresentada ao Programa de Pós-graduação em Linguística da Universidade Federal do Ceará, como requisito parcial para a obtenção do título de Doutor em Linguística.

Aluno: Alisson Hudson Veras Lima  
Orientadora: Profa Dra Maria Elias Soares

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- L696d Lima, Alisson Hudson Veras.  
A distribuição dos nomes repetidos, dos nomes recategorizados, dos pronomes plenos e dos pronomes nulos em função anafórica em Português Brasileiro / Alisson Hudson Veras Lima. – 2020.  
303 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Linguística, Fortaleza, 2020.  
Orientação: Profa. Dra. Maria Elias Soares.  
Coorientação: Profa. Dra. Elisângela Nogueira Teixeira.
1. Correferência. 2. Psicolinguística. 3. Descrição linguística. I. Título.

CDD 410

---

ALISSON HUDSON VERAS LIMA

A DISTRIBUIÇÃO DOS NOMES REPETIDOS, DOS NOMES  
RECATEGORIZADOS, DOS PRONOMES PLENOS E DOS PRONOMES  
NULOS EM FUNÇÃO ANAFÓRICA EM PORTUGUÊS BRASILEIRO

Tese apresentada ao Programa de Pós-  
graduação em Linguística da Universidade  
Federal do Ceará, como requisito parcial  
para a obtenção do título de Doutor em  
Linguística.

Aprovada em: \_\_\_\_/\_\_\_\_/\_\_\_\_

BANCA EXAMINADORA

---

Profa. Dra. Maria Elias Soares (Orientadora)

Universidade Federal do Ceará

---

Prof. Dr. Márcio Martins Leitão

Universidade Federal da Paraíba

---

Profa. Dra. Paula Lenz Costa Lima

Universidade Estadual do Ceará

---

Profa. Dra. Vlândia Maria Cabral Borges

Universidade Federal do Ceará

---

Profa. Dra. Márcia Teixeira Nogueira

Universidade Federal do Ceará

Para minha mãe acadêmica, Maria Elias Soares.

## AGRADECIMENTOS

À minha orientadora Profa. Dra. Maria Elias Soares que, durante o percurso da pós-graduação, esteve acompanhando o meu caminhar, tropeçar, cair, levantar, ensinando-me a ser firme, persistente, lúcido e questionador sobre o meu próprio processo de produção. Foi graças a senhora que compreendi o que é ser crítico sobre si mesmo em busca de melhorias, sempre dizendo as coisas certas e fazendo os questionamentos mais pertinentes para que eu conseguisse chegar até onde eu gostaria com este trabalho. Com seu bondoso e lindo coração, a senhora sempre será a minha *mãe acadêmica*. Por este motivo e por outros que não caberiam nesta página, agradeço infinitamente. Chegando ao final de um ciclo iniciado ao seu lado no Mestrado, consigo acreditar que, enfim, posso trilhar meu caminho acadêmico daqui para a frente com tudo o que a senhora me ensinou, acadêmica e emocionalmente.

Agradeço à Profa. Dra. Elisângela Nogueira Teixeira, coorientadora deste trabalho, pelas conversas de coorientação, sobretudo no início do percurso de doutoramento, quando ainda discutíamos sobre qual o melhor método para encontrar o fenômeno deste estudo e pensávamos em mesclar estudos em *corpora* linguísticos com pesquisa experimental. Seu olhar crítico sobre o texto da introdução deste trabalho foi de grande valia, tendo sido uma pena não ter sido expandido para as demais partes do texto, o que fez com que, ao final do percurso, não fosse de seu conhecimento, as resenhas feitas para a fundamentação teórica, os *corpora* linguísticos utilizados, nossos resultados e conclusões. Fica aqui meu agradecimento pela ajuda dada e por algumas ligações em que foi possível esclarecer algumas dúvidas que pairavam em minha mente.

Agradeço aos professores doutores participantes da banca de defesa deste trabalho, em especial ao Prof. Dr. Márcio Leitão que, com um olhar crítico-construtivo, trouxe questões a serem melhor discutidas na versão final da tese. Prestar meus agradecimentos também à Profa. Dra. Paula Lenz por todas as contribuições e à Profa. Dra. Márcia Nogueira pelas contribuições desde a qualificação do projeto em andamento.

Agradeço, com o maior apreço, à Profa. Dra. Vlândia Borges que, com bastante bondade e profissionalismo, ajudou na compreensão de pontos

importantes para a versão final desta tese, sobretudo no que tange ao tratamento estatístico dado aos resultados da descrição linguística.

Agradeço também ao amigo Prof. Dr. Sávio Cavalcante que, além de muitas conversas sobre o meu objeto de estudo ao longo dos quatro anos de doutorado, esteve à frente da empreitada de, junto a mim, encontrar o *software* estatístico que seria de melhor manuseio para cumprir os objetivos buscados.

A todos que compõem o Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará, desde os colegas da secretaria, representados aqui na pessoa da querida *tia* Antônia, que, mesmo virtualmente, sempre estão a postos para ajudar em qualquer necessidade. Sobretudo, agradeço à coordenadora de nossa pós, Profa. Dra. Rosemeire Selma Monteiro, que, com seu imenso coração e bondade, sempre ajuda a solucionar os problemas que vão aparecendo, sendo um exemplo de garra e perseverança em todo e qualquer momento. Eu gostaria que a senhora soubesse que, para mim, é um exemplo de profissional e de pessoa. Ademais, gostaria de agradecer a todos os professores do PPGL que, direta ou indiretamente, contribuíram para os meus aprendizados sobre Linguística.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de ajuda financeira recebida ao longo de todo o percurso de doutoramento, o que me deu suporte financeiro necessário para a caminhada deste percurso do início ao fim e que, em tempos conturbados como os de nossa vida política atual, soa quase como um benefício que poucos conseguem. Sigo torcendo para que a vida acadêmica brasileira resista e siga firme. Viva a ciência!

À minha amiga-irmã, Julianne Pita, por suportar meus surtos, medos, esperanças e loucuras que só sabe quem convive comigo de perto. Obrigado pela sua paciência e bondade.

À minha mãe, Vernusa Eline Veras Lima, meu exemplo de força e determinação. Obrigado por sempre me apoiar e me encorajar, sobretudo no caminho do conhecimento.

Ao meu companheiro, Leonardo de Moraes Vieira, meu exemplo de bondade e serenidade. Agradeço por me ajudar a reconhecer o lado bom de tudo e por me fazer aprender que devemos sempre ser agradecidos por cada momento e oportunidade que temos.

À luz do meu caminho, meu pequeno, Pietro Lima Veras. É para você e por você que sigo tentando melhorar, crescer e evoluir como pessoa, para lhe ensinar o que a vida tem para oferecer e como devemos ser persistentes e centrados, sempre com um objetivo a conquistar, pois a vida é isso: lutar.

Aos meus familiares, representados, aqui, por minhas tias, Verlane Edna, Verbena Elane, Verlúcia Eneida e por meus tios, Vernaldo Edson e Verberth Edward, que são a continuação do que meus avós ensinaram, Otaviano Edward e Maria de Lourdes (*in memoriam*). Agradeço a vocês e, conseqüentemente, às suas crias, meus primos e primas e aos filhos destes, afinal de contas, cada um de vocês, de uma maneira ou de outra, segue comigo, onde quer que eu vá, pois este, para mim, é o significado da palavra *família*. Minhas conquistas são também de cada um de vocês.

Por fim, agradeço a Deus pelo dom da vida e por me mostrar o quão belo é viver.

## RESUMO

Esta tese tem como objetivo principal explicar como estão distribuídos nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos no processo de correferência, preenchendo centros anafóricos na posição de sujeito em Português Brasileiro (PB). Tomando por base a Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN; 1995) e os trabalhos que explicam os efeitos do processamento das formas correferenciais, como a Penalidade do Nome Repetido (GORDON; GROZS; GUILLIOM, 1993) e a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008), trabalhamos com a hipótese de que, em algum contexto, o nome repetido seria uma forma de uso frequente. Coletando dados de *corpora* a partir de categorias de análise pautadas nos principais efeitos psicolinguísticos sobre o processamento anafórico, procuramos evidências para nossa hipótese em dois estudos, compostos por (i) análise de dados de *corpus* escrito, contendo textos escritos na norma padrão da língua portuguesa e (ii) análise de dados de *corpus* de fala, contendo textos falados em contextos de uso também da norma padrão. Analisando a quantidade de ocorrências de um termo em dados de uso autêntico da língua, contabilizamos, através de formulários, a quantidade de ocorrências de nomes repetidos, de nomes recategorizados, de pronomes plenos e de pronomes nulos na posição de sujeito, levando em consideração a distância entre a expressão correferencial e seu referente, o tipo de sentença ou de parágrafo no qual ocorre a retomada, a posição sintática do referente, o tipo de verbo da sentença anterior à retomada e o tipo de artigo do sintagma nominal ao introduzir o referente. A análise conjunta dos resultados coletados nos *corpora* sugere que há contextos nos quais o nome repetido é mais frequentemente usado e que a escolha da forma correferente a ser usada está diretamente ligada a dois fatores: a distância entre o antecedente e a retomada em quantidade de palavras e a disposição textual no qual a retomada é efetuada.

**Palavras-chave:** Correferência. Psicolinguística. Descrição linguística.

## ABSTRACT

The main objective of this dissertation is to explain how repeated names, recategorized names, overt pronouns and null pronouns are distributed in the co-referential process, filling anaphoric centers in the subject position in Brazilian Portuguese (BP). The study was based on the Centering Theory (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN; 1995) and on the works that explain the effects of processing co-referential forms, such as the Repeated Name Penalty (GORDON; GROZS; GUILLIOM, 1993) and the Overt Pronoun Penalty (GERLOMINI-LEZAMA, 2008). The working hypothesis was that, in some contexts, the repeated name would be the most frequently used form, and this would be the evidence that this form is more easily processed than the overt pronoun or the null pronoun. Data were collected from the analysis categories which were based on the main psycholinguistic effects on anaphoric processing. Two types of analyses were conducted for that purpose: (i) analysis of written *corpus*, containing texts written in the standard Portuguese language and (ii) analysis of speech *corpus*, containing texts spoken in the standard norm. The number of co-referential forms in authentic use in the language was analyzed. In order to analyze the occurrences of these forms, the number of repeated names, recategorized names, overt pronouns and null pronouns in the subject position were counted, taking into account the distance between the co-referential expression and its referent, the type of sentence or paragraph in which the resumption occurs, the syntactic position of the referent, the type of verb in the sentence prior to the resumption and the type of article in the noun phrase when introducing the referent. The combined analyses of the results collected in both *corpora* suggest that there are contexts in which the repeated name is more frequently used, and that the corresponding form to be used is directly linked to two factors: the distance between the antecedent and the co-referent form in number of words and the textual arrangement in which the co-referent form appears.

**Keywords:** Co-reference. Psycholinguistics. Linguistic description.

## LISTA DE IMAGENS

Figura 01 – O modelo clássico da memória de trabalho.....	89
Figura 02 – Sistema de memorização.....	92
Figura 03 – Primeiro modelo de memória de trabalho.....	93
Figura 04 – Modelo atual de memória de trabalho.....	94
Figura 05 – Subdivisões da Memória de Longo Prazo.....	97
Figura 06 – O modelo clássico de memória.....	99

## LISTA DE QUADROS

Quadro 01 – As características das tipologias das línguas naturais a partir do surgimento ou não da categoria vazia <i>pro</i> .....	39
Quadro 02 – As características das tipologias das línguas naturais a partir do surgimento ou não da categoria vazia <i>pro</i> introduzindo a possibilidade de línguas parcialmente <i>pro-drop</i> .....	40
Quadro 03 – Estímulos do segundo experimento.....	52
Quadro 04 – Hierarquia dos marcadores de acessibilidade.....	84
Quadro 05 – Os pronomes nominais e o paradigma verbal em PB.....	102
Quadro 06 – Composição do <i>corpus</i> Lácio-Ref (modificado).....	152
Quadro 07 – Quantidade de palavras e de caracteres sem e com espaço entre o termo antecedente e a expressão correferente.....	156
Quadro 08 – Averiguação do efeito de paralelismo sintático e do efeito de proeminência sintática.....	157
Quadro 09 – Expressão correferente na posição de sujeito .....	158
Quadro 10 – Tipo de sentença entre o antecedente e a retomada.....	159
Quadro 11 – Tipo de verbo presente na sentença anterior à retomada .....	159
Quadro 12 – Tipo de artigo existente no SN.....	160
Quadro 13 – Exemplo da planilha em Excel do parâmetro de contabilização dos dados com ou sem o efeito de Paralelismo Sintático.....	161
Quadro 14 – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em Português Brasileiro de acordo com a quantidade de palavras entre o antecedente a retomada em textos escritos .....	206
Quadro 15 – Escala de posições sintáticas que introduzem o referente quando não há Efeito de Paralelismo Sintático na escrita.....	208
Quadro 16 – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB de acordo com a disposição textual em textos escritos.....	212
Quadro 17 – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB de acordo com a disposição textual em textos escritos (modificado).....	212
Quadro 18 – Estratificação do <i>corpus</i> PORCUFORT.....	218
Quadro 19 – Normas de transcrição adotadas no PORCUFORT.....	220

Quadro 20 – Exemplo da planilha do Excel do parâmetro de contabilização dos dados com ou sem efeito de Paralelismo Sintático.....	223
Quadro 21 - Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB falado de acordo com a quantidade de palavras entre o antecedente a retomada em textos falados.....	256
Quadro 22 – Escala de posições sintáticas que introduzem o referente quando não há Efeito de Paralelismo Sintático na fala.....	258
Quadro 23 – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB de acordo com a disposição textual em textos falados.....	261

## LISTA DE GRÁFICOS

Gráfico 01 – Quantidade de ocorrência de nomes e pronomes como expressões correferentes na posição de sujeito em textos do <i>corpus</i> Lácio-Ref.....	163
Gráfico 02 – Quantidade de ocorrência de nomes, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do <i>corpus</i> Lácio-Ref.....	164
Gráfico 03 – Quantidade de ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do <i>corpus</i> Lácio-Ref.....	166
Gráfico 04 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há até 10 palavras entre o antecedente a retomada em textos do <i>corpus</i> Lácio-Ref.....	167
Gráfico 05 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 11 a 50 palavras entre o antecedente a retomada em textos do <i>corpus</i> Lácio-Ref.....	169
Gráfico 06 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 51 a 100 palavras entre o antecedente a retomada em textos do <i>corpus</i> Lácio-Ref.....	170
Gráfico 07 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há mais de 100 palavras entre o antecedente a retomada em textos do <i>corpus</i> Lácio-Ref.....	172
Gráfico 08 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes em até 10 caracteres sem espaço entre o antecedente a retomada em textos do <i>corpus</i> Lácio-Ref.....	174
Gráfico 09 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes de 11 a 50 caracteres sem espaço entre o antecedente e a retomada em textos do <i>corpus</i> Lácio-Ref.....	175

Gráfico 10 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes de 51 a 100 caracteres sem espaço entre o antecedente e a retomada em textos do <i>corpus</i> Lácio-Ref.....	176
Gráfico 11 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes em mais de 100 caracteres sem espaço entre o antecedente e a retomada em textos do <i>corpus</i> Lácio-Ref.....	177
Gráfico 12 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes em posição ou não de sujeito nos textos do <i>corpus</i> Lácio-Ref.....	183
Gráfico 13 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes em posições sintáticas que não de sujeito nos textos do <i>corpus</i> Lácio-Ref.....	184
Gráfico 14 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações subordinadas nos textos do <i>corpus</i> Lácio-Ref.....	186
Gráfico 15 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas justapostas nos textos do <i>corpus</i> Lácio-Ref.....	188
Gráfico 16 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas em períodos distintos nos textos do <i>corpus</i> Lácio-Ref.....	189
Gráfico 17 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em parágrafos distintos nos textos do <i>corpus</i> Lácio-Ref.....	191
Gráfico 18 – Ocorrência verbos intransitivos, verbos transitivos e verbos de ligação nos textos do <i>corpus</i> Lácio-Ref.....	193
Gráfico 19 – Ocorrência artigos definidos, indefinidos ou outras formas de definitude precedendo o referente nos textos do <i>corpus</i> Lácio-Ref.....	194
Gráfico 20 – Quantidade de nomes e pronomes como expressões correferentes na posição de sujeito em textos do <i>corpus</i> PORCUFORT.....	225

Gráfico 21 – Quantidade de nomes, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do <i>corpus</i> PORCUFORT.....	226
Gráfico 22 – Quantidade de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do <i>corpus</i> PORCUFORT.....	227
Gráfico 23 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há até 10 palavras entre o antecedente a retomada em textos do <i>corpus</i> PORCUFORT.....	229
Gráfico 24 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 11 a 50 palavras entre o antecedente e a retomada em textos do <i>corpus</i> PORCUFORT.....	231
Gráfico 25 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 51 a 100 palavras entre o antecedente e a retomada em textos do <i>corpus</i> PORCUFORT.....	232
Gráfico 26 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há mais de 100 palavras entre o antecedente a retomada em textos do <i>corpus</i> PORCUFORT.....	234
Gráfico 27 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em posição ou não de sujeito nos textos do <i>corpus</i> PORCUFORT.....	236
Gráfico 28 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em posições sintáticas que não de sujeito nos textos do <i>corpus</i> PORCUFORT.....	237
Gráfico 29 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações subordinadas nos textos do <i>corpus</i> PORCUFORT.....	238
Gráfico 30 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas justapostas nos textos do <i>corpus</i> PORCUFORT.....	240

Gráfico 31 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas em períodos distintos nos textos do <i>corpus</i> PORCUFORT.....	241
Gráfico 32 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em parágrafos subsequentes nos textos do <i>corpus</i> PORCUFORT.....	243
Gráfico 33 – Ocorrência verbos intransitivos, verbos transitivos e verbos de ligação nos textos do <i>corpus</i> PORCUFORT.....	244
Gráfico 34 – Ocorrência artigos definidos, indefinidos ou outras formas de definitude precedendo o antecedente nos textos do <i>corpus</i> PORCUFORT.....	245

## LISTA DE SIGLAS E ABREVIATURAS

Cb = *Backward-looking Center* (Centro Anafórico)

Cf = *Forward-looking Center* (Centro Catafórico)

ecs = *empty categories* (Categorias Vazias)

OPP = *Overt Pronoun Penalty* (Penalidade do Pronome Pleno)

PB = Português Brasileiro

RNP = *Repeated Name Penalty* (Penalidade do Nome Repetido)

SNs = Sintagmas Nominais

## SUMÁRIO

1	INTRODUÇÃO .....	20
2	TEORIAS E HIPÓTESES SOBRE A CORREFERÊNCIA ANAFÓRICA.....	28
2.1	Correferência anafórica.....	28
2.1.1	<i>A retomada anafórica</i> .....	29
2.2	Teorias linguísticas sobre o processamento anafórico.....	35
2.2.1	<i>Teorias e hipóteses sintáticas sobre o processamento anafórico</i> .....	36
2.2.1.1	<i>O parâmetro pro-drop</i> .....	37
2.2.1.2	<i>Teoria da Centralização</i> .....	41
2.2.1.3	<i>Penalidade do Nome Repetido</i> .....	45
2.2.1.4	<i>Penalidade do Pronome Pleno</i> .....	48
2.2.1.5	<i>Proeminência Sintática</i> .....	51
2.2.1.6	<i>Paralelismo Sintático</i> .....	54
2.2.2	<i>Teorias e hipóteses semânticas sobre o processamento anafórico</i> .....	60
2.2.2.1	<i>Teoria da sobreposição semântica</i> .....	61
2.2.2.2	<i>Hipótese da Carga Informacional</i> .....	67
2.2.2.3	<i>Teoria do Processamento em Dois Estágios</i> .....	76
2.2.3	<i>Teoria pragmática sobre o processamento correferencial</i> .....	81
2.2.3.1	<i>Teoria da Acessibilidade</i> .....	81
2.3	Memória .....	86
2.3.1	<i>O conceito de memória</i> .....	86
2.3.2	<i>A retenção da informação e os tipos de memória</i> .....	88
2.3.3	<i>Memória de trabalho e memória de longo prazo</i> .....	92
3	OS ESTUDOS SOBRE O PROCESSAMENTO ANAFÓRICO EM PORTUGUÊS.....	100
3.1	Breve panorama sobre a retomada anafórica em dados de <i>corpora</i> linguísticos.....	100
3.1.1	<i>A variação do uso de pronomes na posição de sujeito</i> .....	100
3.1.2	<i>A variação do uso de pronomes na posição de objeto</i> .....	104
3.2	Panorama sobre a retomada anafórica em pesquisas experimentais .....	109

<b>4</b>	<b>FATORES LINGUÍSTICOS LIGADOS AO PROCESSAMENTO CORREFERENCIAL.....</b>	<b>147</b>
4.1	Objetivos e hipóteses.....	147
4.2	Procedimento para a obtenção de dados empíricos no <i>corpus</i> escrito.....	150
4.3	Categorias de análise.....	154
4.4	Resultados.....	160
4.5	Discussão.....	196
<b>5</b>	<b>EVIDÊNCIAS SOBRE O PROCESSAMENTO CORREFERENCIAL EM DADOS DE <i>CORPUS</i> DE FALA.....</b>	<b>216</b>
5.1	Procedimentos para a obtenção de dados empíricos no <i>corpus</i> de fala.....	217
5.2	Resultados.....	223
5.3	Discussão.....	247
<b>6</b>	<b>DISCUSSÃO GERAL.....</b>	<b>266</b>
<b>7</b>	<b>CONCLUSÕES.....</b>	<b>282</b>
	<b>REFERÊNCIAS.....</b>	<b>292</b>

## 1 INTRODUÇÃO

O estudo da coesão/coerência feito pela ótica da Linguística Textual<sup>1</sup> trata de diversas questões, sendo possível analisar e buscar explicações sobre como uma entidade é retomada ao longo do texto, por meio do estudo do mecanismo da *correferência anafórica* (LYONS, 1977; BENVENISTE, 2005; KOCH, 2015). A correferência anafórica é o nome dado ao mecanismo pelo qual um referente previamente inserido em um dado texto<sup>2</sup> é retomado por outro elemento a ele também referente (GORDON; HENDRICK, 1997).

De acordo com Koch (2015), há diversos signos linguísticos que podem retomar uma entidade ao longo do texto, podendo ser encontradas retomadas feitas pelo uso do nome, da mesma maneira usada na introdução do referente, de nome recategorizado, de pronomes, de hiperônimos, de hipônimos, de sinônimos ou de antônimos.

Diversas pesquisas (GROSZ; SIDNER, 1986; GORDON; GROSZ; GILLIOM, 1993; GROSZ; JOSHI; WEINSTEIN, 1995; ARIEL, 1991; KENNISON; GORDON, 1997; CHAMBERS; SMITH, 1998; ALMOR, 1999; GERLOMINI-LEZAMA, 2008), na área da Psicolinguística, têm buscado explicar de que maneira uma forma correferente é capaz de interferir durante o processamento anafórico, sendo mais ou menos custosa cognitivamente a ponto de causar danos na construção de sentido.

Entendendo o processamento correferencial como uma tarefa a ser executada na memória de trabalho (BADDELEY, 2000, 2007; 2017; WILSON, 2009; IZQUIERDO, 2011; LAPOINTE, 2012; BADDELEY; EYSENCK, 2014; WARD, 2015, GROOME; EYSENCK, 2016; AXMACHER; RASCH, 2017), que pode ter formas mais fáceis de serem percebidas e processadas na mente, diversas pesquisas (GROSZ; SIDNER, 1986; GORDON; GROSZ; GILLIOM, 1993; GROSZ; JOSHI; WEINSTEIN, 1995; ARIEL, 1991; KENNISON; GORDON, 1997; CHAMBERS; SMITH, 1998; ALMOR, 1999; GERLOMINI-LEZAMA, 2008) têm dado evidências de que os pronomes, sejam plenos ou

---

<sup>1</sup> Segundo Koch (2015), um dos objetos de estudo da Linguística Textual, desde a chamada Virada Cognitivista, trata sobre como um objeto do discurso é representado por uma expressão anafórica, nominal ou pronominal, sendo necessário compreender como é dado o estabelecimento da relação entre as informações presente na memória discursiva.

<sup>2</sup> Entendemos, neste trabalho, texto como termo sinônimo de sentença ou de segmento discursivo (GROSZ; JOSHI, 1986; GROSZ; JOSHI; WEINSTEIN, 1995; KOCH, 2015)

nulos, tendem a ser as formas mais fáceis de processar quando em função anafórica.

A partir disto, sabendo que em Português Brasileiro (doravante PB), há, dentre as diversas formas linguísticas usadas para retomar um referente (KOCH, 2015), a retomada anafórica, esta pode ser feita pelo uso de (i) repetição de um nome, (ii) de um pronome pleno<sup>3</sup> e (iii) de um pronome nulo<sup>4</sup>, muitas pesquisas (LEITÃO, 2005a, 2005b; QUEIROZ; LEITÃO, 2008; ALBUQUERQUE, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; QUEIROZ, 2009; MAIA; CUNHA-LIMA, 2011; MORGADO, 2011; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; MAIA; CUNHA-LIMA, 2012; VASCONCELOS; LEITÃO, 2012; ALVES, 2012; MAIA, 2013; TEIXEIRA, 2013; SIMÕES, 2014; LIMA, 2014; LIMA, 2015; GONDIM, 2017) têm sido desenvolvidas a fim de observar como se comportam os indivíduos durante a leitura de textos com as formas correferentes supracitadas e, assim, explicar o custo cognitivo causado pelo uso de tal forma durante o processamento correferencial.

As pesquisas desenvolvidas no Brasil que têm buscado explicar o custo cognitivo do processamento anafórico podem ser enquadradas em dois grandes grupos: o grupo que analisa a forma anafórica dentro da mesma sentença, nível intrassentencial; e o grupo que analisa a forma correferente entre sentenças distintas, mas interligadas, como em parágrafos ou sentenças postas lado a lado, nível intersentencial, sendo considerada ou não a distância entre o referente e sua expressão correferencial. Neste grupo, temos os estudos de Maia e Cunha-Lima (2011, 2012, 2014), Maia (2013), Lima (2015) e Almor *et al.* (2017), já os estudos de Leitão (2005) e colaboradores (QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2008; LEITÃO, RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; GONDIM, 2017) foram desenvolvidos tanto no nível intra quanto interssentencial.

Os trabalhos desenvolvidos por Leitão e seus colaboradores têm analisado as retomadas anafóricas em sentenças justapostas ou em sentenças coordenadas, interligadas por conjunção ou separadas por pontos, tanto por

---

<sup>3</sup> Também encontrando na literatura da área como pronome foneticamente realizado (CHOMSKY, 1981).

<sup>4</sup> Também encontrado na literatura da área como categoria vazia, ou seja, um pronome não realizado foneticamente, ou *pro* (leia-se *prozinho*), ou ainda sendo representado por  $\emptyset$  enquanto correspondente a sujeito nulo. (CHOMSKY, 1981; MIOTO; SILVA; LOPES, 2013)

participantes com desenvolvimento normal quanto por participantes com distúrbios, doenças ou afasias<sup>5</sup>.

Em todos os estudos supracitados, os pesquisadores deste grupo têm manipulado o contexto no qual o nome repetido e o pronome pleno têm função de expressão correferencial, a fim de analisar o tempo de leitura da expressão correferente e explicar como os participantes compreendem melhor as sentenças lidas.

De acordo com os resultados reportados nos trabalhos dos pesquisadores que compõem este grupo, sempre são encontrados indícios de que o pronome pleno é mais facilmente compreendido do que o nome repetido nos contextos analisados, apontando para o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) em PB e corroborando o preenchimento do centro anafórico conforme previsto pela Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995).

Na literatura da área, Grosz e Sidner (1986) apontam que, em contextos de sentenças justapostas, no nível discursivo, uma vez introduzida uma entidade a ser retomada novamente em centros anafóricos<sup>6</sup> ao longo do segmento discursivo, esta deve ser retomada por pronome pleno, que é considerado o veículo natural de manutenção da informação ao longo do texto.

Segundo Grosz e Sidner (1986), o uso de um nome repetido em um centro anafórico gera maior custo de processamento, ou seja, torna o discurso mais complexo, pois não há necessidade da repetição do nome em uma região na qual o pronome exerce perfeitamente a mesma função, tornando-se, assim, a forma correferente preferida e menos custosa.

Munidos desta informação e adotando esta concepção, Gordon, Grosz e Guilliom (1993) desenvolveram o conceito da Penalidade do Nome Repetido, que é o nome dado ao custo de processamento causado pelo uso do nome repetido em centros catafóricos que deveriam, preferencialmente, ser preenchidos por pronomes plenos, o que tende a ocorrer quando a retomada preenche a posição de sujeito da sentença, afirmando que este efeito de

---

<sup>5</sup> Para mais informações, ver a subseção 2.5.

<sup>6</sup> Centro anafóricos são os locais do segmento discursivo em que a entidade previamente introduzida deve ser retomada, a fim de suscitar na mente do receptor, a referência discursiva (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995)

penalização<sup>7</sup> do nome repetido não ocorre quando a retomada for executada em outras posições sintáticas.

Contudo, Almor (1999) encontrou indícios de que, quando comparado o tempo de leitura dos nomes repetidos e dos pronomes na posição de objeto, também ocorre o efeito da Penalidade do Nome Repetido, apontando para uma maior extensão das predições desta teoria.

Sabendo que a maioria destas pesquisas foram desenvolvidas para explicar o processamento de nomes e de pronomes em língua inglesa, Gerlomini-Lezama (2008) analisou a aplicabilidade destes conceitos em uma língua que prevê dois tipos de pronomes em centros anafóricos – os pronomes plenos e os pronomes nulos, como ocorre em línguas como o italiano, o espanhol e o português, em vez de apenas um tipo, como ocorre em inglês, alemão e francês.

Quando comparado o tempo de processamento dos nomes repetidos com relação ao pronome pleno e ao pronome nulo, Gerlomini-Lezama (2008) encontrou indícios de que o pronome nulo é a expressão correferencial mais fácil de ser processada em Espanhol Argentino, ocorrendo maior custo de processamento não somente do nome repetido, mas também do pronome pleno. Para este tipo de efeito, o autor criou o conceito da Penalidade do Pronome Pleno.

Considerando que os trabalhos de Leitão (2005) e seus colaboradores (QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; GONDIM, 2017) cruzam, em seus estudos, os custos de processamento somente dos nomes repetidos e dos pronomes plenos, Maia e Cunha-Lima (2011, 2012, 2014) desenvolveram pesquisas, pautados nos resultados de Gerlomini-Lezama (2008) e na proximidade das características linguísticas do Espanhol Argentino e do Português Brasileiro, cruzando o tempo de leitura dos nomes repetidos, dos pronomes plenos e dos pronomes nulos.

---

<sup>7</sup> Gordon, Grosz e Guiliom (1993) usam o termo *penalidade* para se referir ao tempo de processamento da leitura que, quando mais lento, indica que o leitor teve dificuldade de compreender o segmento lido. Neste sentido, Gordon, Grosz e Guiliom (1993) entendem que o nome é penalizado pelo pronome pelo fato de este ser lido mais rapidamente, o que mostra o estabelecimento da ligação do par antecedente-anáfora feito pela ligação nome/pronome como resultado da compreensão mais rápida do que acontece quando a ligação é entre um nome (antecedente) e a repetição do mesmo nome (anáfora).

Os resultados obtidos nas pesquisas de Maia e Cunha-Lima mostram indícios de que a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) é percebida nos contextos analisados, mas não encontraram nenhum tipo de evidência com relação à Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e, assim, passaram a recusar esta penalidade em PB.

Levando em consideração que a discussão feita pelos pesquisadores em geral acerca do processamento correferencial em PB está atrelada aos custos de processamento das formas correferentes em centros anafóricos e a evidenciar quais são as formas mais facilmente processadas, Lima (2015) analisou o tempo de leitura dos nomes repetidos, dos pronomes plenos e dos pronomes nulos em textos.

Lima (2015) observou se, em dados contextos nos quais a forma correferente e seu referente estivessem sintaticamente distantes, haveria ainda penalizações de uma forma em detrimento da outra e encontrou indícios de que, em dados contextos, há Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e em outros, há Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008), mas nunca há coocorrência de ambas as penalidades em um mesmo contexto, o que indica que o efeito de penalização das formas correferentes depende do contexto de uso e não somente do tipo de retomada.

Gondim (2017) analisou se o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) ocorre em PB, por ser um efeito multifatorial, envolvendo desde questões sintáticas, como a posição da retomada, questões semânticas, como a quantidade de entidades retomadas no centro anafórico, até questões metodológicas, como a divisão das sentenças em segmentos durante as análises que levam a encontrar ou não um efeito de penalização.

A pesquisadora encontrou indícios de que as questões metodológicas têm influência nos resultados das pesquisas, pois analisar o tempo de processamento da região na qual a retomada é executada e comparar com o tempo de processamento da leitura de toda a sentença ou segmento discursivo leva a resultados diferentes.

Como a autora se deteve em analisar especificamente o processamento dos nomes repetidos com relação aos pronomes plenos, mais uma vez foram deixadas de lado possíveis evidências de que o pronome nulo poderia ser a forma mais facilmente processada em PB, o que faz com que os resultados não

sejam parametrizados no uso da língua, mas somente na análise do que uma forma linguística pode acarretar como custo durante a leitura de textos bastante manipulados.

Além disso, as pesquisas desenvolvidas em PB têm focado na leitura de sentenças e/ou de segmentos discursivos geralmente curtos sem estar vinculado ao tipo de produção, pois não se sabe se os itens experimentais são criados para estarem mais próximos da língua falada ou da língua escrita, o que pode também influenciar os resultados encontrados.

O que é reportado em cada uma dessas pesquisas é que o grupo dos participantes que são levados para ler o conjunto experimental em laboratório é composto por falantes nativos de PB, mas não há menção alguma sobre o parâmetro que leva à construção dos itens a serem lidos.

Se observarmos as nossas próprias produções, se fizermos uma rápida busca para averiguar se as formas correferentes são usadas como previsto nas pesquisas de Leitão (2005) e seus colaboradores (QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; GONDIM, 2017) ou como previsto por Maia e Cunha-Lima (2011, 2012, 2014), Maia (2013) e Lima (2015), somos induzidos a acreditar que os nomes repetidos não devem surgir ao longo da retomada discursiva, já que os pesquisadores afirmam que o nome repetido é sempre a forma mais difícil de ser processada.

Levando em consideração o conhecimento acerca dos resultados de pesquisas sobre o processamento correferencial em PB, surgem questões como as explicitadas abaixo.

(a) Há diferenças em como os usuários da língua retomam as entidades ao longo do discurso a depender de o texto ser oral ou escrito?

(b) Quando os nomes tendem a ser retomados como forma correferente, a repetição é mais fácil de ser processada por meio de grupos semânticos de sentido próximo (sinônimos) do que quando o nome já usado é meramente repetido?

(c) Se o pronome nulo tende a ser mais frequente para retomar o referente, fatores como a distância entre o referente e a retomada ou o tipo de construção sintática faz surgir o uso do pronome pleno?

Estas perguntas surgem pelo fato de esperarmos que, em determinados contextos, o nome surja como forma frequentemente usada nos dados de *corpus* que não aponte para o padrão explicado por Leitão (2005) e seus colaboradores (QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; GONDIM, 2017), nem para o padrão explicado por Maia e Cunha-Lima (2011, 2012, 2014), Maia (2013) e Almor *et al.* (2017), ou seja, esperamos encontrar contextos em que os nomes não sejam penalizados pelos pronomes, plenos ou nulos.

Assim, percebemos que a análise de como as formas correferentes são distribuídas em contextos de uso autêntico, é extremamente necessária, pois, em Português Brasileiro, as evidências sobre o processamento correferencial, de acordo com os estudos psicolinguísticos, podem não ser capazes de explicar os padrões encontrados quando comparadas a fala e a escrita.

Portanto, levando em consideração as questões que guiam nossa pesquisa, buscamos observar fatores discutidos, sobretudo em pesquisas psicolinguísticas, por meio de dados de *corpora* linguísticos, a fim de encontrar evidências de uma possível distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito em Português Brasileiro.

Para cumprir nossos objetivos, apresentamos, no capítulo 2, as teorias e hipóteses que tratam sobre fatores sintáticos, semânticos e pragmáticos ligados ao processamento anafórico.

No capítulo 3, mostramos as pesquisas que tratam da investigação acerca das penalidades que podem ocorrer durante a resolução correferencial em PB e a possível variação de uso do pronome nulo. Logo, as teorias e hipóteses apresentadas nos capítulos 2 e 3 estão diretamente relacionadas às pesquisas descritas nos capítulos 4 e 5.

No capítulo 4, descrevemos a metodologia, os resultados e a discussão sobre os resultados encontrados a partir da coleta e análise desenvolvida utilizando o *corpus* escrito Lácio-Ref<sup>8</sup>, sendo nosso objetivo explicar o uso de

---

<sup>8</sup> O Lácio-Web é um projeto desenvolvido por pesquisadores da Universidade de São Paulo que tem como objetivo divulgar e disponibilizar livremente na internet dados linguísticos do Português Brasileiro. O projeto Lácio-Web é composto por seis *corpus*, sendo um deles, o *corpus* Lácio-Ref. A principal característica deste *corpus* é ser composto por textos escritos respeitando a norma culta da língua.

nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em Português Brasileiro, levando em conta fatores sintáticos, além de buscar evidências de quais efeitos encontrados em pesquisas psicolinguísticas são percebidos na língua usada em contexto real.

Em seguida, no capítulo 5, continuamos com a descrição da metodologia, resultados e discussão de resultados, mas a partir da coleta e análise desenvolvida utilizando o *corpus* de fala PORCUFORT<sup>9</sup>, tendo os mesmos objetivos que norteiam toda nossa pesquisa e descritos no capítulo 4.

Já no capítulo 6, fazemos uma discussão geral, a fim de comparar os resultados obtidos tanto no capítulo 4 quanto no capítulo 5 e, assim, fazer um apanhado geral de nossos resultados, para, em discussão com as teorias e hipóteses que guiam nosso trabalho, junto às pesquisas realizadas em PB, buscar explicações para os resultados das análises dos *corpora* linguísticos.

Por fim, no capítulo 7, apresentamos nossas conclusões a partir dos resultados encontrados pelas pesquisas conduzidas para esta tese.

---

<sup>9</sup> O PORCUFORT foi idealizado e organizado pelo Prof. Dr. José Lemos Monteiro com o objetivo de organizar um banco de dados da variedade urbana culta falada pelos fortalezenses.

## 2 TEORIAS E HIPÓTESES SOBRE A CORREFERÊNCIA ANAFÓRICA

Nesta seção, apresentaremos os trabalhos que levaram às construções dos conceitos que serão utilizados em nosso estudo, tais como retomada anafórica, os tipos de formas linguísticas que podem ter a função de anáfora e as teorias linguísticas (sintáticas, semânticas e pragmáticas) que tratam do assunto, assim como revisitaremos os trabalhos desenvolvidos em Português Brasileiro na área do processamento correferencial.

### 2.1 Correferência anafórica

Durante o discurso, fazemos referência à determinada entidade do mundo extralinguístico por meio de palavras. Estas estabelecem a referência ao ser feito um elo referencial entre a palavra (termo linguístico) e aquilo o que este termo representa no mundo (entidade extralinguística) sendo, portanto, os referentes entendidos como objetos do discurso (BHAT, 2004).

Segundo Bhat (2004), o ato de fazer referência é desta forma, a base primordial das línguas e, por isso, é necessário entender que, durante o uso de uma língua ora fazemos referência a entidades que estão dentro do contexto linguístico (endofórico), ora a entidades que estão no contexto extralinguístico (exofórico), conforme exemplificado de (1) a (3)<sup>10</sup>.

- (1) Eu<sub>i</sub> trouxe este livro para você<sub>j</sub> e este para você<sub>k</sub>.
- (2) Eu<sub>i</sub> comprei este livro para Ana<sub>j</sub>. Ela<sub>j</sub> ficou muito feliz.
- (3) O policial<sub>i</sub> prendeu o ladrão<sub>j</sub> em sua<sub>ij</sub> casa.

Podemos perceber que, em (1), os pronomes pessoais fazem referência a entidades que estão fora do discurso, ou seja, é uma retomada exofórica e que, além disso, o pronome de segunda pessoa retoma duas entidades distintas, uma vez que é perceptível tratar-se de duas pessoas diferentes e facilmente reconhecíveis no contexto de uso desta sentença.

Já em (2), o pronome pessoal da primeira pessoa faz referência a uma entidade reconhecível no contexto de uso (contexto *pragmático*), assim como o

---

<sup>10</sup> Os exemplos mostrados nesta seção foram inspirados nos exemplos mostrados por Bhat (2004).

nome próprio também. Contudo, o uso do pronome pessoal de terceira pessoa *ela* faz referência a uma entidade previamente inserida no contexto linguístico (endofórico).

Enquanto em (3) são introduzidas duas entidades *o policial* e *o ladrão*, que são retomadas por meio do demonstrativo *sua*, o que gera ambiguidade, já que este pronome pode fazer referência a qualquer uma das duas entidades previamente introduzidas no discurso.

Logo, é notório que o comportamento dos pronomes não é o mesmo já que, em (1), os pronomes pessoais de primeira e de segunda pessoa fazem referência a entidades reconhecidas no contexto de uso, já o pronome pessoal de terceira pessoa, em (2), e o pronome demonstrativo, em (3), fazem referência a entidades previamente introduzidas no contexto linguístico.

Notamos, portanto, que há correferência em (2) e (3) sendo este o mecanismo através do qual um termo linguístico retoma uma entidade previamente introduzida no discurso.

Além disso, é necessário entender que os pronomes são as expressões correferenciais mais facilmente entendidas (GORDON; GROSZ; GUILLIOM, 1993), além de serem apenas uma das formas linguísticas que o falante tem para fazer referência a objetos do discurso e que, mesmo entre os pronomes, há diferenças na maneira como os falantes usam estas formas linguísticas.

### **2.1.1 A retomada anafórica**

Quando um falante/escritor faz referência a um ser do mundo extralinguístico no contexto linguístico, em geral, são usados diversos mecanismos para evitar repetições da mesma forma linguística, já que isto pode acarretar prejuízo à memória durante o processamento. Isto pode acontecer pelo fato de, a depender do tipo de retomada usada para fazer referência à entidade à qual se liga, haver custos de compreensão para a memória a depender da quantidade de informações ou da carga informacional dada (ALMOR, 1999; BADELLEY, 2002, 2007, 2018; WILSON, 2009; IZQUIERDO, 2011; LAPOINTE, 2012; BADDELEY; EYSENCK, 2014; KEREMER, 2014; WARD, 2015, GROOME; EYSENCK, 2016; AXMACHER; RASH, 2017).

Por conta disto, gera estranhamento ao receptor um contexto no qual uma palavra se repete constantemente como exemplificado em (4)<sup>11</sup>, parecendo artificial se pensarmos na língua que vivenciamos no dia a dia.

(4) Ontem a Ana saiu cedo de casa para que a Ana não chegasse atrasada ao trabalho. A Ana correu bastante, mas a Ana acabou perdendo o ônibus. Por conta disso, a Ana chegou novamente atrasada e o chefe da Ana foi reclamar com a Ana na frente dos demais funcionários da empresa.

É muito fácil perceber que qualquer falante de português seria capaz de dizer que não é preciso repetir tanto a palavra *Ana*, pois há outras formas de fazer a retomada deste referente depois que este já tenha sido introduzido no discurso. Neste caso específico, o falante pode optar por usar um pronome ora foneticamente realizado, conhecido como pronome pleno, ou o pronome não realizado foneticamente, ou pronome nulo<sup>12</sup>, representado pela forma  $\emptyset$ , como exemplificado em (5).

(5) Ontem a Ana saiu cedo de casa para que  $\emptyset$  não chegasse atrasada ao trabalho. Ela correu bastante, mas  $\emptyset$  acabou perdendo o ônibus. Por conta disso, ela chegou novamente atrasada e o chefe dela foi reclamar  $\emptyset$  na frente dos demais funcionários.

Entretanto, é preciso atentar para o fato de que esta relação estabelecida entre os pronomes plenos ou os pronomes nulos, como em (5), é possível somente pelo fato de que eles estão diretamente ligados à entidade previamente introduzida constituindo-se, portanto, como anáforas.

O termo 'anáfora' [...] é comumente usado para se referir a uma relação entre dois elementos linguísticos em que a interpretação de um (chamado anáfora) é de alguma forma determinada pela interpretação de outro (chamado antecedente)<sup>13</sup> (HUANG, 2000, p.01, tradução nossa).

<sup>11</sup> Os exemplos mostrados ao longo desta seção foram inspirados nos exemplos de Huang (2000), Bhat (2004), Benveniste (2005), Levinson (2007) e Koch (2015).

<sup>12</sup> Chomsky (1981) propõe que o pronome foneticamente realizado tem uma contraparte chamada de categoria vazia ou pro, comumente usado em pesquisas psicolinguísticas como pronome nulo. Voltaremos a este ponto mais adiante.

<sup>13</sup> The term 'anaphora' (...) is commonly used to a relation between two linguistics elements, wherein the interpretation of one (called an anaphor) is in some way determined by the interpretation of the other (called an antecedent).

Entendendo, pois, o conceito de anáfora é preciso atentar para contextos de retomadas, como exemplificado em (6), (7) e (8), em que os pronomes ora retomam entidades previamente introduzidas no discurso, ora remetem a entidades que são reconhecidas apenas dentro do contexto extralinguístico.

- (6) Eu<sub>i</sub> disse comprei isso hoje.
- (7) Você<sub>i</sub> chegou atrasada.
- (8) Ana<sub>i</sub> disse que Ø<sub>i</sub> é feliz.

O que acontece em (6) e em (7) é que os pronomes de primeira e segunda pessoa fazem referência a uma entidade que é reconhecida somente no contexto de uso, pois são linguisticamente, de acordo com Levinson (2007), denominados de dêixis.

A dêixis diz respeito às maneiras pelas quais as línguas codificam ou gramaticalizam traços do contexto da enunciação ou do evento da fala e, portanto, também diz respeito a maneiras pelas quais a interpretação das enunciações depende da análise desses contextos de enunciação (LEVINSON, 2007).

De acordo com Soares (1991), um dos problemas da distinção dêixis/anáfora é o uso do dêitico de primeira pessoa, que, por ser de natureza exofórica, aponta, na construção textual, para um enunciador que é pragmaticamente reconhecido no papel de locutor, sendo, portanto, um sujeito ligado à situação do discurso, o que pode, por consequência, ser esperado para os pronomes de segunda pessoa.

Segundo Benveniste (2005), o efeito de interpretação que ocorre entre (6) e (7), não é o mesmo que ocorre em (8), pois os pronomes pessoais de primeira e segunda pessoa são, por natureza, diferentes dos pronomes de terceira pessoa, pois aqueles guardam a noção de pessoa, enquanto este a abole, sendo entendido como a não-pessoa.

A 'terceira pessoa' representa de fato o membro não marcado da correlação de pessoa. É por isso que não há truísmo em afirmar que a não-pessoa é o único modo de enunciação possível para as instâncias do discurso que não devam remeter a elas mesmas, mas que predicam o processo de *não importa quem* ou *não importa o que*, exceto a própria instância, podendo sempre esse *não importa quem* ou *não importa o que* ser munido de uma referência objetiva (BENVENISTE, 2005, p.282).

É, por isso, que se pode dizer que há correferência quando a terceira pessoa retoma uma entidade que não é uma pessoa do discurso, mas alguém que não está marcado como enunciador, sendo objetivamente o assunto do discurso proferido pelo *eu*.

Sendo assim, é possível falar que em (6) e (7) não há correferência por dois critérios: (i) a correferência é um mecanismo linguístico no qual um termo correferente está diretamente ligado, no contexto linguístico, a seu referente previamente introduzido e (ii) os pronomes pessoais de primeira e segunda pessoa, ou melhor, os pronomes dêíticos não estabelecem correferência.

Já em (8), há correferência por obedecer ao primeiro critério acima, quando em (i) o pronome nulo retoma a entidade previamente introduzida e por (ii) os pronomes de terceira pessoa, ou melhor, os pronomes anafóricos estabelecerem correferência. E, além do pronome pessoal de terceira pessoa, as chamadas *proformas*<sup>14</sup>, também estabelecem correferência.

Entretanto, mesmo dentro da análise dos pronomes de terceira pessoa, é preciso observar casos de variância de referência, como exemplificado em (9) e (10).

- (9) O João<sub>i</sub> me disse que ele<sub>i/j</sub> estava feliz.  
 (10) Eu disse para o João<sub>i</sub> que ele<sub>i/j</sub> estava feliz.

O pronome pessoal da terceira pessoa *e/e* ocorre em (9) como um possível correferente do enunciador endofórico *João* ou como correferente de qualquer outro referente em terceira pessoa sendo, neste caso, exofórico. Fato bastante similar ao que acontece em (10), já que o pronome *e/e* pode ser correferente de uma entidade endofórica *João* ou de qualquer outra entidade de referência em terceira pessoa exofórica.

Entretanto, parece pertinente buscar explicações plausíveis com nosso uso no cotidiano que poderia explicar que em (9), o pronome *e/e* só pode retomar a entidade previamente introduzida, já que ela é o sujeito da sentença e é dela que se fala, o que parece ter explicações sobre a marcação da não pessoa (BENVENISTE, 2005), enquanto em (10), pode-se buscar a explicação no fato de que o enunciador *eu* já fala de uma terceira pessoa que, por sua vez, fala de outra terceira pessoa e, assim, seria um caso de correferência endofórica.

---

<sup>14</sup> Para mais informações, ver Bhat (2004).

A verdade é que, nem em (9), nem em (10) há evidências que comprovem se o pronome *e/le* é correferente de uma entidade dentro do contexto linguístico ou fora dele, mas podemos dizer que, quando o pronome da terceira pessoa funciona como correferente de uma forma linguística previamente introduzida no discurso, falamos de correferência anafórica, que é o objeto deste estudo.

Desta forma, entendemos que a correferencialidade se dá entre duas ou mais expressões linguísticas, que fazem referência a uma mesma entidade discursiva, referente ou objeto do discurso, sendo os nomes ou os pronomes os meios linguísticos utilizados como formas correferentes<sup>15</sup>.

Ademais, é necessário advertir que não são somente os pronomes que podem estabelecer a correferência anafórica, pois os nomes previamente introduzidos no discurso podem ser retomado por meio de (i) repetição total ou parcial do mesmo nome, como em (11), (ii) de sinonímia ou parassinonímia, como em (12), (iii) de recategorizações como hiperonímias ou hiponímias, como em (13), (iv) de retomada por termo genérico, como em (14) ou, (v) de descrições nominais, como em (15), conforme proposto por Koch (2014).

(11) Ontem, quando a Ana<sub>i</sub> saiu para comprar uma roupa nova, esperávamos que ela chegasse cedo, mas, como sempre, a Ana<sub>i</sub> demorou.

(12) Todos os dias eu fico contando os minutos para que chegue o fim do expediente e eu possa ir para casa<sub>i</sub>, pois, para mim, não há nada como voltar para o lar<sub>i</sub> depois de um dia cansativo.

(13) Minha mãe levou a televisão<sub>i</sub> para o concerto, pois o aparelho<sub>i</sub>, apesar de novo, já apresentou defeito.

(14) O meu irmão comprou um computador<sub>i</sub> tão moderno que eu nem mesmo sei ligar aquela coisa<sub>i</sub>.

(15) Todos os dias as pessoas<sub>i</sub> questionam sobre o andamento do processo. E o que resta a fazer é dar a mesma informação todos os dias para esse *pessoal de memória curta*<sub>i</sub>.

Para Koch (2014, p.22, grifos da autora), há um mecanismo que, durante o estabelecimento da coesão, leva à “repetição do *mesmo item lexical* ou de *sinônimos, hiperônimos, nomes genéricos*”, mostrando que o nome pode ser substituído por outras formas nominais<sup>16</sup> dentro da cadeia anafórica.

<sup>15</sup> Por tratarmos sobre os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos em função anafórica, ou seja, como formas correferentes, simplificaremos a ligação correferencial aos termos antecedente ou referente e correferente, seguindo o que comumente é usado em pesquisas psicolinguísticas, conforme veremos adiante neste capítulo.

<sup>16</sup> Nesta tese, os termos nome, repetido ou recategorizado, são entendidos como sinônimo do termo chomskyano *sintagma nominal*.

Segundo Basseto (2015), esta possibilidade de mudanças de nomes servindo como expressão anafórica de um referente, cumpre propósitos comunicativos do emissor, mas cria instabilidade no discurso, podendo ocorrer perdas de ligação entre o par antecedente/anáfora.

Tais seleções lexicais são realizadas com o papel de introduzir, retomar e qualificar o referente a partir de um ponto de vista, pois, a cada categorização ou recategorização do referente, o sujeito opera no material linguístico de modo a imprimir uma perspectiva sobre o referente, com vistas à sua finalidade discursiva (BASSETO, 2015, p.10).

De acordo com Cavalcante e Brito (2016, p.125), durante a análise da cadeia referencial de um texto, “não é suficiente explicitar as transformações sofridas pelo referente”, pois dentre outros fatores, é possível perceber que o referente foi recategorizado. Por conta disto, acreditamos que, quando o emissor produz texto, o nome é usado como um recategorizador do referente, entretanto, quando o texto é dado ao receptor o processo de recategorização está feito e o nome usado se torna estável no todo da cadeia referencial, fazendo emergir na superfície textual os nomes recategorizados.

Assim, podemos ver que o fenômeno da anáfora é um campo extremamente fértil não somente para as pesquisas linguísticas, mas também para filósofos, psicólogos, cientistas cognitivos e pessoas que trabalham com inteligência artificial (HUANG, 2000), por diversas razões, dentre as quais podemos enumerar:

Em primeiro lugar, a anáfora representa um dos mais complexos fenômenos da linguagem natural, que, em si, é a fonte de fascinantes problemas. (...) Em segundo lugar, a anáfora tem sido, por algum tempo, considerada uma das poucas ‘sondas extremamente boas’ para aprofundar nossa compreensão da natureza da mente/cérebro humano e, assim, facilitar na busca de respostas para o que Chomsky (1981, 1982, 1986a, 1995) considera ser o problema fundamental da linguística, ou seja, o problema lógico da aquisição da linguagem, um caso especial do princípio de Platão. (...) Em terceiro lugar, a anáfora tem mostrado criar interações entre fatores de ordem sintática, semântica e pragmática. Consequentemente, a anáfora tem fornecido um campo de testes para uma série de hipóteses concorrentes que envolvem a relação entre sintaxe, semântica e pragmática na teoria linguística (HUANG, 2000, p 01-02, tradução nossa)<sup>17</sup>.

---

<sup>17</sup> In the first place anaphora represents one of the most complex phenomena of natural language, which, in itself, is the source ODF fascinating problems, (...) Secondly anaphora has for some time been regarded as one of the few ‘extremely good probes’ in furthering our understanding of the nature mind/brain and thus in facilitating an answer to what Chomsky (1981, 1982, 1986<sup>a</sup>, 1995) considers to be the fundamental problem of linguistics, namely the logical problem of language acquisition, a special case of Plato’s problem. Thirdly anaphora has been shown to

Além disso, podemos perceber que a maneira como a correferência anafórica ocorre depende da expressão anafórica em uso e que, segundo Huang (2000), pode acontecer de quatro grandes maneiras: (i) categorias sintáticas, (ii) condições de verdade, (iii) contextos e (iv) rastreamento de referência do discurso<sup>18</sup>.

Para este estudo, buscamos compreender, especificamente, como ocorrem as anáforas enquanto (i) categorias sintáticas e (ii) contextos de uso da anáfora, pois tratamos de compreender como nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos, em saliência sintática, ou seja, na posição de sujeito da sentença, podem ser usados enquanto expressões anafóricas das quais o usuário da língua dispõe para efetuar retomadas ao longo do discurso.

Para melhor compreender como expressões anafóricas são ligadas entre si e como são compreendidas pelos interlocutores, descrevemos, a seguir, estudos já realizados sobre o tema e a partir dos quais temos as chamadas teorias do processamento anafórico.

## 2.2 Teorias linguísticas sobre o processamento anafórico

As mais variadas propostas linguísticas, apoiadas em diversos quadros teóricos, têm sido criadas, a fim de descrever a maneira como ocorrem as relações entre expressões ou sintagmas nominais (SN), exemplificados de (16) a (20)<sup>19</sup>.

- (16) O João<sub>i</sub> se<sub>i</sub> feriu.
- (17) O João<sub>i</sub> adora-o<sub>j</sub>.
- (18) A Ana<sub>i</sub> falou que ela<sub>j</sub> estava apressada.
- (19) A Ana<sub>i</sub> gostaria que ele<sub>j</sub> chegasse cedo.
- (20) A Ana<sub>i</sub> comprou um lanche e Ø<sub>i</sub> comeu rápido.

---

interact with syntactic, semantic, and pragmatic factor. Consequently it has provided a testing ground for a number of competing hypotheses concerning the relationship between syntax, semantics, and pragmatics in linguistic theory.

<sup>18</sup> Para mais informações, ver Huang (2000).

<sup>19</sup> Os exemplos desta seção são inspirados nos exemplos de Bhat (2004).

As expressões usadas para efetuar uma retomada são, em geral, agrupadas de duas maneiras: (i) expressões linguísticas com potencial para se referir a um objeto do discurso<sup>20</sup>, como *João* e *Ana* em (17) e (18), respectivamente; e, (ii) as expressões que são referencialmente dependentes, estabelecendo um laço de correferência com a entidade à qual ela se refere no contexto linguístico, como *-se*, *-o*, *ela*, *ele* e  $\emptyset$  em (16), (17), (18), (19) e (20), respectivamente.

Dentre as formas correferentes, é possível perceber que o comportamento dos pronomes reflexivos não é o mesmo dos pronomes pessoais e que, ainda há, em Português Brasileiro, assim como em outras línguas latinas, a possibilidade de a expressão correferente ser uma forma pronominal não realizada foneticamente (CHOMSKY, 1981).

Por conta destas particularidades, as mais diversas teorias têm sido criadas visando explicar o que acontece para que cada uma destas formas correferentes seja utilizada em determinado contexto em vez de outra forma possível.

Luegi (2012) apresenta algumas destas teorias subdividindo-as em dois grupos, a saber: (i) teorias linguísticas e (ii) teorias psicolinguísticas. Em nosso estudo, preferimos analisar as teorias baseados em Huang (2000) por considerarmos que a anáfora deve ser estudada estabelecendo correlação entre a sintaxe, a semântica e a pragmática.

Portanto, dividiremos as teorias em três grupos: inicialmente, apresentaremos as teorias apoiadas no estudo da estrutura sintática; em seguida, as teorias voltadas para o estudo semântico das formas correferentes; por fim, será mostrada uma teoria de base pragmática que busca explicar o uso das formas correferenciais.

### **2.2.1 Teorias e hipóteses sintáticas sobre o processamento anafórico**

As teorias e as hipóteses<sup>21</sup> sintáticas que estudam a maneira como nomes e pronomes com função anafórica são ligados a seus referentes partem dos princípios básicos da Teoria da Ligação (CHOMSKY, 1981) como a definição do

---

<sup>20</sup> Ver Bhat (2004).

<sup>21</sup> Chamaremos as hipóteses sintáticas e semânticas do processamento correferencial de teoria ao longo deste projeto de tese.

parâmetro *pro-drop* até estudos mais recentes que buscam entender de que maneira o falante compreende a informação da anáfora durante o processamento linguístico, como a Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995) e a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008).

Nesta seção, revisitaremos algumas destas teorias e hipóteses, a fim de buscar mais informações que possam lançar luz em nosso estudo.

### 2.2.1.1 O parâmetro *pro-drop*

Segundo Chomsky (1981), parâmetro *pro-drop* é o nome dado às línguas naturais que possuem em sua sintaxe a opção de omitir o sujeito da sentença, como ocorre em línguas como o italiano, o espanhol e o português, mas não em línguas como o francês, o inglês e o alemão, conforme exemplificado na sequência de sentenças em (44)<sup>22</sup>.

- (21) a. [*pro*] Sta piovento.  
 b. [*pro*] Está llovento.  
 c. [*pro*] Está chovendo.  
 d. Il pleut.  
 e. It's raining.  
 f. Es regnet.

Ao analisar a estrutura sintática de base das sentenças que compõem o agrupamento (21), podemos perceber que, em (21a), (21b) e (21c), o SN da posição de sujeito está vazio e deve, portanto, segundo o Princípio da Projeção Estendida (CHOMSKY, 1981), ser preenchido por uma das quatro categorias vazias. Já ao observar os exemplos (21d), (21e) e (21f), notamos que, nas sentenças em francês, em inglês e em alemão, respectivamente, a mesma posição é ocupada pelo pronome pessoal *il*, *it* e *es*. A possibilidade de preencher a posição de sujeito em (21a), (21b) e (21c) com a categoria vazia contraposta ao pronome pleno é que nos leva a introduzir, nesta estrutura sintática, a forma *pro*.

---

<sup>22</sup> Os exemplos mostrados ao longo desta seção foram inspirados nos exemplos de Chomsky (1981, p.260) e de Miotto, Silva e Lopes (2013, p.228).

Contudo, faz-se necessário entender que, na estrutura sintática das sentenças em (21), todas as posições argumentais<sup>23</sup> em aberto, ou seja, a serem preenchidas por uma categoria vazia, tem atribuição de caso, mais especificamente, o *pro* expletivo em (21a), (21b) e (21c) e os pronomes expletivos *il* (21d), *it* (21e) e *es* (21f).

Observemos agora o agrupamento de sentenças exemplificado em (22):

- (22) a. [*pro*] Ventou muito forte ontem na minha cidade.  
 b. [*pro*] Dancei demais ontem à noite.  
 c. Eu bebi muito [*pro*].

Precisamos, agora, compreender o que fez surgir *pro* em cada uma das sentenças acima e, assim, sermos capazes de entender que, em (22a), *pro* surge para satisfazer o Princípio da Projeção Estendida, enquanto em (22b), *pro* desempenha a função de sujeito nulo e, em (22c) de objeto nulo.

O fato mais interessante é observar que, em (22b), o aparecimento de *pro* depende, assim como no agrupamento (22), da atribuição de caso e, assim, *pro* passa a depender da Teoria do Caso<sup>24</sup>, sendo atribuído a *pro*, em (22b), o Caso Nominativo, fazendo com que esta categoria vazia seja diferente das demais por desempenhar papel paramétrico nas línguas naturais.

Analisemos o exemplo (23):

- (23) A Maria<sub>i</sub> disse que [*pro*]<sub>i</sub> adorou o almoço de hoje.

Em (23), muito facilmente um falante nativo de português diria que é possível colocar o pronome pleno *ela* sem gerar nenhum tipo de prejuízo sintático à sentença, enquanto um falante nativo de inglês ou francês, por exemplo, provavelmente teria problemas em assimilar esta construção sintática.

Isto acontece porque, em inglês, não se pode apagar o pronome pleno, o que resultaria em uma sentença como *Mary<sub>i</sub> said she<sub>i</sub> loved the lunch today* assim como em francês, em que teríamos a sentença *Maria<sub>i</sub> a dit qu'elle<sub>i</sub> aimait le déjeuner aujourd'hui*.

<sup>23</sup> Chomsky (1981) postula que há três tipos de argumentos: os argumentos verdadeiros, os quasi-argumentos e os não-argumentos. Estes últimos são aqueles que não recebem papel temático, como é o caso dos argumentos expletivos.

<sup>24</sup> Para mais informações sobre a Teoria do Caso, ver Chomsky (1981).

Esta concepção aponta que *pro* é um parâmetro<sup>25</sup> que depende de (i) presença do pronome pleno e (ii) presença do pronome nulo, o que resulta (i) nas línguas não *pro-drop*, como é o caso do francês, do inglês e do alemão e (ii) nas línguas *pro-drop*, como é o caso do italiano, do espanhol e do português.

As línguas naturais podem, segundo Chomsky (1981), *a priori*, ser divididas em dois grupos, conforme disposto no quadro 01.

**Quadro 01** – As características das tipologias das línguas naturais a partir do surgimento ou não da categoria vazia *pro*

Tipo de NP	Tipo de língua
[+Pronome Pleno] [-Pronome Nulo]	Não <i>pro-drop</i>
[-Pronome Pleno] [+Pronome Nulo]	<i>Pro-drop</i>

(Fonte: Os autores)

É preciso entender que, segundo Chomsky (1981), podemos criar o quadro acima pensando nas características do pronome pleno e do pronome nulo para classificar as línguas naturais, entendendo que os critérios [+pronome pleno] [-pronome nulo] diz respeito às línguas que aceitam o uso do pronome pleno, mas excluem, completamente, o surgimento do pronome nulo. Já as características [-pronome pleno] [+pronome nulo] diz respeito às línguas que aceitam o uso tanto do pronome pleno quanto do pronome nulo.

Ao verificar mais atentamente esta nossa proposta de caracterização das línguas naturais, inspirados na categorização dos SNs, de acordo com Chomsky (1981), podemos perceber que, na teoria chomskyana, parece não haver a contraparte às línguas *pro-drop* ou às línguas *não pro-drop*, o que resultaria em uma língua parcialmente *pro-drop* com características [+pronome pleno] [+pronome nulo].

O fato é que, de acordo com o previsto por Chomsky (1981), quando uma língua *pro-drop* é usada, o falante deve evitar o uso do pronome pleno, já que existe, na estrutura profunda, a forma lexical do pronome nulo, enquanto em línguas não *pro-drop*, esta permuta de itens lexicais não acontece porque o falante não dispõe do pronome nulo na estrutura profunda e, assim, conhece apenas o uso do pronome pleno.

<sup>25</sup> Chomsky (1981) postula que a noção de parâmetro diz respeito a uma regra, um padrão, um princípio por meio do qual pode ser estabelecida uma relação de comparação entre termos das línguas sendo, neste caso, o uso do pronome pleno ou do pronome nulo funcionando como argumento a ser atribuído um caso.

Contudo, há línguas que parecem ficar no intermédio deste parâmetro, como é o caso do Português Brasileiro (PB), que dispõe do parâmetro *pro-drop*, mas que tem os falantes, aparentemente, fazendo uso cada vez menos do parâmetro conforme previsto por Chomsky, marcando cada vez mais o sujeito com o pronome pleno por conta de uma possível mudança no paradigma verbal desta língua (TARALLO, 1987; SILVA, 1988; DUARTE, 1996, 2000, 2007; CAVALCANTE, 2008), estando o pronome nulo como a forma de marcação do sujeito nulo de sentenças encaixadas dependendo da noção de c-comando (KATO; NEGRÃO, 2000; KATO; DUARTE, 2014; VERISSIMO, 2017).

Esta possível mudança no parâmetro *pro-drop* faria com que fosse aceita a ideia de que o Português Brasileiro é uma língua parcialmente *pro-drop* (HOLMBERG; NAYUDU; SHEEHAN, 2009). Seguindo esta concepção, poderíamos avançar no estudo do parâmetro *pro-drop*, pois, se tal fenômeno acontece com o PB, muito provavelmente deve acontecer em outras línguas naturais, o que nos levaria a um panorama de três possíveis tipos de línguas naturais a partir do uso do pronome pleno e do pronome nulo, conforme mostrado no quadro 02.

**Quadro 02** – As características das tipologias das línguas naturais a partir do surgimento ou não da categoria vazia *pro* introduzindo a possibilidade de línguas parcialmente *pro-drop*.

Tipo de NP	Tipo de língua
[+Pronome Pleno] [-Pronome Nulo]	Não <i>pro-drop</i>
[+Pronome Pleno] [+Pronome Nulo]	Parcialmente <i>pro-drop</i>
[-Pronome Pleno] [+Pronome Nulo]	<i>Pro-drop</i>

(Fonte: Os autores)

Neste contexto, teríamos, portanto, uma língua que dispunha de pronomes plenos e pronomes nulos podendo ocupar a mesma posição argumental e que poderiam ser usados ao acaso. Seria, portanto, necessário explicar como ocorre a distribuição destes pronomes e se há restrições de uso, mesmo em línguas nas quais estes itens lexicais parecem intercambiáveis<sup>26</sup>.

O fato é que as categorias vazias *pro* e PRO, apesar de surgirem como categorias vazias resultantes não dos movimentos de SNs, mas como formas

<sup>26</sup> Estudos que buscam observar a distribuição dos pronomes plenos e dos pronomes nulos já foram desenvolvidos em italiano (CARMINATI, 2002) e em Português Brasileiro (TEIXEIRA, 2013), mas são necessários mais estudos, a fim de mais comprovações empíricas sobre como os falantes realmente fazem uso destes itens em línguas *pro-drop*.

vazias que são o contraponto de anáforas e de pronomes plenos, conforme afirma Chomsky (1981), apresentam características bem marcadas que os diferenciam e, assim, leva a estudos sintáticos distintos.

Por ser objeto deste estudo, interessa-nos tratar também sobre *pro* no PB a partir das características atribuídas a esta categoria, conforme Chomsky (1981), pois é de nosso interesse atentar para como ocorre a correferência anafórica. Entretanto, é de nosso interesse analisar os pronomes pessoais de terceira pessoa do singular, o que Chomsky define como pronomes, apenas, e não como anáfora, já que em sua teoria, somente os reflexivos e os recíprocos contêm características de anáfora.

Ao atentar para este fato, recorreremos à classificação de Huang (2000) e passamos a entender a anáfora<sup>27</sup> como o termo que “carrega de volta” no sentido de estabelecer correferência a um termo antecedente e chamamos, portanto de anáfora, os pronomes pessoais de terceira pessoa, pautados também na classificação dada aos pronomes por Benveniste (2005) e Levinson (2007), pois esta concepção tem sido a mais trabalhada em estudos linguísticos e psicolinguísticos a partir da Teoria da Centralização e que nos é de interesse como aparato teórico em nosso trabalho.

Revisitaremos, a seguir, estudos posteriores à publicação da Teoria da Ligação (CHOMSKY, 1981), que têm trazido informações importantes sobre como as anáforas e os pronomes são processados durante tarefas linguísticas.

### *2.2.1.2 Teoria da Centralização*

A Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995) foi desenvolvida no âmbito da investigação em inteligência artificial, a fim de relacionar foco de atenção e a escolha da expressão referencial para explicar a coerência de um texto, apontando que, no texto, há (i) a coerência local – fator que mantém a coerência dentro do segmento discursivo, e (ii) a coerência global – fator que mantém a coerência do texto todo.

---

<sup>27</sup> Fazemos uma distinção ao que Chomsky (1981) chama de anáfora, por enquadrar nesta classificação somente os pronomes reflexivos e, por isso, tratamos do pronome da terceira pessoa como anáfora, como fazem Huang (2000), Benveniste (2005) e Levinson (2007). É neste conceito de anáfora que estão feitas as pesquisas em psicolinguística, mas sem abandonar os conceitos de ligação dos pronomes, conforme previsto na Teoria da Ligação (CHOMSKY, 1981).

De acordo com Grosz e Sidner (1986), para compreender o funcionamento de um segmento discursivo, é preciso atentar para a maneira com as partes estão em correlação. Para estes autores, o segmento discursivo pode é composto pela junção de três elementos, sendo eles: (i) a estrutura linguística, (ii) a estrutura intencional e (iii) o foco de atenção.

O primeiro componente, segundo Grosz e Sinder (1986), é a estrutura linguística, ou seja, a organização sequencial dos enunciados que formam o discurso. É a partir da estrutura linguística podem ser notadas as ligações entre os segmentos, já que “as declarações em um segmento, como as palavras em uma frase, desempenham funções específicas tanto dentro do próprio segmento quanto do todo textual”<sup>28</sup> (GROSZ; SIDNER, 1986, p.177, tradução nossa, adaptado).

Para Grosz e Sinder (1986), o segundo componente do segmento discursivo, a estrutura intencional, serve para mostrar os objetivos dos envolvidos no discurso, pois, a depender da intenção dada a determinados segmentos, o falante é capaz de torná-los coerentes ou incoerentes.

Já o terceiro elemento, o foco de atenção, serve para, segundo Grosz e Sidner (1986), guiar os participantes durante a compreensão do desenrolar do segmento discursivo, sendo, de acordo com os autores, uma propriedade do texto, não do discurso dos participantes.

Com relação ao foco de atenção, Grosz e Sidner (1986) defendem ser um elemento dinâmico e em constante mudança ao longo do segmento discursivo, já que serve para relacionar os referentes e as relações entre os segmentos, levando os participantes a definirem o que é ou não saliente em cada parte do discurso.

Ao levar em consideração os três componentes do segmento discursivo, Grosz e Sinder (1986) afirmam que a centralização é um processo dinâmico, mas que diz respeito a locais do segmento discursivo como, por exemplo, nas regiões de retomadas de elementos já mencionados ao longo do texto.

A fim de aprofundar os conceitos criados por Grosz e Sidner (1986), Grosz, Joshi e Weinstein (1995) mostraram que, durante a construção de um texto, o falante/escritor busca opções para preencher a posição na qual usará uma expressão anafórica para criar coerência local e, conseqüentemente,

---

<sup>28</sup> The utterances in a segment, like the words in a phrase, serve particular roles with respect that segment [...] or the overall discourse.

coerência global. Se ocorrerem casos nos quais o preenchimento do local da anáfora se der de forma inequívoca, o foco de atenção do texto será afetado, como exemplificado por Grosz, Joshi e Weinstein (1995, p.206, tradução nossa) em (24)<sup>29</sup> e (25)<sup>30</sup>.

- (24) a. João foi à sua loja de música favorita para comprar um piano.  
 b. Ele frequenta a loja há anos.  
 c. Ele estava muito feliz por ele finalmente poder comprar um piano.  
 d. Ele chegou à loja que estava fechada naquele dia.
- (25) a. João foi à sua loja de música favorita para comprar um piano.  
 b. Era a loja que João tem frequentado há anos.  
 c. Ele estava muito feliz por ele finalmente poder comprar um piano.  
 d. Ela estava fechada quando João chegou.

De acordo com os exemplos acima, Grosz, Joshi e Weinstein (1995) explicam que o texto (24) é mais coerente do que (25) pelo fato de haver apenas uma entidade discursiva sendo retomada ao longo do texto, o que mantém o foco de atenção na mesma entidade *João*, enquanto no texto (25) há duas entidades concorrentes para serem retomadas pelo pronome que funciona como expressão anafórica *João* e *loja*, fazendo com que o foco de atenção seja mudado ao longo da continuação do texto e, assim, torna-se menos coerente do que o texto anterior.

Levando em consideração a mudança no foco de atenção em determinada região do segmento discursivo, Grosz, Joshi e Weinstein (1995) passam a chamar de centro do discurso as formas anafóricas utilizadas para ligar as entidades de um texto, sendo que o centro discursivo pode tanto remeter a uma informação que ainda será dada no texto (centro catafórico) quanto a uma informação já dada no texto (centro anafórico).

De acordo com Grosz, Joshi e Weinstein (1995), para que se crie um centro anafórico no texto, é preciso que seja introduzido apenas um termo que

---

<sup>29</sup> (24) a. John went to his favorite music store to buy a piano.  
 b. He had frequented the store for many years.  
 c. He was excited that he could finally buy a piano.  
 d. He arrived just as the store was closing for day.

<sup>30</sup> (25) a. John went to his favorite music store to buy a piano.  
 b. It was a store John had frequented for many years.  
 c. He was excited that he could finally buy a piano.  
 d. It was closing just as John arrived.

possa indiciar tal centro e que ocorra, preferencialmente, na mesma posição sintática do termo previamente introduzido, como em (25).

É postulada, então, a Regra 1 da Teoria da Centralização, afirmando que “se um elemento de um centro anafórico (de dado segmento) é realizado por um pronome, então, no(s) próximo(s) centro(s) discursivo(s), o(s) centro(s) anafórico(s) devem ser também realizados por pronome(s)”<sup>31</sup> (GROSZ; JOSHI; WEINSTEIN, 1995, p.214, tradução nossa, adaptado).

Entretanto, ao ser criado mais de um centro anafórico, podemos notar incoerência a depender da posição sintática a ser retomada (posição de sujeito e/ou posição de objeto), sendo que, ao usar um pronome como expressão anafórica, este deve preferencialmente ser usado para preencher a posição de sujeito, a fim de tornar o texto mais coerente, enquanto tende-se a preencher o centro anafórico na posição de objeto com a repetição do nome, como em (26)<sup>32</sup>, exemplo retirado de Grosz, Joshi e Weinstein (1995, p.211, tradução nossa).

- (26) a. Susan deu um hamster para Betsy.  
b. Ela se lembrou de que os hamsters são bastante tímidos.

Os autores mostram que em discursos subsequentes, o *status* de equivalência do centro anafórico não se torna claro, sobretudo se for usado um pronome que funcione como correferente anafórico, possível de executar a retomada tanto a posição de sujeito quanto a posição de objeto, como em (27), (28), (29) e (30), exemplificado por Grosz, Joshi e Weinstein (1995, p.211-212, tradução nossa).

- (27) a. Susan deu um hamster para Betsy.  
b. Ela se lembrou de que os hamsters são bastante tímidos.  
c. Ela perguntou para Betsy se ela tinha gostado do presente<sup>33</sup>.
- (28) a. Susan deu um hamster para Betsy.  
b. Ela se lembrou de que os hamsters são bastante tímidos.

---

<sup>31</sup> Rule 1

If any element of  $C_i(U_n)$  is realized by a pronoun in  $U_{n+1}$ , then the  $C_b(U_{n+1})$  must be realized by a pronoun also.

<sup>32</sup> (26) a. Susan gave Betsy a pet hamster.  
b. She reminded her that such hamsters were quite shy.

<sup>33</sup> (27) a. Susan gave Betsy a pet hamster.  
b. She reminded her that such hamster were quite shy.  
c. She asked Betsy whether she liked the gift.

c. Betsy disse para ela que ela tinha adorado o presente<sup>34</sup>.

- (29) a. Susan deu um hamster para Betsy.  
b. Ela se lembrou de que os hamsters são bastante tímidos.  
c. Susan perguntou para ela se ela tinha adorado o presente<sup>35</sup>.

- (30) a. Susan deu um hamster para Betsy.  
b. Ela se lembrou de que os hamsters são bastante tímidos.  
c. Ela disse para Susan que ela tinha adorado o presente<sup>36</sup>.

Grosz, Joshi e Weinstein (1995, p.212, traduções nossas) alertam que “o problema não é meramente a mudança de um pronome por um nome próprio”, tampouco pode ser “atribuído somente à mudança da posição gramatical de sujeito e objeto”<sup>37</sup>, o que pode ser entendido pelo fato de que a pronominalização na posição de sujeito nos exemplos de (26) a (30) é viável, enquanto na posição de objeto há uma tendência à incoerência local.

Por fim, Grosz, Joshi e Weinstein (1995, p.214, tradução nossa) afirmam que “o uso de um pronome no centro anafórico sinaliza ao ouvinte que o falante continua a falar da mesma coisa”<sup>38</sup>, mantendo, assim, o foco de atenção na mesma entidade a ser retomada no texto por meio de um pronome, preferencialmente na posição de sujeito.

### 2.2.1.3 Penalidade do Nome Repetido

Seguindo os conceitos propostos na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), Gordon, Grosz e Guilliom (1993) criaram um *desing*<sup>39</sup> para uma pesquisa experimental a ser executado

<sup>34</sup> (28) a. Susan gave Betsy a pet hamster.  
b. She reminded her that such hamster were quite shy.  
c. Betsy told her that she really liked the gift.

<sup>35</sup> (29) a. Susan gave Betsy a pet hamster.  
b. She reminded her that such hamster were quite shy.  
c. Susan asked her whether she liked the gift.

<sup>36</sup> (30) a. Susan gave Betsy a pet hamster.  
b. She reminded her that such hamster were quite shy.  
c. She told Susan that she really liked the gift.

<sup>37</sup> The problem is not merely a change from a pronoun back to a proper name (...) it also cannot be attributed solely to a change from grammatical subject to grammatical object position.

<sup>38</sup> The use of a pronoun (...) signal the hearer that the speaker is continuing to talk about the same thing.

<sup>39</sup> Denomina-se de *design* experimental o(s) grupo(s) de sentenças construídas para serem lidas ou escutadas em pesquisas psicolinguísticas feitas em laboratório por meio de métodos experimentais.

por meio de uma leitura automonitorada<sup>40</sup>, a fim de testar a proeminência sintática e os custos de processamento de nomes repetidos e pronomes plenos durante a leitura. O principal objetivo do estudo desenvolvido por Gordon, Grosz e Guilliom (1993) era comprovar se o efeito proposto por Grosz, Joshi e Weinstein (1995) ao postular Regra 1 da Teoria da Centralização poderia ser percebido nos resultados de seu experimento.

O *desing* experimental criado por Grosz, Joshi e Weinstein (1995) buscava, portanto, evidências de que os pronomes são usados para preencher os centros anafóricos na posição de sujeito da sentença e que, quando comparado com o tempo de leitura de nomes sendo usados para preencher o mesmo local da sentença, fosse percebida diferença de tempo de processamento da informação entre estas duas formas linguísticas.

Caso encontrassem tais resultados, Grosz, Joshi e Weinstein (1995) esperavam corroborar a tese de que os pronomes são elos linguísticos que mantêm a coerência local, sendo, conseqüentemente, fatores decisivos para manutenção da coerência global do texto por serem entendidos como veículos naturais para a manutenção de uma entidade discursiva previamente introduzida.

Segundo Gordon, Grosz e Guilliom (1993), na Regra 1 da Teoria da Centralização, Grosz e Sinder (1986) descrevem

os pronomes como mecanismos linguísticos que indicam a continuidade e a coerência. Parte da informação fornecida pelo uso de um pronome da terceira pessoa é um link para o discurso anterior. Portanto, os pronomes devem ser interpretados em relação às informações já apresentadas no discurso. Por conta disso, os pronomes são um mecanismo natural para o estabelecimento da coerência (GORDON; GROSZ; GUILLIOM, 1993. p.313-314, tradução nossa, adaptado)<sup>41</sup>.

Para testar a Regra 1 da Teoria da Centralização (GROSZ; SIDNER, 1986), Gordon, Grosz e Guilliom (1993) compararam, em seu *design* experimental, as retomadas anafóricas por meio de três casos: (i) pronome/pronome – dois referentes sendo previamente introduzidos, um na posição de sujeito e outro na posição de objeto, e, nos segmentos subsequentes,

<sup>40</sup> Para mais informações, ver Straub e Rayner (2007)

<sup>41</sup> [...] pronouns are linguistic mechanism for indicate continuity and coherence. Part of the information conveyed by the use of the third-person pronoun is a link to previous discourse. Pronouns must be interpreted in relation to information that has already been presented in a discourse. Thus, they provide a natural mechanism for coherence.

ambos os referentes sendo substituídos por pronome pleno, como em (31); (ii) pronome/nome – dois referentes sendo previamente introduzidos, um na posição de sujeito e outro na posição de objeto, mas somente o referente na posição de sujeito sendo substituído por pronome pleno, enquanto o referente na posição de objeto foi retomado pela repetição do mesmo nome usado em sua introdução, isto é, sendo usado o nome repetido, como em (32); e, (iii) nome/nome – dois referentes sendo previamente introduzidos, um na posição de sujeito e outro na posição de objeto e ambos os referentes sendo retomados por nome repetido, como em (33). Os exemplos de (31) a (33) podem ser encontrados no trabalho de Gordon, Grosz e Guilliom (1993, p.317-318, tradução nossa, grifos dos autores).

- (31) a. *Bruno* era o valentão do bairro.  
 b. *Ele* perseguiu *Tommy* por todo o caminho de casa para a escola um dia.  
 c. *Ele* o observou escondido atrás de uma grande árvore e começou a chorar.  
 d. *Ele* gritou tão alto com *ele* que todos os vizinhos saíram de casa<sup>42</sup>.

- (32) a. *Bruno* era o valentão do bairro.  
 b. *Ele* perseguiu *Tommy* por todo o caminho de casa para a escola um dia.  
 c. *Ele* observou *Tommy* escondido atrás de uma grande árvore e começou a chorar.  
 d. *Ele* gritou tão alto com *Tommy* que todos os vizinhos saíram de casa<sup>43</sup>.

- (33) a. *Bruno* era o valentão do bairro.  
 b. *Bruno* perseguiu *Tommy* por todo o caminho de casa para a escola um dia.  
 c. *Bruno* observou *Tommy* escondido atrás de uma grande árvore e começou a chorar.  
 d. *Bruno* gritou tão alto com *Tommy* que todos os vizinhos saíram de casa<sup>44</sup>.

---

<sup>42</sup> (31) Pro-Pro

- a. *Bruno* was the bully of the neighborhood.  
 b. *He* chased *Tommy* all the way home from school one day.  
 c. *He* watched *him* hide behind a big tree and start to cry.  
 d. *He* yelled at *him* só loudly that all the neighbors came outside.

<sup>43</sup> (32) Pro-Name

- a. *Bruno* was the bully of the neighborhood.  
 b. *He* chased *Tommy* all the way home from school one day.  
 c. *He* watched *Tommy* hide behind a big tree and start to cry.  
 d. *He* yelled at *Tommy* só loudly that all the neighbors came outside.

<sup>44</sup> (33) Name-Name

1. *Bruno* was the bully of the neighborhood.  
 2. *Bruno* chased *Tommy* all the way home from school one day.  
 3. *Bruno* watched *Tommy* hide behind a big tree and start to cry.  
 4. *Bruno* yelled at *Tommy* só loudly that all the neighbors came outside.

Neste experimento, Gordon, Grosz e Guilliom (1993) atentam para o fato de haver, segundo a Teoria da Centralização (GROSZ; SIDNER, 1986), uma regra postulando que:

A condição pronome/nome testará diretamente a Regra 1 da Teoria da Centralização. Uma vez que a Regra 1 é baseada na proclamação que o centro anafórico faz como importante link de manutenção da coerência do enunciado e na conjectura que, tal como, deve ser realizado por um pronome porque a forma é inerente à construção da coerência (GORDON; GROSZ; GUILLIOM, 1993, p.318, tradução nossa)<sup>45</sup>.

Após testarem seu experimento, Gordon, Grosz e Guilliom (1993) concluíram que, quando o leitor passou de uma sentença para outra durante a leitura dos segmentos discursivos, o nome repetido demorou mais tempo para ser processado do que o pronome na mesma condição. Este resultado aponta para o fato de o nome repetido demandar maior custo cognitivo durante o processamento anafórico quando comparado a um pronome pleno desempenhando a mesma função, levando Gordon, Grosz e Guilliom (1993) a nomear este efeito de Penalidade do Nome Repetido<sup>46</sup>.

O efeito de penalidade entre o nome repetido e o pronome pleno, conforme postulado por Gordon, Grosz e Guilliom (1993), levou Gerlomini-Lezama (2008) a comparar o mesmo efeito poderia ser percebido em Espanhol Argentino, já que esta é uma língua *pro-drop*, enquanto os resultados que levaram ao postulado da Penalidade do Pronome Pleno foram encontrados em inglês, uma língua não *pro-drop*.

#### 2.2.1.4 Penalidade do Pronome Pleno

Gerlomini-Lezama (2008) desenvolveu, em Espanhol Argentino, um *design* linguístico inspirado no que fora criado por Gordon, Grosz e Guilliom (1993), a fim de averiguar se o efeito da Penalidade do Nome Repetido pode ser percebido em uma língua *pro-drop*.

---

<sup>45</sup> Rule 1 is based on the claim that the C<sub>b</sub> is an important coherence link back to the previous utterance and on the conjecture that, as such, it should be realized with a pronoun because that form is an inherent cue to coherence.

<sup>46</sup> Repeated Name Penalty (GORDON; GROSZ; GUILLIOM, 1993).

O principal objetivo de Gerlomini-Lezama (2008) era observar como ocorre o preenchimento de um centro anafórico quando o falante tem, a seu dispor, como opções linguísticas, o nome repetido, o pronome pleno e o pronome nulo. Para tal, o autor, seguindo os preceitos da Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), que também nortearam a pesquisa que levou à postulação do efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), levou em consideração a estrutura linguística e o foco de atenção como fatores que, junto da escolha da expressão anafórica, constroem a coerência de um segmento discursivo.

Para observar tal fenômeno, Gerlomini-Lezama (2008) criou um *desing* experimental, a ser lido como uma tarefa de leitura automonitorada, constituído por duas sentenças postas lado a lado sendo que, na primeira sentença, eram introduzidos dois referentes de gêneros distintos e, na segunda sentença, o referente que ocupava a posição sintática de sujeito era retomado por meio de (i) nome repetido, como em (34), (ii) pronome pleno, como em (35) ou (iii) pronome nulo, como em (36), conforme exemplificado por Gerlomini-Lezama (2008, p.19, tradução nossa).

(34) João<sub>i</sub> se encontrou com Maria<sub>j</sub>. João<sub>i</sub> a<sub>j</sub> viu triste<sup>47</sup>.

(35) João<sub>i</sub> se encontrou com Maria<sub>j</sub>. Ele<sub>i</sub> a<sub>j</sub> viu triste<sup>48</sup>.

(36) João<sub>i</sub> se encontrou com Maria<sub>j</sub>. Ø<sub>i</sub> A<sub>j</sub> viu triste<sup>49</sup>.

Como resultado deste experimento, Gerlomini-Lezama (2008) encontrou que, as sentenças nas quais o uso do pronome nulo retoma a entidade previamente introduzida, houve menor custo cognitivo durante o processamento do que quando a retomada é feita usando tanto o nome repetido quanto o pronome pleno.

Gerlomini-Lezama (2008) afirma que, assim como ocorreu na pesquisa desenvolvida por Gordon, Grosz e Guilliom (1993) ao postularem o efeito da Penalidade do Nome Repetido, o nome repetido foi a forma mais custosa de ser processada quando levando em consideração o tempo de leitura. Este fato o levou a afirmar que “a PNR se estende ao espanhol e que esse fenômeno por

<sup>47</sup> Juan<sub>i</sub> se encontró con Maria<sub>j</sub>. Juan<sub>i</sub> la<sub>j</sub> vio triste.

<sup>48</sup> Juan<sub>i</sub> se encontró con Maria<sub>j</sub>. El<sub>i</sub> la<sub>j</sub> vio triste.

<sup>49</sup> Juan<sub>i</sub> se encontró con Maria<sub>j</sub>. Ø<sub>i</sub> La<sub>j</sub> vio triste.

muito bem ser o reflexo de um universal linguístico” (GERLOMINI-LEZAMA, 2008, p.23, tradução nossa)<sup>50</sup>.

Além disso, pelo fato de os resultados do experimento desenvolvido por Gerlomini-Lezama (2008) terem mostrado que o pronome nulo, no contexto estudado, é a forma mais facilmente processada quando comparado ao tempo de processamento do pronome pleno, foi postulado, pelo autor, um efeito denominado de Penalidade do Pronome Pleno<sup>51</sup>.

O fenômeno da Penalidade do Pronome Pleno indica a leitura mais lenta de sentenças em que há pronomes plenos se referindo a antecedentes salientes, quando comparados à mesma condição contendo pronomes nulos. Esse fenômeno pode ser refletido como um desequilíbrio entre custo e função porque, nesse contexto, as características semânticas do pronome pleno não contribuem substancialmente para a coerência do discurso. Nesses casos, sentenças com pronome nulo são processadas mais rapidamente porque a expressão reduzida é eficiente o bastante para que seja identificado o referente em questão (GERLOMINI-LEZAMA, 2008, p. 59, tradução nossa)<sup>52</sup>.

Percebemos, portanto, que, quando o falante tem, a seu dispor, diversas opções linguísticas que podem funcionar como anáfora, é interessante verificar o possível efeito de suas contrapartes, conforme Chomsky (1981), pois diversos efeitos podem ser encontrados em uma mesma língua, como afirma Gerlomini-Lezama (2008) ao mostrar que, em Espanhol Argentino, tanto há o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GILLIOM, 1993) quanto o efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008).

Como ambos os efeitos acima foram encontrados a partir do que é postulado na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), outros efeitos de ordem linguística devem ser levados em consideração por pesquisadores que desenvolvem estudos a partir desta teoria. Esta necessidade se dá pelo fato de tanto Grosz e Sidner (1986) quanto Grosz, Joshi e Weinstein (1995) falarem sobre (i) as entidades estarem ou não em posição de proeminência sintática, além de (ii) induzirem à observação de o

---

<sup>50</sup> The RNP extends to Spanish and that this phenomenon may well be a reflection of a linguistic universal.

<sup>51</sup> Overt Pronoun Penalty (GERLOMINI-LEZAMA, 2008)

<sup>52</sup> The Overt Pronoun Penalty phenomenon, that is, the slower reading of sentences containing overt pronouns referring to salient antecedents, relative to comparative sentences with null pronouns. The phenomenon can be interpreted as reflecting an imbalance between cost and function because in OPP context the additional semantic features of the overt pronoun do not make a substantive contribution to the coherence of the discourse. In such cases, sentences with a null pronoun are processed faster because this reduced expression is efficient enough to identify the referent in question.

sujeito anafórico retomar preferencialmente uma entidade também na posição de sujeito, fazendo emergir o possível efeito gerado pelo paralelismo sintático entre as sentenças.

### 2.2.1.5 *Proeminência Sintática*

Com vista a analisar o padrão linguístico durante uma tarefa de processamento correferencial<sup>53</sup>, Gordon e Hendrick (1997) desenvolveram uma série de testes de julgamento de gramaticalidade de sentenças e de segmentos discursivos, a fim de averiguar os princípios da Teoria da Ligação (CHOMSKY, 1981) com relação ao nível intrassentencial e o conceito de centro discursivo da Teoria da Centralização (GORDON, GROSZ e GILLION, 1993) com relação ao nível interssentencial.

De acordo com postulados da Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), o pronome deve preencher o local de centro anafórico quando retoma uma entidade proeminente. Contudo, a noção de proeminência sintática não é bem esclarecida dentro dos postulados desta teoria, o que fez com que, no nível intrassentencial, Gordon e Hendrick (1997) buscassem explicações sobre como os sujeitos julgam sentenças nas quais são aplicados os princípios da Teoria da Ligação e, além de observarem como as entidades previamente introduzidas tendem a ser retomadas em determinada posição sintática, a fim de desenvolverem o conceito de proeminência sintática.

Em uma série de cinco experimentos, todos de teste de julgamento de gramaticalidade, os pesquisadores buscaram explicações sobre (i) a precisão do Princípio C da Teoria da Ligação no primeiro experimento, (ii) a relação de c-comando entre um pronome e um nome no segundo experimento, (iii) a relação existente entre as anáforas e as posições sintáticas em que elas devem ocorrer quando há correferência no terceiro experimento, (iv) a variação da entidade em c-comando no quarto experimento com vistas a apoiar os resultados primeiro experimento e (v) a posição sintática do pronome e do nome em segmentos

---

<sup>53</sup> A correferência linguística é um mecanismo por meio do qual duas formas linguísticas se referem à mesma entidade semântica (GORDON; HENDRICK, 1997, p.336, tradução nossa) (Linguistic co-reference is the mechanism by which two linguistic forms refer to the same semantic entity).

discursivos no quinto experimento com vistas a apoiar os resultados do terceiro experimento.

Todos os experimentos envolviam a possível correferência entre um nome e um pronome e seguia os padrões já discutidos nas teorias nas quais Gordon e Hendrick (1997) usaram de apoio para esta pesquisa e, assim, foram desenvolvidos, *grosso modo*, experimentos que buscavam explicar quando os nomes e os pronomes estão ou não em uma relação de c-comando, além de averiguar como os nomes e pronomes podem ser distribuídos ao longo de um segmento discursivo<sup>54</sup>.

Para nosso estudo, interessa-nos o segundo experimento, mostrado a seguir, de acordo com o que fora proposto por Gordon e Hendrick (1997, p.340, tradução nossa). O interesse neste experimento específico se dá pelo fato de que o objetivo principal era observar de que forma as construções gramaticais permitem a concretização da correferência entre nomes e pronomes. Para averiguar este efeito, Gordon e Hendrick (1997) examinaram a aceitabilidade da correferência entre um nome e um pronome não c-comandado quando o pronome ocorre em uma oração anteposta.

**Quadro 03** – Estímulos do segundo experimento<sup>55</sup>

NP <sub>i</sub>	NP <sub>j</sub>	C-comanda	Sentença experimental
Nome	Pronome	Não	Antes de a Susan começar a cantar, ela se levantou.
Nome	Pronome	Sim	Susan se levantou antes de ela começar a cantar.
Nome	Nome	Não	Antes de a Susan começar a cantar, a Susan se levantou.
Nome	Nome	Sim	Susan se levantou antes de a Susan começar a cantar.
Pronome	Nome	Não	Antes de ela começar a cantar, a Susan se levantou.
Pronome	Nome	Sim	Ela se levantou antes de a Susan começar a cantar.

(Fonte: Gordon; Hendrick, 1997)

De maneira geral, Gordon e Hendrick (1997) afirmam que, com relação ao processamento da correferência no nível intrassentencial, quando um nome é retomado por um pronome, o processamento é mais rápido do que quando um

<sup>54</sup> Para mais informações, ver Gordon e Hendrick (1997).

<sup>55</sup>

NP <sub>1</sub>	NP <sub>2</sub>	C-command	Sample sentence
Name	Pronoun	No	Before Susan began to sing, she stood up.
Name	Pronoun	Yes	Susan stood up before she began to sing.
Name	Name	No	Before Susan began to sing, Susan stood up.
Name	Name	Yes	Susan stood up before Susan began to sing.
Pronoun	Name	No	Before she began to sing, Susan stood up.
Pronoun	Name	Yes	She stood up before Susan began to sing.

nome retoma este mesmo nome, o que mostra que o tipo de SN interfere durante uma tarefa linguística de teste de julgamento de gramaticalidade de sentenças.

Além disso, os pesquisadores afirmam que a relação de c-comando entre os SNs reduz a aceitabilidade de correferência entre nomes repetidos, sendo mais propícia a retomada anafórica por meio de um pronome quando o antecedente está na posição de sujeito da sentença matriz.

Por fim, no que diz respeito ao processamento correferencial no nível intrassentencial, Gordon e Hendrick (1997) afirmam que o processamento das formas anafóricas da esquerda para a direita, conforme proposto por Chomsky (1981) não é preciso, pois nos casos em que o pronome surge antes do nome, a correferência tende a não ser aceita, o que aponta para uma possível não aplicação do Princípio C.

Desta forma, Gordon e Hendrick (1997) concluem que os resultados dos experimentos por eles desenvolvidos são consistentes com os postulados do Princípio A e do Princípio B da Teoria da Ligação, mas não totalmente concordantes com os ditames do Princípio C desta teoria.

No que diz respeito à proeminência sintática, conceito pré-concebido na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995) Gordon e Hendrick (1997) afirmam que a posição sintática de sujeito ou de objeto da entidade previamente introduzida tende a desempenhar algum efeito sobre a região do c-comando quando do processamento correferencial.

A proeminência linguística é inversamente relacionada à sentença encaixada. Isto significa dizer que os antecedentes do c-comando tenderão a ter maior destaque como constituintes de sujeito em relação a constituintes de objeto. (GORDON e HENDRICK, 1997, p. 362, tradução nossa)<sup>56</sup>.

De acordo com os autores, a concepção da proeminência sintática como a preferência de retomada de entidades na posição de sujeito com relação a entidades na posição de objeto leva ao processamento mais rápido de um pronome na posição de sujeito que retoma um nome também na posição de sujeito do que quando um pronome retoma um nome na posição de objeto, contribuindo, assim, para a ordenação das entidades durante o discurso.

---

<sup>56</sup> Linguistic prominence is inversely related to depth of embeddedness. This means that c-commanding antecedents will tend to have greater prominence, as will constituents of subjects relative to comparable constituents of objects.

A proeminência afeta a construção das representações do discurso de duas maneiras: (1) Influencia a ordem dos gatilhos de processamento, com representações sintáticas mais proeminentes, levando ao processamento anterior dos gatilhos do que as representações menos proeminentes; (2) Contribui para a ordenação das entidades mencionadas na representação do discurso, e essa ordenação influencia a eficiência com a qual as regras de construção para o estabelecimento da correferência operam. (Esse segundo efeito de destaque é equivalente à noção na Teoria de Centralização de que a representação de uma expressão em discurso coerente inclui um conjunto ordenado de centros anafóricos) (GORDON; HENDRICK, 1997, p.362, tradução nossa)<sup>57</sup>.

Estes resultados apontam que, durante o processamento correferencial, os falantes tendem a distinguir o núcleo gramatical como um dos fatores responsáveis pela distribuição das expressões correferenciais tanto no nível intrassentencial quanto no nível interssentencial.

#### 2.2.1.6 Paralelismo Sintático

A fim de investigar os efeitos gerados pela estrutura sintática dos enunciados e observar como ocorre o processamento correferencial, Chambers e Smith (1998) desenvolveram três experimentos de leitura analisados através do paradigma do rastreamento ocular tomando como base estudos anteriores sobre o Paralelismo Estrutural<sup>58</sup> e o que é dito pelos autores da Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995) sobre o processamento de nomes repetidos e de pronomes.

Levando em consideração que a estrutura do segmento discursivo tem influência sobre a interpretação feita sobre o que é lido, Chambers e Smith (1998) tentam identificar fatores que tornam uma sequência discursiva coerente. Entre os fatores considerados, os autores aceitavam que os pronomes devem ser entendidos como as melhores expressões correferentes a uma entidade proeminente, enquanto o uso de uma expressão correferente mais informativa, como é o caso dos nomes, torna a interpretação mais complexa (GORDON, GROSZ e GILLIOM, 1993).

---

<sup>57</sup> Prominence affects the construction of discourse representations in two ways: (1) It influences the order of processing triggers, with more prominent syntactic representations leading to earlier processing of triggers than less prominent representations; (2) It contributes to an ordering of the entities referred to in the discourse representations, and this ordering influences the efficiency with which the construction rules establishing co-reference operate. (This second effect of prominence is equivalent to the notion in centering theory that the representation of an utterance in coherent discourse includes an ordered set of forward-looking centers).

<sup>58</sup> Para mais informações sobre os estudos anteriores que deram origem a este trabalho, ver Chambers e Smith (1998).

De acordo com Chambers e Smith (1998), isto ocorre por conta da distribuição das expressões correferenciais dentro de um enunciado, já que certas posições sintáticas são mais frequentemente reservadas para entidades mais salientes, criando, assim, a coerência discursiva.

É proposto na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995) que a coerência discursiva está relacionada ao uso de uma anáfora restrita ao termo com o qual estabelece correferência, tendo de ambos ocorrerem nas sentenças nas mesmas posições sintáticas. Além disso, foi mostrado por Grosz, Joshi e Weinstein que pronomes devem ocupar, preferencialmente, o centro anafórico na posição de sujeito quando retomando um termo antecedente na mesma posição.

Para Chambers e Smith (1998), embora o estabelecimento da ligação entre um termo antecedente e sua forma correferencial de fato influenciem na interpretação do pronome, não é somente por questões de centralização que a correferência é resolvida.

Em sentenças como (37), exemplificadas em Chambers e Smith (1998, p.594, tradução nossa), é mostrada a chamada função paralela por meio da qual se pressupõem que em (37a) o pronome *ele*, que está na posição de sujeito da segunda sentença, retoma o nome *John* que está na mesma posição da sentença matriz, assim como ocorre em (37b), já que o pronome oblíquo *o* retoma o nome *Paul* que se encontra na mesma posição sintática na sentença anterior.

- (37) a. John<sub>i</sub> criticou Paul<sub>j</sub> e então ele<sub>i</sub> insultou Marie<sub>k</sub><sup>59</sup>.  
 b. John<sub>i</sub> criticou Paul<sub>j</sub> e Maria<sub>k</sub> insultou-o<sub>j</sub>.

De acordo com Chambers e Smith (1998), a interpretação feita em (37a) e em (37b) é facilitado pela semelhança estrutural entre as sentenças, o que pode ser dificultado quando esta semelhança é menor como em (38), exemplificado em Chambers e Smith (1998, p.594, tradução nossa), o que não é explicado pela Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995).

---

<sup>59</sup> (37) a. John criticized Paul and the he insulted Marie. (he = John)  
 b. John criticized Paul and Marie insulted him. (him = Paul)

(38) John<sub>i</sub> criticou Paul<sub>j</sub> e, em seguida, Marie<sub>k</sub> perguntou se ele<sub>ij</sub> iria sair<sup>60</sup>.

O primeiro experimento desenvolvido por Chambers e Smith (1998, p.597, tradução nossa), exemplificado em (70)<sup>61</sup>, neste trabalho visava avaliar se os efeitos de paralelismo sintático são atuantes durante o processamento do pronome que retoma ou não o tópico discursivo e analisar se os resultados são consistentes ou não com as previsões propostas pela Teoria da Centralização em segmentos discursivos.

(39)

Preâmbulo: a. Leonardo e seus amigos estavam dividindo seus almoços.

Sentença de contexto:

O sujeito é o tópico	b. Leonardo entregou a Micael um sanduíche.
O sujeito não é o tópico	c. Micael entregou a Leonardo um sanduíche.

Sentença alvo:

Sujeito pronominal	d. Então ele passou para Carla uma maçã.
Sujeito não pronominal	e. Então Carla passou para ele uma maçã.

Segundo Chambers e Smith (1998), os resultados do primeiro experimento apontam para uma forte influência do paralelismo sintático tanto para sujeitos pronominais quanto para sujeitos não pronominais.

No que diz respeito à Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), Chambers e Smith (1998) afirmam que o dado importante era observar a resolução dos sujeitos não pronominais, já que, de acordo com esta teoria, um pronome que não ocupe a posição de sujeito é tipicamente julgado como correferente ao assunto da sentença anterior, pois a primeira entidade mencionada é o antecedente mais acessível, o que não foi confirmado por Chambers e Smith (1998).

De acordo com Chambers e Smith (1998), quando um enunciado contém um pronome que é estruturalmente congruente com enunciado anterior, um pronome que não ocupa a posição de sujeito é mais comumente interpretado

<sup>60</sup> (38) a. Josh criticized Paul and then Marie asked him to leave. (him = ?)

<sup>61</sup> (39) Preamble: Leonard and his friends were sharing their lunches.

Context Sentence:

Topic realized as subject:	Leonard handed Michael a sandwich.
Topic realizes as nonsubject:	Michael handed Leonard a sandwich.

Target Sentence:

Subject pronoun:	Then he passed Carla an apple.
Nonsubject pronoun:	Then Carla passed him an apple.

como correferente a uma entidade que também não esteja na posição de sujeito, mesmo quando o sujeito da sentença anterior é o tópico discursivo.

No segundo experimento, exemplificado em (40)<sup>62</sup>, Chambers e Smith (1998, p.600, tradução nossa, grifos do autor) manipularam as formas correferentes usando nomes repetidos e pronomes para investigar o papel desempenhado pelas expressões anafóricas na criação de coerência discursiva em enunciados estruturalmente congruentes.

(40)

Preâmbulo                    a. Um grupo de celebridades estava com problemas no carro.  
                                      b. Martin Miles disse a Liz Lovejoy para verificar o óleo.

Sentença alvo

Antecedente paralelo

Sujeito anafórico:        c. Então *ele/Martin Miles* disse a Dean Morgan para inspecionar o carburador.

Objeto anafórico:        d. Então Dean Morgan disse a *ela/Liz Lovejoy* para inspecionar o carburador.

Antecedente não paralelo

Sujeito anafórico:        e. Então *ela/Liz Lovejoy* disse a Dean Morgan para inspecionar o carburador.

Objeto anafórico:        f. Então Dean Morgan disse a *ele/Martin Miles* para inspecionar o carburador.

Sentença final:            g. O motor tinha superaquecido.

Os resultados do segundo experimento desenvolvido por Chambers e Smith (1998) apontam para a consolidação da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), foi percebido que usar um pronome como forma correferente em enunciados estruturalmente coerentes torna mais fácil de processar o segmento discursivo, independente do papel gramatical deste pronome.

---

<sup>62</sup> (40) Preamble:                    A group of celebrities were having car trouble.  
    Martin Miles told Liz Lovejoy to check the oil.

Target Sentence:

Parallel Antecedent:

Subject Anaphor:            Then *he/Martin Miles* told Dean Morgan to inspect the coolant.

Nonsubject Anaphor:        The Dean Morgan told *her/Liz Lovejoy* to inspect the coolant.

Nonparallel Antecedent:

Subject Anaphor:            The *she/Liz Lovejoy* told Dean Morgan to inspect the coolant.

Nonsubject Anaphor:        The Dean Morgan told *him/Martin Miles* to inspect the coolant.

Final Sentence:

The engine had overheated.

Assim, Chambers e Smith (1998) mostraram que a congruência estrutural dos segmentos discursivos favorece o estabelecimento da coerência semântica entre uma anáfora e seu termo antecedente realizado na mesma posição sintática. Contudo, Chambers e Smith (1998) consideraram necessário observar o que acontece quando um enunciado tem mais de uma entidade previamente introduzida no discurso sendo retomada por pronomes ou quando apenas uma destas entidades é pronominalizada, observando, portanto, o efeito da pronominalização como fator de coerência discursiva, o que aconteceu por meio do terceiro experimento, exemplificado em (41)<sup>63</sup> (CHAMBERS; SMITH, 1998, p. 603, tradução nossa).

(41)

Preâmbulo:	Uma briga estava acontecendo no quintal. Debbie deu um soco no nariz de David.
Sentença alvo:	
Nome – Nome:	Então Debbie deu um soco nas costelas de David.
Pronome – Nome:	Então ela deu um soco nas costelas de David.
Nome – Pronome:	Então Debbie deu um soco nas costelas dele.
Pronome – Pronome:	Então ela deu um soco nas costelas dele.

De acordo com Chambers e Smith (1998), os resultados do terceiro experimento apontam que a coerência resultante do uso de formas correferentes não é limitado a uma única anáfora que atenda à definição de centro anafórico. Além disso, Chambers e Smith (1998) encontraram um efeito de penalidade para nomes repetidos quando comparados os tempos de processamento das sentenças nas quais as duas entidades são retomadas por nomes em vez de pronomes.

Contudo, Chambers e Smith (1998) chamam a atenção para o fato de não haver diferenças significativas no tempo de processamento das sentenças com retomadas pronome-nome e nome-pronome, pois a falta de significância para o tempo de processamento destas sentenças aponta para uma possível

---

<sup>63</sup> Preamble: A fight was in full swing in the back yard.  
Debbie punched David in the nose.

Target Sentence:

Name-Name	Then Debbie slugged David in the ribs.
Pronoun-Name	Then she slugged David in the ribs.
Name-Pronoun	Then Debbie slugged him in the ribs.
Pronoun-Pronoun	Then she slugged him in the ribs.

relatividade na retomada da primeira ou da segunda entidade por meio de um pronome, o que é inconsistente com os resultados de Gordon, Grosz e Guilliom (1993) que afirmaram ser um pronome preferencialmente a ocupar a posição de sujeito, salvos os casos em que a posição de sujeito será preenchida por uma nova informação.

Assim, observados os resultados de cada um dos experimentos acima descritos, Chambers e Smith (1998) propõem que a coerência dos segmentos discursivos foi facilitada quando um pronome se refere a um antecedente que ocupe a mesma posição sintática que ele e, assim, adotam<sup>64</sup> “a concepção de paralelismo estrutural como estando restrito à congruência dos enunciados” (CHAMBERS; SMITH, p.604, tradução nossa).

Além disso, Chambers e Smith (1998) demonstraram que a maneira como a expressão anafórica contribui para a compreensão do segmento discursivo não é explicado totalmente na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), alegando que o paralelismo estrutural impõe fortes restrições às potenciais anáforas e que as propriedades semânticas dos enunciados devem ser levadas em conta para determinar como a inferência das formas correferentes influencia a construção da coerência dos segmentos discursivos.

A partir dos resultados encontrados por Chambers e Smith (1998), Gordon e Hendrick (1998) discutiram, pautados no formalismo proposto pela Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), os possíveis efeitos gerados pelo paralelismo sintático durante o processamento correferencial e corroboraram os achados de Chambers e Smith (1998).

Segundo Gordon e Hendrick (1998), a estrutura sintática e a sequência das sentenças dentro dos segmentos discursivos afetam o processamento correferencial. Gordon e Hendrick (1998) afirmam que, segundo o modelo por eles proposto, a função primária do pronome é retomar uma entidade já introduzida no discurso, enquanto a função primária dos nomes é introduzir novas entidades no discurso.

Além disso, Gordon e Hendrick (1998) destacam que, para resultados confiantes acerca do processamento anafórico, é preciso que sejam empregadas técnicas que levem o pesquisador a resultados com dois tipos de

---

<sup>64</sup> The definition of parallelism adopted was the strict congruence of the overall constituent structure of the utterances.

dados, sendo tanto dados de julgamentos intuitivos como dados resultantes do tempo de compreensão da linguagem.

Tal necessidade, segundo os autores, ocorre pelo fato de que, segundo Gordon e Hendrick (1998), a interpretação dos pronomes depende da ligação inicial entre o pronome em função anafórica e seu antecedente e, posteriormente, pelo fato de emergirem os aspectos semânticos que levam à resolução do par anáfora-antecedente.

Levando em consideração as teorias sintáticas revisitadas nesta seção, observamos que, em muitos dos trabalhos citados ao longo desta seção, há a concordância de que, uma vez estabelecidos os padrões dos centros anafóricos a serem preenchidos sintaticamente pelos nomes ou pelos pronomes, há a necessidade de resolver a correferência semanticamente.

Por conta disto, revisaremos na seção a seguir as teorias semânticas sobre o processamento correferencial.

### **2.2.2 Teorias e hipóteses semânticas sobre o processamento anafórico**

As pesquisas sobre o processamento correferencial sob a ótica da sintaxe<sup>65</sup> têm se debruçado em explicar como os centros anafóricos tendem a privilegiar o uso do pronome em detrimento do nome através do tempo de processamento do pronome em tarefas linguísticas.

Contudo, a maioria dos pesquisadores<sup>66</sup> afirma que, uma vez estabelecidas as posições sintáticas a serem retomadas, a resolução correferencial depende de critérios semânticos e, levando em considerações tais afirmações, alguns trabalhos têm se debruçado em analisar como os pronomes em função anafórica são processados.

Os trabalhos sobre a influência da semântica no processamento correferencial são focados em explicar a tipicidade da relação entre o par anáfora-antecedente e serão revistados nesta seção.

---

<sup>65</sup> Para mais informações, ver a seção anterior desta tese.

<sup>66</sup> *Idem.*

### 2.2.2.1 Teoria da sobreposição semântica

Garrod e Sanford (1977) investigaram, por meio de quatro experimentos de leitura automonitorada, os fatores que afetavam o tempo de processamento das sentenças de um texto.

Para Garrod e Sanford (1977), a compreensão leitora resulta de tarefas que o leitor é capaz de executar, a fim de integrar os elementos do texto de forma coerente, armazenando as informações lidas na memória. O papel do leitor é, portanto, organizar as informações dadas sobre objetos, eventos, pessoas, etc. e ser capaz de organizar as informações dadas ao longo do texto em um local da memória para a qual ele retornará sempre que uma nova informação dada tenha correlação a outra já armazenada neste local.

Segundo Garrod e Sanford (1977), um dos problemas de conceber um local da memória que armazena todas as informações dadas em um texto e ao qual o leitor é capaz de recorrer quando uma passagem do texto retoma uma destas informações, é compreender como o usuário de uma língua é capaz de retornar, por meio de palavras que não se repetem, mas que possuem significados interligados, estabelecendo a correferência entre os termos.

Garrod e Sanford (1977) buscavam entender como um objeto específico e que possui certa variedade de palavras que remetem a ele em um texto pode ser retomado ao longo de uma sequência de sentenças, criando correferência entre os termos que estão armazenados na memória, como em (42)<sup>67</sup> (GARROD; SANFORD, 1977, p.77, tradução nossa).

- (42) a. Um ônibus apareceu na esquina  
b. O veículo quase atropelou um pedestre.

A fim de compreender como falantes nativos de dado idioma compreendem estas possibilidades de variação semântica entre o termo antecedente e a forma correferente usada, Garrod e Sanford (1977) desenvolveram uma sequência de quatro experimentos de leitura para observar como ocorre a ligação entre estes termos durante o estabelecimento da coerência de um texto.

---

<sup>67</sup> (73) a. A *bus* came roaring round the corner.  
b. The *vehicle* nearly flattened a pedestrian.

O primeiro experimento foi criado, a fim de analisar se a frequência de uso de determinadas palavras teria algum efeito durante o processamento de sentenças que exigissem o mínimo de reconhecimento de sentido entre o termo antecedente e a forma anafórica usada para retomá-lo. Para isso, Garrod e Sanford (1977) manipularam, na primeira sentença, palavras que podem ser ora mais comumente usadas, ou seja, com alta frequência de uso, ora menos comumente usadas, ou seja, com baixa frequência de uso na segunda sentença, como exemplificado em (43)<sup>68</sup> (GARROD; SANFORD, 1977, p.79, tradução nossa).

- (43) a. Um pintassilgo<sub>i</sub>/um ganso<sub>i</sub> às vezes perambulava por dentro da casa.  
b. O pássaro<sub>i</sub> foi atraído para a dispensa.

Com o objetivo de observar como a introdução da entidade pode interferir durante o estabelecimento da correferência, Garrod e Sanford (1977) optaram por mostrar, durante a realização da tarefa de leitura, uma segunda versão dos estímulos, colocando como primeira sentença a classe de palavras geral e, na segunda sentença, os exemplares citados, como em (44)<sup>69</sup> (GARROD; SANFORD, 1977, p.79, tradução nossa).

- (44) Um pássaro<sub>i</sub> às vezes perambulava por dentro da casa.  
O pintassilgo<sub>i</sub>/o ganso<sub>i</sub> foi atraído para a dispensa.

Segundo Garrod e Sanford (1977), os resultados do primeiro experimento mostraram que houve efeito significativo durante o tempo de processamento da segunda sentença, quando na primeira surge um exemplar específico a ser retomado de maneira mais geral na segunda, o que aponta para uma integração de informação da primeira sentença conjuntamente à segunda para o processamento da correferência.

Ao compararem o tempo de leitura das sentenças, Garrod e Sanford (1977) afirmaram que, ao estabelecer ligação entre a anáfora e um antecedente com alta frequência de uso, o tempo de processamento da correferência é menor

---

<sup>68</sup> (43) a. A *robin/goose* would sometimes wander into the house.  
b. The *bird* was attracted by the larder.

<sup>69</sup> (44) a. A *bird* would sometimes wander into the house.  
b. The *robin/goose* was attracted by the larder.

do que quando o par inverso, ou seja, entre um termo antecedente de menor frequência com a sua forma correferente. Segundo Garrod e Sanford (1977), este tipo de efeito foi observado tanto no segmento discursivo em que primeiro foi introduzido o exemplar, sendo retomado pela classe mais geral, quanto na forma inversa, com a classe geral como entidade previamente introduzida, sendo retomada por uma palavra menos frequente como termo correferente.

Assim, Garrod e Sanford (1977) sugerem que o processamento da leitura de sentenças com ligação correferencial entre um termo antecedente mais ou menos específico e sua classe mais geral ocorre no momento em que os pares de sentença ficam disponíveis ao leitor, o que sugere um efeito de *priming* semântico-lexical que rege a resolução do par antecedente/anáfora.

No segundo experimento, Garrod e Sanford (1977) buscaram observar se o mesmo efeito encontrado no primeiro experimento aconteceria na presença de uma sentença interveniente e, assim, utilizaram como estímulo, segmentos discursivos como (45)<sup>70</sup> (GARROD; SANFORD, 1977, p.79, tradução nossa).

- (45) a. Um veículo; surgiu barulhento na esquina.  
 b. Ele teve uma falha no freio.  
 c. O ônibus; quase atropelou um pedestre.

Para a análise do tempo de leitura dos segmentos discursivos utilizados como estímulos do segundo experimento, Garrod e Sanford (1977) excluíram a retomada anafórica da segunda sentença por ser ela considerada apenas como material com informação sobre uma entidade previamente introduzida no discurso<sup>71</sup>, recaindo o interesse sobre o processamento da forma correferente usada na terceira sentença.

De acordo com Garrod e Sanford (1977), a informação da terceira sentença é novamente resultante da integração conjunta das informações dadas na primeira e na segunda sentença, o que confirma os resultados do primeiro experimento, ou seja, há efeito de *priming* semântico-lexical mesmo na presença de uma sentença interveniente entre o antecedente e sua forma correferente.

---

<sup>70</sup> (45) a. A *vehicle* came roaring round the corner.  
 b. It had had a brake failure.  
 c. The *bus* nearly flattened a pedestrian.

<sup>71</sup> Para mais detalhes acerca desta discussão, ver Garrod e Sanford (1977).

Já no terceiro experimento, Garrod e Sanford (1977), tomando como base que no primeiro e no segundo experimento haviam observado o que acontece quando palavras diversas que podem se referir a uma mesma entidade, com ou sem sentença interveniente, tem sua correferência estabelecida ao longo do processamento do texto, investigaram qual seria o efeito resultante do uso de diversas palavras que se referem a uma mesma entidade, sem haver correferência entre o termo antecedente e a sua retomada anafórica.

Este experimento foi desenvolvido por meio de dois pares de sentenças com uma pergunta de controle, sendo que metade dos estímulos experimentais continha palavras de alta ou de baixa frequência de uso a serem retomadas na segunda sentença na posição de objeto<sup>72</sup> por meio de um pronome anafórico e, posteriormente, por meio da classe de palavras mais geral com a qual era correferente, como em (46)<sup>73</sup> (GARROD; SANFORD, 1977, p.85, tradução nossa).

- (46) a. Um ônibus/um tanque; surgiu barulhento na esquina.  
 b. Ele quase esmagou alguns veículos.  
 O ônibus esmagou alguns veículos?

Já a outra metade dos itens experimentais do terceiro experimento continha, na segunda sentença, uma nova entidade na posição de sujeito para que o termo não-correferente surgisse novamente na posição de objeto<sup>74</sup>, como em (47)<sup>75</sup> (GARROD; SANFORD, 1977, p.85, tradução nossa).

- (47) a. Um ônibus/um tanque; surgiu barulhento na esquina.  
 b. Um pedestre quase foi morto pelo veículo.  
 O pedestre quase foi morto pelo veículo?

---

<sup>72</sup> Há indícios de que, para Garrod e Sanford (1977) a correferência deve ser estabelecida por meio de questões de cunho sintático, já que eles afirmam que a correferência no terceiro experimento deixa de ser obedecida pelo fato de uma entidade na posição de sujeito ser, posteriormente, retomada na posição de objeto. Para mais informações, ver Paralelismo Sintático (CHAMBERS; SMITH, 1998) na seção anterior.

<sup>73</sup> (46) a. A bus/tank came roaring round the corner.  
 b. It nearly smashed some vehicles.  
 Did the bus smash some vehicles?

<sup>74</sup> Pela tradução em português, o termo correferente aparece como agente da passiva e não como objeto. Para os originais, ver Garrod e Sanford (1977).

<sup>75</sup> (47) a. A bus/tank came roaring round the corner.  
 b. A pedestrian was nearly killed by the vehicle.  
 Was the pedestrian nearly killed by the bus?

Segundo Garrod e Sanford (1977), foram encontrados, neste experimento, resultados que mostram que os leitores parecem fazer um tipo de processo de checagem da relação existente entre as palavras, já que a maioria dos participantes parece aceitar que *ônibus* pode ser retomado pela forma *o veículo*, como em (47), mas não pela forma *alguns veículos*, como em (46).

Isto, para Garrod e Sanford (1977), confirma os resultados do primeiro e do segundo experimento, apontando para um efeito de *priming* semântico-lexical, já que os participantes pareceram associar uma única palavra com o item ao qual se liga, tornando a sentença mais fácil de ser processada, ocorrendo o oposto quando os participantes tentaram associar mais de uma palavra a um item previamente mencionado, tornando, por conta disto, a sentença mais difícil de ser processada.

Por conta destes resultados, Garrod e Sanford (1977) desenvolveram um quarto experimento, a fim de observar os mecanismos de *priming* semântico-lexical e avaliação implícita da frequência uso das palavras associadas entre si, pois, para eles, poderia haver uma diferença básica entre estes dois mecanismos, já que “eles implicam no envolvimento de mecanismos totalmente diferentes”<sup>76</sup> (GARROD e SANFORD, 1977, p.87, tradução nossa).

Com vistas a cumprir o objetivo do quarto experimento, Garrod e Sanford (1977) criaram um *design* que permitisse a escolha entre formas correferentes possíveis para, assim, observar o comportamento dos participantes com relação a palavras que tem correlação entre si, mas não apresentam frequência de uso na associação feita no texto, como em (48)<sup>77</sup> (GARROD e SANFORD, 1977, p.87, tradução nossa).

- (48) a. Um ônibus;/um tanque; surgiu barulhento na esquina.  
b. Ele quase atingiu um veículo; puxado por cavalos.<sup>78</sup>

<sup>76</sup> They imply the involvement of totally different mechanisms.

<sup>77</sup>(48) a. A tank/bus came roaring round the corner.

b. It nearly hit a horse-drawn vehicle.

<sup>78</sup> Com estes itens experimentais, os autores acreditavam ser possível encontrar a diferença entre o efeito de *priming* semântico-lexical e o efeito de frequência conjunta (quando a frequência de uso entre o termo antecedente e sua anáfora é grande). Entretanto, é preciso entender que o efeito de *priming* tem como base o princípio de excitação, por meio do qual a primeira palavra tem capacidade de facilitar na compreensão da segunda. Por conta disto, é necessário ter em mente que a segunda sentença do experimento original tem o grupo semântico *puxado por cavalos* antecedendo a palavra *veículos* (*It nearly hit a horse-drawn vehicle*).

De acordo com Garrod e Sanford (1977), os resultados do quarto experimento apontam para a realização do possível efeito de *priming* como decorrente do uso de apenas uma palavra como correferente à entidade previamente introduzida, confirmando os resultados do terceiro experimento, já que estes mostraram a possibilidade de um processo de busca executado pelos participantes, a fim de descobrir as relações existentes entre as palavras usadas como possíveis correferentes dentro do texto.

Após a bateria de experimentos citados, Garrod e Sanford (1977) afirmaram que os casos em que termos mais específicos são usados como expressões anafóricas configuram uma raridade em textos naturais, pelo fato de construir limitações na mente do leitor.

Isto, de acordo com Garrod e Sanford (1977), acontece porque o leitor, durante uma tarefa de leitura automonitorada, demonstrou ser capaz de atribuir informação à expressão anafórica por meio de um mecanismo de conferência das informações conjuntas entre o antecedente e a anáfora, o que ocorre por meio de dois processos.

O primeiro processo é originado de uma análise preliminar das sentenças em si, por meio de uma observação sintática, como a presença de um artigo definido que indica a presença de um antecedente ou um segundo processo, levando o leitor a procurar de maneira ativa as informações relevantes mencionadas anteriormente, a fim de estabelecer a correferência entre os termos, prevalecendo, então, questões de cunho semântico.

Garrod e Sanford (1977) afirmam que, durante a resolução da correferência, opera na mente do leitor um sistema automático de checagem de informações entre o antecedente e a anáfora. Deste sistema, podemos observar que ao lançar uma entidade no texto, outros lugares serão sintaticamente abertos para que esta entidade seja retomada e, então, o leitor passará a buscar as possíveis formas anafóricas que poderão ocupar essa posição, mantendo o tópico discursivo ativo na mente do leitor ao longo do texto. Quando o tópico mudar, este local, antes aberto, passará a ser psicologicamente visto como fechado e só será reaberto quando, em alguma sentença, as novas informações não coincidirem com o tópico anteriormente citado, ou seja, a reabertura do local de ligação correferencial passa a depender da introdução de um novo tópico discursivo.

Pensamos que a identificação das informações relativas a um local aberto pode ser refletida como um efeito da frequência conjunta (...). Além disso, sugerimos que o mecanismo pelo qual a informação passa a ser aceita como relevante depende do grau de características da sobreposição semântica entre o antecedente e a anáfora (GARROD e SANFORD, 1977, p. 89, tradução nossa)<sup>79</sup>.

No modelo de Sobreposição Semântica, conforme proposto por Garrod e Sanford (1977), a sintaxe exerce um papel fraco durante a checagem das palavras que estão à disposição do preenchimento dos locais abertos para o estabelecimento da correferência ao longo do texto, ficando a cargo da semântica explicar como uma forma anafórica pode ser mais ou menos fácil de processar.

Por tratar do quão típico é o uso de nomes mais gerais com relação a nomes mais específicos quando da retomada anafórica, Garrod e Sanford (1977) mostraram que o par anáfora-antecedente, para ser interpretado coerentemente, depende dos critérios semânticos entre os termos correferentes.

Entretanto, os resultados que possibilitaram à postulação da Teoria da Sobreposição Semântica (GARROD; SANFORD, 1977), levaram Almor (1999) a observar a tipicidade da anáfora com relação ao seu referente, encontrando resultados que vão de encontro às predições da teoria criada por Garrod e Sanford (1977).

#### *2.2.2.2 Hipótese da Carga Informacional*

O processamento de SNs e anáforas no discurso dependem das formas linguísticas utilizadas para retomar a entidade previamente introduzida e dependem da ativação/otimização de informações na memória de trabalho.

Na busca de evidências sobre as motivações psicológicas que levam à escolha de uma expressão anafórica durante o processamento da leitura, Almor (1999) desenvolveu uma série de cinco experimentos utilizando a técnica de leitura automonitorada, a fim de explicar os custos de processamento<sup>80</sup> envolvidos na resolução anafórica.

---

<sup>79</sup> We would think of the identification of information relating to an open location as being reflected in the conjoint frequency effect. (...) Further, we suggest that one mechanism whereby information is accepted as relevant is the degree of semantic feature overlap between the instance and the category.

<sup>80</sup> Almor (1999) defende que, em sua teoria, o princípio de custo de processamento diz respeito aos custos de tempo adicionais para que o leitor possa entender à qual entidade a anáfora se refere, pois, para o autor, naturalmente, o leitor não precisa refletir sobre o termo antecedente

Ao discutir teorias que tratam sobre o processamento anafórico<sup>81</sup>, Almor (1999) aponta para o fato de que as teorias têm discutido sobre como acontece a retomada de SNs por pronomes anafóricos que, em geral, tendem a retomar uma entidade em foco discursivo. Contudo, Almor (1999) chama a atenção para o fato de que o uso de anáforas repetidas descarta a possibilidade de que a expressão anafórica possa introduzir novas informações sobre a entidade à qual ela se refere, como em (49)<sup>82</sup> (ALMOR, 1999, p.749, tradução nossa, grifos do autor).

- (49) a. Um sabiá comeu uma fruta. O *sabiá* parecia muito satisfeito.  
 b. Um sabiá comeu uma fruta. O *passarinho* parecia muito satisfeito.  
 c. Um sabiá comeu uma fruta. O *pássaro* parecia muito satisfeito.  
 d. Um sabiá comeu uma fruta. Ele *parecia* muito satisfeito<sup>83</sup>.

Segundo Almor (1999), a teoria por ele proposta, denominada Hipótese da Carga Informacional, sugere que o processamento de expressões anafóricas reflete um princípio pragmático parecido com a máxima de Grice (1975) da quantidade máxima. Além da relação com esta máxima, Almor (1999) afirma ser necessário entender que (i) a complexidade da resolução anafórica é representada pela medida da carga informacional trazida pela anáfora para a memória de trabalho e (ii) a informação dada pela expressão anafórica pode fazer com que o leitor identifique o antecedente, adicione novas informações sobre o antecedente ou ambos.

Almor (1999) defende que a memória de trabalho desempenha papel fundamental em sua teoria pelo fato de esta memória possuir uma capacidade limitada e que, esta capacidade, é usada tanto para armazenar quanto para processar informação. E, segundo Almor (1999), é justamente a capacidade da

---

ao qual a anáfora está ligada e, se ocorrer de o leitor ter de buscar informações adicionais para resolver a ligação entre antecedente e anáfora, o tempo de processamento será maior e, necessariamente, há custo cognitivo maior, gerando o chamado custo de processamento. Portanto, o custo de processamento, segundo o autor, está diretamente ligado à Semântica.

<sup>81</sup> Para mais informações acerca das teorias discutidas e que levaram à formulação da Hipótese da Carga Informacional ver Almor (1999)

<sup>82</sup> (49) a. A robin ate the fruit. The *robin* seemed very satisfied.  
 b. A robin ate the fruit. The *the wet little bird* seemed very satisfied.  
 c. A robin ate the fruit. The *bird* seemed very satisfied.  
 d. A robin ate the fruit. *It* seemed very satisfied.

<sup>83</sup> É importante notar que, para Almor (1999), o estudo da anáfora se dá no nível discursivo, compartilhando noções de discurso e Pragmática conforme concebidos na Teoria da Acessibilidade (ARIEL, 1991). Desta forma, entendemos que o trabalho de Almor (1999) deve ser classificado na interface semântico-pragmática.

memória de trabalho que leva muitos pesquisadores a acreditarem que o uso de expressões anafóricas no discurso é o reflexo desta limitação.

Muitos pesquisadores<sup>84</sup> acreditam que a memória de trabalho é responsável por dar conta de todos os recursos necessários para o processamento da linguagem, o que, de acordo com Almor (1999), leva o leitor a atribuir as restrições expressas na teoria da Hipótese da Carga Informacional quando estão executando tarefas linguísticas que envolvam o processamento anafórico.

Visando formalizar sua teoria, Almor (1999) mostra que o processamento de SNs, de acordo com a teoria da Hipótese da Carga Informacional, depende de três fatores: (i) foco discursivo, (ii) novas informações adicionadas pela anáfora ao antecedente e (iii) a carga informacional do par anáfora-antecedente para, assim, determinar o custo de processamento do par.

A Hipótese da Carga Informacional implica que se uma anáfora adicionar nova informação sobre o referente, quanto menor for a carga informacional do par anáfora-antecedente, mais fácil será o processamento, desde que essa carga informacional seja suficiente para identificar o antecedente. Isto significa que, nos casos em que o antecedente está em foco e é, portanto, o antecedente 'padrão', uma anáfora de alta carga informacional seria mais difícil de processar se ela não adicionar nenhuma informação ao seu antecedente (ALMOR, 1999, p. 753, tradução nossa).<sup>85</sup>

Esta percepção levou Almor (1999) a fazer previsões que norteiam a teoria por ele criada, conforme descritas a seguir:

*Predição 1.* Anáforas com baixa carga informacional que dizem respeito a um único antecedente são mais fáceis de processar quando o antecedente está em foco do que quando não estão. *Predição 2.* Anáforas com alta carga informacional que dizem respeito a um único antecedente e que adicionam nova informação sobre seu referente são mais fáceis de processar quando o antecedente está em foco do que quando não está. *Predição 3.* Anáforas com alta carga informacional que dizem respeito a um único antecedente são mais fáceis de processar quando o antecedente não está em foco do que quando está. *Predição 4.* Havendo um antecedente e uma anáfora que não adiciona nova informação, quanto mais carga informacional tiver a

<sup>84</sup> Para ler os trabalhos e autores discutidos para dar origem a Hipótese da Carga Informacional, ver Almor (1999).

<sup>85</sup> The ILH implies that if an anaphor does not add new information about its referent, then the smaller the informational load of the anaphor-antecedent pair, the easier it is to process as long as this informational load is sufficient for identifying the antecedent. This means that in cases in which the antecedent is focused and is, therefore, the "default" antecedent, an anaphor with high informational load with respect to that antecedent would be harder to process if it does not add any new information about its referent.

anáfora com relação ao antecedente, mas difícil será de processá-la (ALMOR, 1999, p.753, tradução nossa)<sup>86</sup>.

O primeiro experimento desenvolvido por Almor (1999) visava analisar, com base na primeira predição, se as anáforas que não adicionam novas informações ao antecedente seriam mais fáceis de processar quando possuem menor carga informacional, referindo-se a um antecedente que é o foco discursivo. Exemplificamos as quatro condições propostas pelo autor em (50) (ALMOR, 1999, p.755, tradução nossa)<sup>87</sup>.

(50)

Construção	Primeiro	Posição do antecedente Segundo
Foco no primeiro nome	(i)	(ii)
a. Foi o <u>sabiá</u> que comeu a maçã.		a. Foi o <u>sabiá</u> que comeu a maçã.
b. O pássaro parecia muito satisfeito.		b. A fruta já estava meio podre.
c. O pássaro tocou a fruta?		c. O pássaro tocou a fruta?
Foco no segundo nome	(i)	(ii)
a. O que o <u>sabiá</u> comeu foi a <u>maçã</u> .		a. O que o <u>sabiá</u> comeu foi a <u>maçã</u> .
b. O pássaro parecia muito satisfeito.		b. A fruta já estava meio podre.
c. O pássaro tocou a fruta?		c. O pássaro tocou a fruta?

Os resultados do primeiro experimento confirmam a predição feita por Almor (1999) e mostram que quando a anáfora é um termo mais geral, se ela tiver uma baixa carga informacional com relação ao seu antecedente, sem adicionar novas informações sobre o referente, o fato de o antecedente ser o foco do discurso, facilita o processamento.

Já o segundo experimento desenvolvido por Almor (1999), visando analisar a segunda predição da Hipótese da Carga Informacional, testava se as

<sup>86</sup> *Prediction 1.* Anaphors with low informational load with respect to a particular antecedente are easier to process when that antecedent is focused that when it is not. *Prediction 2.* Anaphor with high informational load with respect to a particular antecedent and that add new information about their referent are easier to process when that antecedent is focused that when it is not. *Prediction 3.* Anaphor with high informational load with respect to a particular antecedent, which do not add new information about their antecedents are easier to process when their antecedent is not focused that when it is. *Prediction 4.* Given focused antecedent and an anaphor that does not add new information, the more informationally loaded that anaphor is with respect to that antecedent, the harder it is to process.

87(50)

Condition	First	Antecedent position Second
It cleft (focus on first noun)	(i)	(ii)
a. It was <u>the robin</u> that ate the apple.		a. It was <u>the robin</u> that ate the apple.
b. The bird seemed very satisfied.		b. The fruit was already half rotten.
c. Did the bird touch the fruit?		c. Did the bird touch the fruit?
Wh-cleft (focus on second name).	(iii)	(iv)
a. What the robin ate was <u>the apple</u> .		a. What the robin ate was <u>the apple</u> .
b. The bird seemed very satisfied.		b. The fruit was already half rotten.
c. Did the bird touch the fruit?		c. Did the bird touch the fruit?

anáforas com grande carga informacional, mas que adicionam novas informações sobre seu antecedente, seriam mais fáceis de processar quando seu antecedente é o foco do discurso.

De acordo com Almor (1999), as anáforas, quando mais específicas do que seus antecedentes, têm alta carga informacional com relação a estes, mas, por servirem para adicionar informações, sendo mais facilmente processadas quando seus antecedentes são o foco discursivo.

Neste experimento, Almor (1999) utilizou os mesmos itens experimentais do primeiro experimento, sendo modificada a ordem do antecedente e da anáfora, pois, neste experimento, o termo mais geral *o pássaro* era o antecedente e o termo mais específico *o sabiá* era a anáfora, gerando passagens experimentais como *Foi o pássaro que comeu a maçã. O sabiá parecia muito satisfeito*<sup>88</sup> (ALMOR, 1999. p.757, tradução nossa).

Os resultados do segundo experimento, segundo Almor (1999), mostram que a segunda predição feita na Hipótese da Carga Informacional está correta, pois quando as anáforas com alta carga informacional servem para introduzir novas informações sobre o antecedente, elas são facilmente processadas, desde que seu antecedente seja o foco do discurso. Para Almor (1999), isto mostra que a escolha das anáforas pode ser justificada de acordo com o seu uso a partir dos fatores que entram em jogo durante o processamento anafórico.

Segundo Almor (1999), esta justificativa funcional para o uso de anáforas de alta carga informacional reside no fato de que, quando a antecedente serve como foco do discurso, a ligação do par anáfora-antecedente facilita o reconhecimento deste antecedente. Por outro lado, quando uma anáfora de alta carga informacional não adiciona novas informações a um antecedente que é foco do discurso, então o seu processamento deve ser inversamente proporcional, ou seja, será mais difícil de processar, de acordo com a terceira predição da Hipótese da Carga Informacional.

Almor (1999) desenvolveu o terceiro experimento para observar a influência da mudança no foco do discurso ao qual a anáfora se referia, conforme exemplificado em (51) (ALMOR, 1999, p.158, tradução nossa).

---

<sup>88</sup> It was the bird that ate the apple. The robin seemed very satisfied.

- (51)<sup>89</sup>
- (a) Foi o pássaro que comeu a fruta.  
O *pássaro* parecia muito satisfeito.
  - (b) O que o pássaro comeu foi a fruta.  
O *pássaro* parecia muito satisfeito.
  - (c) Foi o pássaro que comeu a fruta.  
A *fruta* já estava meio podre.
  - (d) O que o pássaro comeu foi a fruta.  
A *fruta* já estava meio podre.

O terceiro experimento mostra, segundo Almor (1999), resultados proporcionalmente inversos aos encontrados nos dois experimentos anteriores, pois as anáforas repetidas foram lidas mais lentamente quando se referiam a um antecedente que era o foco do discurso.

De acordo com Almor (1999), os resultados deste experimento corroboram a terceira predição da Hipótese da Carga Informacional, mas vai de encontro aos resultados de Gordon, Grosz e Guilliom (1993) e as previsões que levaram ao efeito da Penalidade do Nome Repetido.

Segundo Gordon, Grosz e Guilliom (1993), as anáforas feitas pela repetição de nomes são mais difíceis de serem processadas quando seu antecedente é o foco do discurso. Para Almor (1999), há uma interpretação ruim para a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), pois, pode-se notar que este efeito de penalidade ocorre quando um nome repetido anafórico é lido tanto quando seu antecedente é o foco do discurso como quando não é.

No quarto experimento, Almor (1999) observou como as descrições indefinidas são processadas, já que o uso de descrições definidas nos experimentos anteriores pode ter violado o princípio de que os participantes deveriam estar familiarizados com as entidades introduzidas no discurso, posteriormente, retomadas por meio da anáfora, o que poderia interferir no processamento anafórico.

Almor (1999) desenvolveu, então, o quarto experimento usando materiais que não violassem o princípio da descrição definida, para, assim, comprovar se

---

<sup>89</sup> (51) a. It was the bird that ate the fruit.  
*The bird* seemed very satisfied.  
b. What the bird ate was the fruit.  
*The bird* seemed very satisfied.  
c. It was the bird that ate the fruit.  
*The fruit* was already half rotten.  
d. What the bird ate was the fruit.  
*The fruit* was already half rotten.

a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) está associada à facilitação de familiaridade com as referências feitas ao longo do discurso conforme exemplificado em (52) (ALMOR, 1999, p.759, tradução nossa)<sup>90</sup>.

(52)

Posição do antecedente	Construção Foco na primeira categoria Anáforas repetidas
Primeira	(i) a. Um homem e uma garota caminhavam com um cachorro e um papagaio b. Era a garota que caminhava com o papagaio c. A garota era muito pequena para andar com o cachorro.
	(ii) a. Um homem e uma garota caminhavam com um cachorro e um papagaio b. O que a garota caminhava era com o papagaio. c. A garota era muito pequena para andar com o cachorro.
Segunda	(iii)

---

<sup>90</sup> (52) Antecedent set position	Construction It cleft (focus on first category) Repetition anaphors
First	(i) a. A man and a girl were walking a dog and a parrot. b. It was the girl that walked the parrot. c. The girl was too small to walk the dog.
Second	(ii) a. A man and a girl were walking a dog and a parrot. b. It was the girl that walked the parrot. c. The parrot was small enough for the girl to carry.
First	(iii) a. A man and a girl were walking a dog and a parrot. b. It was the girl that walked the parrot. c. The man was strong enough to walk the big dog.
Second	(iv) a. A man and a girl were walking a dog and a parrot. b. It was the girl that walked the parrot. c. The dog was too big for the small girl to walk.
Antecedent set position	Construction Wh-cleft (focus on second category) Contrast anaphors
First	(i) a. A man and a girl were walking a dog and a parrot. b. What the girl walked was the parrot. c. The girl was too small to walk the dog.
Second	(ii) a. A man and a girl were walking a dog and a parrot. b. What the girl walked was the parrot. c. The parrot was small enough for the girl to carry.
First	(iii) a. A man and a girl were walking a dog and a parrot. b. What the girl walked was the parrot. c. The man was strong enough to walk the big dog.
Second	(iv) a. A man and a girl were walking a dog and a parrot. b. What the girl walked was the parrot. c. The dog was too big for the small girl to walk.

- a. Um homem e uma garota caminhavam com um cachorro e um papagaio
- b. Era a garota que caminhava com o papagaio
- c. A garota era muito pequena para andar com o cachorro.

(iv)

- a. Um homem e uma garota caminhavam com um cachorro e um papagaio
- b. O que a garota caminhava era com o papagaio.
- c. O papagaio era muito pequeno para ser carregado pela garota.

Posição do antecedente	Construção Foco na segunda categoria Anáforas de contraste
Primeira	<ul style="list-style-type: none"> <li>a. Um homem e uma garota caminhavam com um cachorro e um papagaio</li> <li>b. Era a garota que caminhava com o papagaio</li> <li>c. O homem era forte o suficiente para caminhar com o cachorro grande.</li> </ul>
	(v)
	<ul style="list-style-type: none"> <li>a. Um homem e uma garota caminhavam com um cachorro e um papagaio</li> <li>b. O que a garota caminhava era com o papagaio.</li> <li>c. O homem era forte o suficiente para caminhar com o cachorro grande.</li> </ul>
	(vi)
Segunda	<ul style="list-style-type: none"> <li>a. Um homem e uma garota caminhavam com um cachorro e um papagaio</li> <li>b. Era a garota que caminhava com o papagaio</li> <li>c. O cachorro era muito grande para a pequena garota levar.</li> </ul>
	(vii)
	<ul style="list-style-type: none"> <li>a. Um homem e uma garota caminhavam com um cachorro e um papagaio</li> <li>b. O que a garota caminhava era com o papagaio.</li> <li>c. O cachorro era muito grande para a pequena garota levar.</li> </ul>
	(viii)

De acordo com Almor (1999), os resultados do quarto experimento apontam que a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) pode ser explicada como uma idiosincrasia das anáforas repetidas, sem relação à carga informacional delas, que é o fato relevante para a Hipótese da Carga Informacional (ALMOR, 1999).

O quinto experimento, exemplificado em (53) (ALMOR, 1999, p.760, tradução nossa, grifos do autor)<sup>91</sup>, visava analisar a quarta predição da Hipótese da Carga Informacional (ALMOR, 1999) que diz: “havendo um antecedente e uma anáfora que não adiciona nova informação, quanto mais carga informacional tiver a anáfora com relação ao antecedente, mas difícil será de processá-la” (ALMOR, 1999, p.753, tradução nossa)<sup>92</sup>.

<sup>91</sup>(53) a. The professor and her student arranged the transportation for their field trip. It was the student that rented *the car*. *The vehicle* was necessary for getting to the exploration site.  
 b. The professor and her student arranged the transportation for their field trip. It was the student that rented *the boat*. *The vehicle* was necessary for getting to the exploration site.  
 c. The professor and her student arranged the transportation for their field trip. What the student rented was *the car*. *The vehicle* was necessary for getting to the exploration site.  
 d. The professor and her student arranged the transportation for their field trip. What the student rented was *the boat*. *The vehicle* was necessary for getting to the exploration site.

<sup>92</sup> *Prediction 4*. Given focused antecedent and an anaphor that does not add new information, the more informationally loaded that anaphor is with respect to that antecedent, the harder it is to process.

(53)

(a) A professora e sua aluna organizaram o transporte para a viagem de campo. Foi a aluna que alugou *o carro*. *O veículo* era necessário para chegar ao local da exploração.

(b) A professora e sua aluna organizaram o transporte para a viagem de campo. Foi a aluna que alugou *o barco*. *O veículo* era necessário para chegar ao local da exploração.

(c) A professora e sua aluna organizaram o transporte para a viagem de campo. O que a aluna alugou foi *o carro*. *O veículo* era necessário para chegar ao local da exploração.

(d) A professora e sua aluna organizaram o transporte para a viagem de campo. O que a aluna alugou foi *o barco*. *O veículo* era necessário para chegar ao local da exploração.

Segundo Almor (1999) ao postular a quarta predição, uma dada anáfora correferente de uma forma mais específica será lida mais rapidamente e mais facilmente processada quando estiver semanticamente distante de seu referente.

O efeito que é gerado pela distância semântica neste experimento pode ser explicado pelo fato de que, quando o antecedente está em foco e a anáfora não adiciona novas informações sobre o seu referente, a anáfora é facilmente processada, segundo a previsão da Hipótese da Carga Informacional (ALMOR, 1999).

A quarta previsão da Hipótese da Carga Informacional (ALMOR, 1999) foi corroborada pelos resultados do quinto experimento ao ser demonstrado que as anáforas com antecedente atípico são processadas mais facilmente do que com antecedente típico.

Estes resultados vão de encontro aos resultados de Garrod e Sanford (1977), que encontraram resultados inversamente proporcionais. Almor (1999) afirma que estes resultados podem ter sido encontrados pelo fato de o foco ter sido diretamente manipulado tanto no desenvolvimento deste experimento quanto nos experimentos desenvolvidos por Garrod e Sanford (1977). Além disso, Almor (1999) supõe que estas duas pesquisas mediram o tempo de maneiras distintas, sendo, nesta pesquisa, manipulado do tempo de leitura da região crítica, isto é, a região da anáfora; enquanto naquela pesquisa, pode ter sido estudado o tempo de leitura da sentença inteira, o que ajudaria a explicar os resultados opostos.

Desta forma, ao final de uma sequência de cinco experimentos de leitura automonitorada, Almor (1999) afirma que todas as predições da Hipótese da Carga Informacional foram corroboradas, apontando para o uso psicologicamente motivado das anáforas durante o discurso.

Isso nos leva a entender que o uso de uma anáfora com alta carga informacional em relação ao seu antecedente é justificado apenas quando são acrescentadas novas informações ao referente discursivo ou quando ajuda a identificá-lo.

E, segundo Almor (1999), há três grandes achados reportados neste estudo que embasam a Hipótese da Carga Informacional.

- (i) Sobre a Penalidade do Nome Repetido: anáforas executadas por meio de nomes repetidos são mais lentamente processadas quando seu antecedente está em foco do que quando não está, (ii) anáforas executadas por nomes não repetidos são lidas mais rapidamente quando o seu antecedente está em foco do que quando não está e (iii) anáforas nominais mais gerais do que o antecedente focalizado são lidas mais rapidamente quando elas estiverem mais distantes de seus antecedentes (ALMOR, 1999, p. 762-763, tradução nossa)<sup>93</sup>.

De fato, podemos perceber que noções pragmáticas como a introdução de novas informações e foco discursivo podem motivar a escolha psicológica das anáforas, de acordo com a Hipótese da Carga Informacional (ALMOR, 1999). Almor (1999) defende que a combinação entre função linguística e princípios cognitivos pode constituir a base de uma explicação psicológica de um importante aspecto do uso da linguagem, a ligação do par anáfora-antecedente.

### 2.2.2.3 Teoria do Processamento em Dois Estágios

Levando em consideração o fato de os resultados encontrados por Garrod e Sanford (1977) e Almor (1999) serem inversamente proporcionais e contraditórios entre si, Van Gompel, Liversedge e Pearson (2004) analisaram como se dá o processamento correferencial, tendo como base as predições da Teoria da Sobreposição Semântica (GARROD; SANFORD, 1977 e da Hipótese da Carga Informacional (ALMOR, 1999).

Segundo Van Gompel, Liversedge e Pearson (2004), durante a leitura de um texto, os leitores geralmente processam a informação de uma sentença conjuntamente às informações dadas em outras sentenças, isto é, durante o

---

<sup>93</sup> (i) The Repeat Name Penalty: repeated name RN anaphors are read slower when their antecedent is focused than when it is not focused. (ii) nonrespective NP anaphors are read faster when their antecedent is focused than when it is not focused. (iii) NP anaphors which are more general than their focused antecedent were faster read the more semantically distant they were from their antecedent.

processamento natural do texto não há isolamento de sentenças, sendo o todo textual necessário para a construção da coerência discursiva.

Van Gompel, Liversedge e Pearson (2004) chamam a atenção para o fato de um dos mecanismos mais importantes para o estabelecimento e manutenção da coerência discursiva ficar a cargo da correferência e a relação existente entre os elementos do texto que constituem o par antecedente-anáfora.

Para Van Gompel, Liversedge e Pearson (2004), em um pequeno texto como em (54)<sup>94</sup> (VAN GOMPEL; LIVERSEGE; PEARSON, 2004, p.119, tradução nossa), o leitor precisa entender a relação existente entre a anáfora *o pássaro* e o antecedente *o sabiá*, de modo que seja capaz de perceber que ela trata do mesmo referente.

(54) Um sabiá<sub>i</sub> às vezes entrava na casa. O pássaro<sub>i</sub> foi atraído para a dispensa.

A partir deste exemplo, Van Gompel, Liversedge e Pearson (2004) mostram que se o leitor não for capaz de compreender a relação existente no par anáfora-antecedente, ele não será capaz de construir uma representação coerente do texto.

Tomando como base estudos psicolinguísticos<sup>95</sup>, Van Gompel, Liversedge e Pearson (2004) apontam que diversos fatores estão envolvidos durante o processamento correferencial, sendo um destes fatores a tipicidade<sup>96</sup> da anáfora.

Pautados sobretudo nos resultados do estudo de Garrod e Sanford (1977) e de Almor (1999) sobre a tipicidade da anáfora, Van Gompel, Liversedge e Pearson (2004) trabalham com ideias completamente excludentes entre si, já que, para aqueles, é mais fácil processar uma anáfora mais geral, com maior carga semântica<sup>97</sup> com relação ao seu antecedente, enquanto este prediz justamente o contrário. Segundo Almor (1999), é mais fácil de ser processada a

<sup>94</sup> (54) A robin would sometimes wander into the house. The bird was attracted by the larder.

<sup>95</sup> Muitos dos estudos preliminares que levaram Van Gompel, Liversedge e Pearson a este trabalho foram discutidos ao longo desta tese. Entretanto, para informações mais completas acerca do trabalho, ver Van Gompel, Liversedge e Pearson (2004).

<sup>96</sup> De acordo com Van Gompel, Liversedge e Pearson (2004), a tipicidade do antecedente diz respeito às palavras que são mais comuns a serem retomadas entre si como, por exemplo, ser mais fácil (ou mais comum) retomar o antecedente *o pássaro* por uma anáfora como *o sabiá* do que *o ganso*.

<sup>97</sup> Nos estudos semânticos sobre o processamento, os termos *carga semântica* e *carga informacional* são sinônimos, mas optamos por manter os termos usados por cada autor.

anáfora mais específica, com menor carga semântica com relação ao seu antecedente<sup>98</sup>.

Além de compreender que fatores semânticos agem durante o processamento correferencial, Van Gompel, Liversedge e Pearlson (2004) trabalham com as noções de marcadores linguísticos mais acessíveis, conforme a Teoria da Acessibilidade (ARIEL, 1991) e as previsões da Penalidade do Nome Repetido (GORDON, GROSZ e GUILLIOM, 1993).

A fim de testar as previsões contrastantes da Teoria da Sobreposição Semântica (GARROD e SANFORD, 1977) e da Hipótese da Carga Informacional (ALMOR, 1999), Van Gompel, Liversedge e Pearlson (2004) desenvolveram um experimento utilizando a técnica do rastreamento ocular para investigar o efeito da tipicidade do antecedente durante o processamento da leitura.

Os itens experimentais foram criados com vistas a analisar se o efeito da tipicidade do antecedente ocorre quando o antecedente típico ou atípico era o foco da sentença anterior à anáfora, conforme exemplificado em (55) (VAN GOMPEL; LIVERSEEDGE; PEARLSON, 2004, p.129, tradução nossa)<sup>99</sup>.

(55)

- (a) Foi o carvão que foi encomendado em grandes quantidades antes do inverno.  
Obviamente, o combustível estava com grande procura.
- (b) Foi a turfa que foi encomendada em grandes quantidades antes do inverno.  
Obviamente, o combustível estava com grande procura.
- (c) O carvão foi encomendado em grandes quantidades antes do inverno.  
Obviamente, o combustível estava com grande procura.
- (d) A turfa foi encomendada em grandes quantidades antes do inverno.  
Obviamente, o combustível estava com grande procura.

Tomando como base as sentenças experimentais que levaram à postulação da teoria da Hipótese da Carga Informacional (ALMOR, 1999), Van Gompel, Liversedge e Pearlson (2004) usaram os mesmos tipos de construções

---

<sup>98</sup> Para mais detalhes sobre as pesquisas desenvolvidas por Garrod e Sanford (1977) e Almor (1999), ver a subseção 3.3 deste projeto.

<sup>99</sup>(55) a. It was the coal that was ordered in great quantities before the winter.  
Obviously, the fuel was in great demande.  
b. It was the peat that was ordered in great quantities before the winter.  
Obviously, the fuel was in great demande.  
c. The coal that was ordered in great quantities before the winter.  
Obviously, the fuel was in great demande.  
d. The peat that was ordered in great quantities before the winter.  
Obviously, the fuel was in great demande.

daquele trabalho, colocando os antecedentes sempre em foco e analisando o tempo de leitura tanto de partes da sentença como o tempo total<sup>100</sup>, já que uma das críticas aos resultados tão dispares entre o trabalho de Almor (1999) e Garrod e Sanford (1977) é o fato de como foram analisados os tempos de leitura, se de regiões específicas da sentença, sobretudo o tempo de processamento da região da anáfora, ou se do texto inteiro.

Os resultados encontrados pelos pesquisadores indicam dois tipos de processos gerando efeitos durante o processamento da anáfora: durante um estágio inicial do processamento, as anáforas com antecedentes mais típicos são mais fáceis de processar do que as anáforas com antecedentes atípicos, mas durante um estágio subsequente do processamento, o padrão é invertido.

De acordo com Van Gompel, Liversedge e Pearlson (2004), a vantagem das anáforas com antecedentes típicos sendo formas mais fáceis de serem processadas pode ser explicada de duas maneiras: (i) o efeito é o resultado de um efeito de *priming* léxico-semântico entre o antecedente e a anáfora e, quando o antecedente é uma forma típica para a anáfora, a sobreposição semântica entre os dois nomes é maior do que quando é usada uma forma atípica. Contudo, aceitar estas explicações implica que o efeito da tipicidade do antecedente é um efeito puramente lexical.

E, (ii) o efeito da tipicidade pode ser o reflexo inicial da identificação do antecedente, o que poderia explicar a sobreposição semântica que ocorre entre a anáfora e o antecedente quando a relação de tipicidade entre o par é grande, sendo a forma mais típica mais facilmente de ser processada do que quando o antecedente é atípico.

Esta explicação é próxima da explicação dada por Garrod e Sanford (1977), mas não ajuda a entender o que ocorre quando o efeito se torna inverso, o que leva Van Gompel, Liversedge e Pearlson (2004) a acreditar que, no estágio final do processamento, há um efeito de super especificação semântica.

De acordo com Van Gompel, Liversedge e Pearlson (2004), isto implica dizer que, depois que o processador<sup>101</sup> estabelece a relação de referência, ele

---

<sup>100</sup> Van Gompel, Liversedge e Pearlson (2004) analisaram os tempos de leitura em quatro regiões específicas da sentença (a região precedente à anáfora, a região da anáfora, a região subsequente à anáfora e a região final da sentença) e, com base nos dados da soma dos tempos de leitura de cada região, analisaram também o tempo de leitura do texto inteiro.

<sup>101</sup> É importante notar que Van Gompel, Liversedge e Pearlson (2004) trabalham com o conceito de processador (*parser*) como Garrod e Sanford (1977) e não com o conceito de memória e os efeitos que o uso de uma anáfora gera à memória de trabalho, como Almor (1999).

verifica se a anáfora é semanticamente apropriada para ser usada como expressão a retomar o antecedente. Se a anáfora for muito semelhante ao seu antecedente, ela passa a ser recusada pelo processador pelo fato de as duas expressões de referência fornecerem informações muito semelhantes sobre a entidade discursiva a que elas se referem, o que já se aproxima da explicação de Almor (1999) para o efeito da tipicidade do termo antecedente.

Contudo, Van Gompel, Liversedge e Pearlson (2004), apesar de terem os resultados de sua pesquisa próximos aos resultados encontrados por Garrod e Sanford (1977) e por Almor (1999), propõem um conceito de processamento que levou à Teoria do Processamento Correferencial em Dois Estágios:

De acordo com nossa descrição, o processador anafórico opera em (pelo menos) dois estágios: um estágio de identificação do antecedente e um estágio subsequente durante o qual a forma semântica da anáfora é verificada. (VAN GOMPEL, LIVERSEGE e PEARSON, 2004, p.133, tradução nossa)<sup>102</sup>.

É a concepção do processamento ocorrendo como dois processos distintos que faz com que Van Gompel, Liversedge e Pearlson (2004) prevejam, inicialmente, efeitos de ligação durante a identificação do par anáfora-antecedente<sup>103</sup> e, posteriormente, a observância dos requisitos semânticos necessários para o estabelecimento da coerência do par.

Assim, ao levarmos em consideração as descobertas da interpretação do par anáfora-antecedente por meio da observação dos traços semânticos, temos de analisar qual é a forma mais fácil de processar entre o par e, temos resultados sobre a ligação entre uma entidade sendo introduzida por meio de um nome e sendo retomada, também, por meio de um nome.

Contudo, os estudos que trabalham com questões semânticas, sobretudo o estudo de Almor (1999), levam a crer que, para se entender os critérios semânticos que geram efeitos no processamento correferencial, é preciso aceitar, também, que questões de cunho pragmático podem influenciar em como as formas do par anáfora-antecedente são usadas pelos falantes de uma língua.

---

<sup>102</sup> According to our account, the anaphoric processor operates in (at least) two stages: An antecedent identification stage and a subsequent stage during which the semantic felicity of the anaphor is checked.

<sup>103</sup> Ao longo do estudo, Van Gompel, Liversedge e Pearson (2004) não deixam explícitos se os efeitos de ligação aos quais eles se referem são de cunho sintático, como nos estudos de Gordon, Grosz e Guilliom (1993), ou de cunho pragmático, como no estudo de Ariel (1991) e de Almor (1999).

### **2.2.3 Teoria pragmática sobre o processamento correferencial**

As teorias e hipóteses sobre o processamento correferencial têm sido observadas, majoritariamente, por meio de análise das questões de cunho sintático que influenciam na ligação e interpretação do par anáfora-antecedente<sup>104</sup>.

Para além dos estudos de cunho sintático, há trabalhos que analisam o quão típico ou não podem ser as formas que funcionam como antecedente e correferente por meio dos traços semânticos que ligam o par, sobretudo na ligação nome-nome<sup>105</sup>.

Porém, os trabalhos que tratam das questões semânticas que podem influenciar o processamento correferencial, sobretudo o trabalho de Almor (1999), acreditam que pode haver influência da pragmática durante o processamento correferencial, o que leva em consideração a acessibilidade dos termos correferentes, de acordo com as propostas de Ariel (1991) ao postular a Teoria da Acessibilidade.

#### **2.2.3.1 Teoria da Acessibilidade**

Com o propósito de apresentar uma proposta sólida acerca de como ocorre o processo do acesso mental feito por falantes nativos de dada língua durante a manutenção das entidades ao longo do discurso, Ariel (1991) propõe a Teoria da Acessibilidade como forma de explicar a maneira como os SNs são retomados, a fim de estabelecer a relação entre antecedente e retomada e, assim, manter a coerência discursiva.

De acordo com Ariel (1991), as representações mentais das entidades introduzidas no discurso e que devem, posteriormente, e durante o discurso ser retomadas, não são acessadas na mente do usuário da língua da mesma maneira.

Para Ariel (1991), as formas utilizadas pelo falante para recuperar uma entidade já introduzida no discurso variam de acordo com o grau de

---

<sup>104</sup> Para mais detalhes sobre as pesquisas desenvolvidas a partir de critérios sintáticos envolvidos durante o processamento correferencial, ver a subseção 2.1 desta tese.

<sup>105</sup> Para mais detalhes sobre as pesquisas desenvolvidas a partir de critérios semânticos envolvidos durante o processamento correferencial, ver a subseção 2.2 desta tese.

acessibilidade do marcador linguístico<sup>106</sup> que é tomado como forma correferente ao longo do texto e que serve para manter em foco o tópico retomado, além de auxiliar o falante na recuperação desta entidade quando acessa às informações em sua memória de trabalho.

Na busca de determinar quais os fatores geradores destes efeitos sobre como o falante pode acessar os marcadores linguísticos em sua memória de trabalho, Ariel (1991) determina que o grau de acessibilidade depende de três fatores: (i) informatividade, (ii) rigidez e (iii) atenuação.

Segundo Ariel (1991), os critérios de análise do grau de acessibilidade dos marcadores linguísticos levam a uma observação proporcionalmente inversa. Assim, um marcador será mais acessível quando ele for mais informativo, mais rígido e menos atenuado e, do contrário, menos acessível ele será quando for menos informativo, menos rígido e mais atenuado.

O primeiro critério elencado por Ariel (1991), a informatividade, aponta para o fato de que, dada a entrada de uma entidade em um texto, a forma lexical mais informativa e de retomada preferencial ao longo do discurso será aquela que contém menor quantidade de material lexical apontando para outra entidade. Logo, será preferencialmente usado como marcador linguístico correferente à entidade previamente introduzida a forma linguística que contém menor informação lexical independente, o que pode apontar para as marcações feitas pelo uso de um pronome.

Ariel (1991) mostra que o segundo critério, a rigidez, diz respeito ao fato de certos marcadores linguísticos estarem mais presos a determinadas construções contextuais, sendo necessário atentar que, para certos contextos, parece haver melhor acessibilidade a determinados marcadores do que a outros, como ocorre ao se comparar a recuperação de nomes próprios ou descrições definidas.

Por fim, Ariel (1991) diz que o terceiro critério, a atenuação, é um critério problemático pelo fato de não ser o critério de atenuação que marca a acessibilidade, mas por desempenhar efeito sobre ela.

Neste critério, Ariel (1991) afirma aproximar-se do conceito de tamanho fonológico, conforme proposto por Givón (1983). Ariel (1991) diz que, para determinar o grau de acessibilidade de um marcador linguístico na memória de

---

<sup>106</sup> Ariel (1991) chama de marcador a forma linguística usada ao longo do texto, como nomes ou pronomes, por exemplo.

trabalho do falante, o tamanho da palavra usada como forma de retomada do termo antecedente não desempenha o papel de marcador de acessibilidade, mas sim a atenuação da pronúncia.

Por meio da observação destes três critérios, Ariel (1991) aponta que, a partir da análise de cada um separadamente, haverá formas linguísticas menos ou mais marcadas a serem acessadas na mente do falante. Contudo, Ariel (1991) propõe uma escala geral de acessibilidade e, para esta escala, ela afirma haver um padrão universal, a variar entre as línguas, mas sendo seguida, em geral, uma gradação das formas com maior ou menor acessibilidade

Após as investigações empíricas desenvolvidas por Ariel (1991) que levaram à formulação da Teoria da Acessibilidade, é interessante ter em mente afirmações que tratam especificamente do que ocorre com os pronomes, entendidos como marcadores linguísticos a serem buscados na memória de trabalho dos falantes, pois, de acordo com a autora, dentro dos tipos de pronomes, é possível observar graus diferentes de acessibilidade.

Em diferentes línguas, os pronomes<sup>107</sup>, por exemplo, podem se comportar de maneiras diferentes. O que é constante é o fato de eles sinalizarem maior acessibilidade do que os demonstrativos e menor acessibilidade do que os zeros (ARIEL, 1991, p. 451, tradução nossa)<sup>108</sup>.

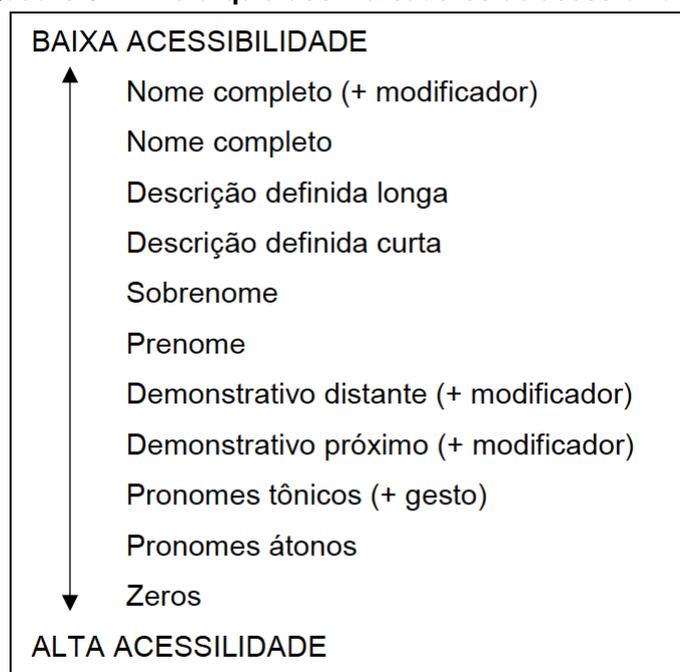
Mostraremos, a seguir, o quadro com a hierarquia dos marcadores linguísticos, de acordo com a possibilidade do menos acessível ao mais acessível, de acordo com Ariel (1991).

---

<sup>107</sup> Para entender melhor a distinção entre demonstrativos, pronomes e zeros, de acordo com o proposto pela Teoria da Acessibilidade, ver Ariel (2004).

<sup>108</sup> In different languages, pronouns, for example, behave differently. What is constant in that they signal higher Accessibility than demonstratives and lower Accessibility than zeros.

**Quadro 04 – Hierarquia dos marcadores de acessibilidade**



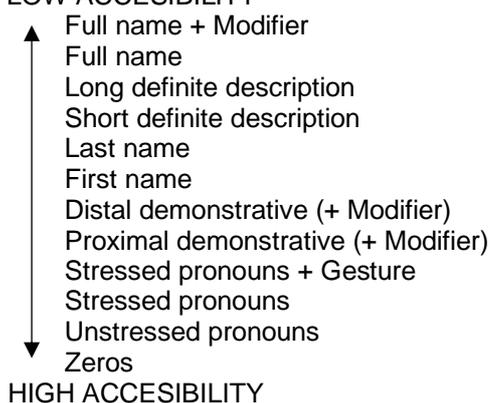
(Fonte: ARIEL, 1991, p.449, tradução nossa)

É possível notar que, durante a postulação do quadro de hierarquizados marcadores de acessibilidade em sua teoria, conforme visto no quadro acima<sup>109</sup>, Ariel (1991) previu as formas que retomam o antecedente de acordo com a possibilidade de, em línguas quaisquer, haver variações entre as formas, gerando uma hierarquia dos marcadores de acordo com relação à acessibilidade na mente do falante, conforme veremos na figura a seguir.

Além disso, ao longo das discussões que Ariel (1991) promove durante a construção de sua teoria, ela chama a atenção para o fato de que, com relação à distribuição gramatical entre pronomes e zeros, o conceito de Princípio de Evite Pronome (CHOMSKY, 1981) traz em si um pequeno problema.

Este problema, segundo Ariel (1991), ocorre pelo fato de “não ser o uso do zero que necessita de um princípio especial para contabilizá-lo. O problema

<sup>109</sup> LOW ACCESIBILITY



surge quando os pronomes *plenos* são preferidos” (ARIEL, 1991, p. 453, tradução nossa, grifo nosso)<sup>110</sup>, o que, de fato, seria o caso que necessitaria de observação e explicação.

O problema com o Princípio Evitar Pronome (assim como as outras teorias gramáticas quando não mediadas pela Teoria da Acessibilidade) é a previsão de que as diferenças são unidirecionais (apenas os zeros são favorecidos pelos pronomes, nunca o contrário). Isto não é verdade no discurso real (ARIEL, 1991, p.453, tradução nossa)<sup>111</sup>.

A partir de então, Ariel (1991) afirma que os problemas com as teorias de cunho gerativista estão no fato de os estudiosos debruçarem-se apenas na resolução das categorias vazias e não darem atenção aos sujeitos nulos, pois, a explicação de como estes são retomados ao longo do discurso envolve, de alguma maneira, questões de cunho pragmático.<sup>112</sup>

Além da discussão feita acerca das formas anafóricas pronominais, Ariel (1991) mostra em sua pesquisa que os pronomes, quando comparados com descrições definidas, tendem a ser preferencialmente usados dentro das mesmas orações. Já o uso do pronome como a forma que retoma o tópico no início de um parágrafo tende a ser menor do que o uso da descrição definida, o que pode indicar que o falante precise atualizar o tópico discursivo ao longo da leitura, já que o tópico tendo a ser retomado por diversas formas linguísticas, obedecendo a gradação proposta na Teoria da Acessibilidade<sup>113</sup>.

Assim, Ariel (1991) propõe que a forma como os falantes acessam certas informações na memória de trabalho depende do grau de acessibilidade que dado marcador linguístico tem dentro do contexto de uso.

Contudo, dentro da discussão que levou à formulação da Teoria da Acessibilidade, a autora não deixa de considerar que há outros fatores que

---

<sup>110</sup> No need for a special principle in order to account for the zero choices. A problem does arise when pronouns are preferred.

<sup>111</sup> The problem with the Avoid Pronoun principle (as well as the other theories, when not mediated by Accessibility theory) is the prediction that preferences are unidirectional (only zeros are favored over pronouns, never the reverse). This is not true in actual discourse.

<sup>112</sup> Ao longo da discussão feita no trabalho que deu origem à Teoria da Acessibilidade, Ariel (2001) afirma que, se feito um quadro comparativo entre teorias sintáticas, semânticas e pragmáticas, a sua enquadra-se melhor como uma teoria de cunho semântico-pragmática e mais voltada ainda para a Pragmática do que a para a Semântica. Contudo, ela não deixa de discutir padrões gramaticais que a levaram a tais afirmações.

<sup>113</sup> A maneira como estão dispostos demais marcadores linguísticos, conforme proposto por Ariel (2001), não aparecem nesta discussão por não serem de interesse para o desenvolvimento desta tese.

podem determinar o quão acessível é determinada forma, como é o caso dos traços de animacidade ou a saliência discursiva que parece ser inerente a algumas entidades, o que pode contribuir para o estabelecimento da correferência.

Contudo, muitas das pesquisas citadas ao longo desse capítulo tem buscado dar explicações de como a mente processa a informação de acordo com a expressão anafórica que é utilizada em um dado texto, sendo, portanto, necessário compreender melhor a memória e os tipos de memória que usamos para desenvolver determinadas atividades.

## **2.3 Memória**

Nesta seção, descreveremos como a memória está intrinsecamente ligada a atividades cognitivas, sobretudo quando o indivíduo está em execução de uma tarefa que demande atenção.

Por conta disto, mostraremos o conceito de memória, os tipos de memória e seus componentes, além de apontar as diferenças entre o modelo clássico (BADDELEY; HITCH, 1974) e o modelo atual (BADELLEY, 2000, 2007, 2017) de memória de trabalho, que será de grande valia para a nossa pesquisa.

Pelo fato de nosso trabalho ser pautado em textos produzidos por usuários da língua em contextos autênticos, acreditamos que seja necessário trazer o conceito de memória de trabalho para auxiliar as explicações que possamos vir a dar a partir dos resultados encontrados por meio da pesquisa e, com isso, sermos capazes de dar possíveis evidências de como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos são distribuídos em Português Brasileiro quando desempenham função anafórica na posição de sujeito.

### **2.3.1 O conceito de memória**

As diversas áreas cerebrais são responsáveis por cada tipo de resposta dos seres vivos para os estímulos ambientais aos quais entram em contato. Entretanto, o cérebro de cada espécie, apesar de ter semelhanças entre si, difere, principalmente, no tipo de resposta enviada tanto para as outras áreas da mente quanto para o corpo (WILSON, 2009).

Na literatura clínica (BADDELEY; KOPEMAN; WILSON, 2003; 2004; WILSON, 2009; LAPOINTE, 2012; BADDELEY, 2007, 2013, 2017), além de todas as regiões cerebrais mapeadas, sabe-se que a área medial do cérebro é onde podem ser encontradas áreas distintas do córtex circundante. E, é nesta área onde os neurônios que fazem as sinapses que são responsáveis pelas informações relacionadas ao comportamento sexual e emocional, bem como pelo processamento da memória.

Sabendo que, durante o desenvolvimento do feto<sup>114</sup> e ao longo de toda a sua vida, quando os *inputs* cerebrais são enviados ao encéfalo por meio da interação com o meio ambiente geram o aprendizado e, o aprendizado, por sua vez, depende da memória, que é a base da individualidade.

Contudo, é necessário entender que (i) o aprendizado está intimamente relacionado com (ii) a memória, sendo que (i) é entendido como a aquisição de novos conhecimentos e (ii) como a responsável pela retenção do aprendizado.

O papel básico da memória é codificar os conhecimentos, armazená-los e evocá-los (BADELLEY, 2013; BADDELEY; EYSENCK; ANDERSON, 2014), mas, além disso, é a memória a responsável pela individualidade ao passo que cada pessoa recorda de informações diferentes das lembranças de outras pessoas, mesmo que diga respeito a uma situação em comum entre alguns indivíduos (WILSON, 2009).

No entanto, apesar da individualidade das memórias de cada pessoa, todos recordam fatos, eventos, emoções e desempenhos por um período curto de tempo, ou mesmo por toda a vida.

As lembranças armazenadas na memória dos indivíduos são determinantes para a formação da personalidade e acabam influenciando o âmbito cognitivo, pois é a partir da evocação das memórias que a vida adquire significado.

De acordo com Axmacher e Rasch (2017), sem a memória, não existiria uma identidade particular a cada ser, pois não existiriam recordações das pessoas mais próximas e das habilidades que são usadas no cotidiano, tendo

---

<sup>114</sup> Biologicamente, sabe-se que o desenvolvimento da memória ocorre durante a fase pré-natal e que, após o nascimento, ela desempenha umas das funções mais básicas da mente, pois cabe à memória a determinação de como o indivíduo se desenvolverá ao longo de sua vida. É a memória o módulo cerebral responsável por permitir que o indivíduo adquira novos conhecimentos e elabore estratégias para resolver desde os mais simples aos mais complexos problemas cognitivos, juntamente com a maturação do sistema nervoso central (BADDELEY, 2000; IZQUIERDO, 2011)

assim que serem constantemente reaprendidas ou, talvez, nem mesmo aprendidas, pois a capacidade de memorização é importante para qualquer processo de obtenção e elaboração de informação.

Assim, entendemos que a cognição e a memória estão intimamente ligadas, pois não há aprendizado sem capacidade de armazenamento na memória e, sobretudo, acreditamos que não haja processamento de informações sem acesso à memória durante qualquer atividade na vida dos indivíduos.

### **2.3.2 A retenção da informação e os tipos de memória**

Sabendo que os seres humanos não seriam capazes de pensar a não ser por meio de dados que são enviados ao cérebro para que as informações sejam retidas e resgatadas quando necessário, é preciso compreender que, iniciar o processo de envio de informações ao cérebro, depara-se com a percepção, que é o primeiro passo para a formação das memórias<sup>115</sup>.

Os olhos, por exemplo, são receptores que captam imagens luminosas e as guardam durante alguns segundos na retina, além de serem os responsáveis por armazenar a percepção visual em frações de segundo no máximo possível de detalhes (GROOME; EYSENCK, 2016).

Estes detalhes são enviados ao encéfalo que analisa o estímulo quanto ao seu significado e importância, descartando as informações desnecessárias, seleção esta feita pelo sistema reticular para impedir que o estímulo chegue ao cérebro com excesso de informações que tornariam incapaz a compreensão. Uma vez concluída a seleção de informações, são liberadas reações emocionais antes que o córtex conclua a percepção.

Quando concluída a percepção, as informações entram em contato com a memória, integrando os seus mais diversos módulos para que todos os inputs sejam processados, armazenados ou descartados.

É por isso que entendemos a memória como estando relacionada à maneira que o indivíduo usa para colocar algo novo na mente, o tempo de retenção de uma informação contida na memória e, ainda, a maneira como esta informação pode ser extraída para o uso que o indivíduo deseja fazer.

---

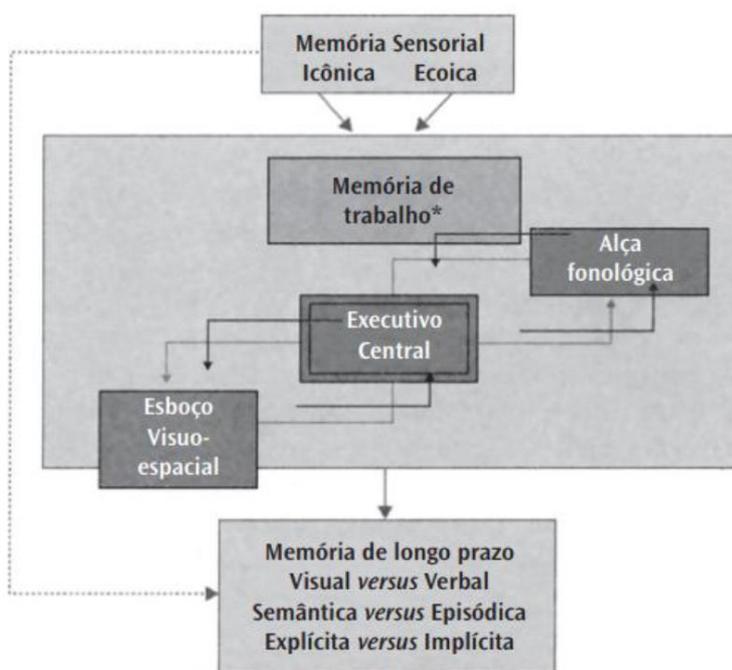
<sup>115</sup> Como as lembranças provêm das experiências que o indivíduo tem ao longo da vida, pode-se dizer que existem diferentes tipos de memórias e cada tipo de memória envolve diferentes sistemas neurais, que serão classificadas de acordo com o seu conteúdo e sua duração (BADDELEY, 2000).

Por conta disso, de acordo com Wilson (2009) quatro aspectos devem ser levados em conta por serem determinantes para que se entenda que tipo de memória está em uso durante uma atividade cognitiva, como a leitura, por exemplo.

Estes aspectos dizem respeito ao *input* ou como o material é recebido durante a percepção, quais as informações que devem ser mantidas na memória, por quanto tempo podem ser mantidas e como se dá a recuperação de um conhecimento ou de uma recordação na memória quando necessário, já que cada tarefa é armazenada e processada por um tipo diferente de memória durante atividades que demandem o envio/reconhecimento/evocação de informações ao cérebro<sup>116</sup>.

O conceito dos diferentes tipos de memória, que funcionam como módulos do componente da memória cerebral, levou Baddeley e Hitch (1974) a proporem o modelo da memória de trabalho.

**Figura 01:** O modelo clássico da memória de trabalho



(Fonte: Baddeley; Hitch, 1974)

<sup>116</sup> A ideia que se tem sobre a memória é de que há, em nossa mente, um local a ser armazenado todo o tipo de informação e, a este local, damos o nome de memória. Entretanto, uma informação, quando chega ao cérebro, passa a ser analisada sobre como deve ser guardada para, posteriormente, ser evocada ou não, fazendo com que haja sublocais de memória, dentro do campo da memória e, cada tipo é responsável por armazenamento e processamento de maneiras diferentes.

De antemão, é preciso entender o fato de que todas as informações enviadas ao cérebro passam por uma seleção do que deve ser guardado e do que deve ser descartado para não sobrecarregar a memória (IZQUIERDO, 2011; LAPOINTE, 2012; KEMERER, 2014), pois uma informação visual demanda componentes da memória que são distintos dos componentes demandados por uma informação auditiva ou tátil, por exemplo, e, por isso, é necessário que tenhamos o mínimo de conhecimento acerca dos tipos de memória.

De acordo com Baddeley e Hitch (1974), a memória poderia ser dividida em três categorias, levando em consideração o tempo que a mente leva para armazenar na memória as informações recebidas sendo, pois, (i) a memória sensorial (250 ms), (ii) a memória de curto prazo (retém a informação por alguns segundos) e (iii) a memória de longo prazo (retém a informação de minutos a anos).

Segundo Wilson (2009, p. 22), a memória sensorial “é a que usamos quando vamos ao cinema e percebemos o que parece ser uma imagem em movimento”, pois vemos, de fato, uma série de fotos que são mantidas por curtíssimos espaços de tempo na memória e, então, são processadas como imagens em movimento.

Compreender o movimento envolve a chamada memória icônica<sup>117</sup>, enquanto a memória envolvida em tarefas com estímulos auditivos é chamada de memória ecoica (BADDELEY, 2002, 2007; WILSON, 2009; GROOME; EYSENCK, 2014).

Já a memória de curto prazo<sup>118</sup> (MCP), responsável por manter uma informação por alguns segundos na memória, é o que, segundo Wilson (2009, p.23) “usamos quando discamos um número de telefone com sete dígitos, por exemplo”, já que somos capazes de estar atentos aos números por tempo o suficiente para discar.

Contudo, se levarmos em consideração que a memória de curto prazo nos permite guardar a sequência de números do telefone por alguns minutos, é fácil compreender o motivo de, caso a pessoa para a qual estejamos telefonando não atenda ou se, por acaso, o telefone estiver ocupado durante muito tempo,

---

<sup>117</sup> Na literatura clínica, pode-se encontrar também o termo *memória visual* como sinônimo para *memória icônica*.

<sup>118</sup> Podem-se encontrar os termos *memória imediata*, *memória primária* ou *memória de trabalho* como sinônimos para *memória de curto prazo*.

termos de olhar em alguma anotação ou perguntar para alguém o número quando formos discar novamente.

Pelo fato de a MCP ser uma memória de uso imediato e que está diretamente ligada ao raciocínio ou à execução de uma tarefa<sup>119</sup>, é comum ver os pesquisadores se referirem a este tipo de memória como memória de trabalho (WILSON, 2009; LAPOINTE, 2012; KEREMER, 2014), enquanto outros preferem defender que a memória de trabalho é um dos componentes da MCP (WARD, 2015; GROOME; EYSENCK, 2016; AXMACHER; RASH, 2017).

O fato é que, *grosso modo*, os estudiosos da memória concordam que a MCP (i) atua no momento em que a informação está sendo adquirida e (ii) depende do perfeito funcionamento do córtex pré-frontal.

Já a memória de longo prazo (MLP) serve para armazenar as informações por longos períodos que podem variar de minutos a décadas.

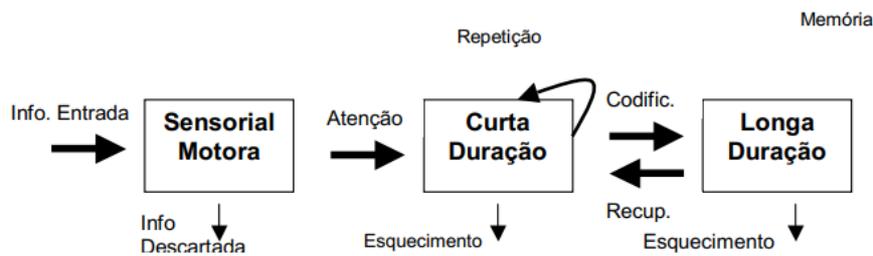
De acordo com Wilson (2009), o meio pelo qual as informações são armazenadas tanto na MCP quanto na MLP parece ser o mesmo. Entretanto, isso não significa que as pessoas podem se lembrar de eventos acontecidos anos atrás da mesma maneira que se lembram de eventos que aconteceram recentemente e, de acordo com a pesquisadora, não há nenhuma distinção clara sobre o motivo que leva a este tipo de evocação ser tão distinta entre si.

O fato consensual da área é que, para a consolidação de uma recordação ou de um conhecimento pela memória envolve desde a entrada do estímulo na memória sensorial, sendo descartadas as informações desnecessárias após a seleção que ocorre em milésimos de segundos. Depois do descarte, é preciso ter atenção nos *inputs* ambientais para incorporar a informação na MCP até que o indivíduo seja capaz de codificar e armazenar na MLP. Uma vez armazenada a recordação ou o conhecimento na MLP, o indivíduo é capaz de recuperar a informação quando necessário pela MCP (por isso que é, para muitos, a própria memória de trabalho) e, caso não haja recuperação das informações, cair em esquecimento, conforme a figura abaixo.

---

<sup>119</sup> Segundo Goldman-Rakic (1995), a memória de trabalho armazena informações durante o tempo em que está sendo realizada uma determinada tarefa e só acontece pelo fato um grupo de neurônio ficar disparando potenciais de ação durante alguns segundos, retraindo temporariamente a informação somente durante o tempo em que ela for necessária e descartando-a em seguida. Para o autor, este tipo de atividade cerebral é efêmera e não forma traços bioquímicos.

Figura 02: Sistema de memorização



(Fonte: Dividino; Faigle, 2014)

Na literatura clínica<sup>120</sup>, é comum haver distinções entre os tipos de memória quando há referência a MLP, sendo usados os termos *memória tardia*, *memória recente* e *memória remota*<sup>121</sup> para localizar, de acordo com o tempo ao qual a memória remonta, o tipo de memória evocada durante uma tarefa.

Como não há total concordância entre os estudiosos da área sobre o fato de a memória de trabalho ser ou não distinta da MCP, consideraremos que ambas desempenham a mesma função e, para além da nomenclatura, é necessário que entendamos quais são as distinções que há entre estes dois a memória de curto prazo e a memória de longo prazo e seus componentes, descreveremos a seguir a distinção básica entre a memória de trabalho<sup>122</sup> e a memória de longo prazo.

### 2.3.3 Memória de trabalho e memória de longo prazo

Para muitos autores (BADDELEY, 2013, 2007, 2018; BADDELEY; EYSENCK; BADDELEY; KOPELMAN; WILSON, 2005; WILSON, 2009; LAPOINTE, 2012; KEREMER, 2014) está cada vez mais claro que a memória de curto prazo e a memória de trabalho é, senão o mesmo tipo de memória, parte integrante e indissociável entre si.

O modelo de memória de trabalho, conforme inicialmente proposto por Baddeley e Hitch (1974), contempla três componentes interligados entre si, no qual há um componente responsável pelo controle das informações que chegam

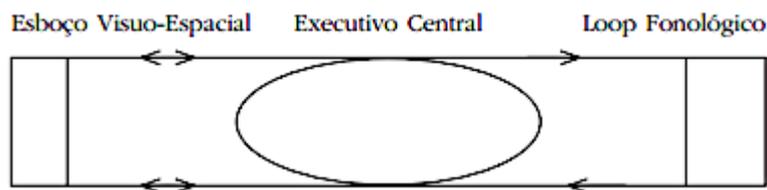
<sup>120</sup> Ver Baddeley, Kopelman e Wilson (2004), Wilson (2009), Kemerer (2014), Ward (2015) e Baddeley (2017).

<sup>121</sup> “Memória tardia é a que se refere a eventos ou informações que ocorreram alguns minutos antes, memória recente a que se refere a coisas que aconteceram há poucos dias ou semanas e memória remota a que se refere a eventos ou informações que aconteceram anos antes” (WILSON, 2009, p 25).

<sup>122</sup> Optamos por usar o termo memória de trabalho como correspondente ao termo memória de curto prazo, entendendo os dois tipos como um só, conforme afirmam WILSON, 2009; LAPOINTE, 2012 e KEREMER, 2014.

à memória – o sistema executivo central – e dois componentes dependentes das informações controladas feitas pelo regulador – o esboço espaço-visual e a alça fonológica.

**Figura 03:** Primeiro modelo de memória de trabalho



(Fonte: Baddeley; Hitch, 1974)

Contudo, ao longo dos anos, o modelo de memória de trabalho sofreu alterações em sua estrutura (BADDELEY; HITCH, 1974; BADDELEY, 2000) e passou a ser estudada até o modelo proposto atualmente (BADELLEY, 2017) não como composta por três, mas por quatro componentes: o executivo central, a alça fonológica, o esboço espaço-visual e o retentor episódico<sup>123</sup>.

Segundo Baddeley (2000, 2007, 2017), a função do executivo central é (i) focar a atenção do indivíduo em informações relevantes do *input*, descartando as informações irrelevantes, (ii) coordenar múltiplas atividades cognitivas que estejam ocorrendo ao mesmo tempo, (iii) selecionar e executar estratégias de memorização, (iv) direcionar o módulo da memória para o qual será destinada a informação e (v) recuperar informações armazenadas na memória de longo prazo.

Já a alça fonológica<sup>124</sup> é o componente responsável por armazenar e processar as informações codificadas verbalmente sejam elas provenientes de *inputs* visuais ou auditivos.

A alça fonológica é responsável por transformar o conhecimento e as lembranças em fonemas e em palavras a partir da combinação dos códigos fonológicos armazenados em dois subcomponentes: a memória fonológica de curto prazo<sup>125</sup> e o ensaio articulatorio subvocal.

<sup>123</sup> Mais comumente encontrado como *buffer* episódico na literatura da área.

<sup>124</sup> Também conhecido como *loop* fonológico.

<sup>125</sup> A memória fonológica de curto prazo armazena as informações verbais, escritas ou faladas, enquanto o ensaio articulatorio subvocal permite resgatar informações verbais não mais usadas pelo indivíduo.

O esboço espaço-visual realiza o processamento e a manutenção de informações visuais e espaciais dos objetos no mundo<sup>126</sup>, bem como as relações espaciais entre eles, sendo também o componente responsável pela formação e manipulação das imagens mentais (BADDELEY, 2007, 2017)

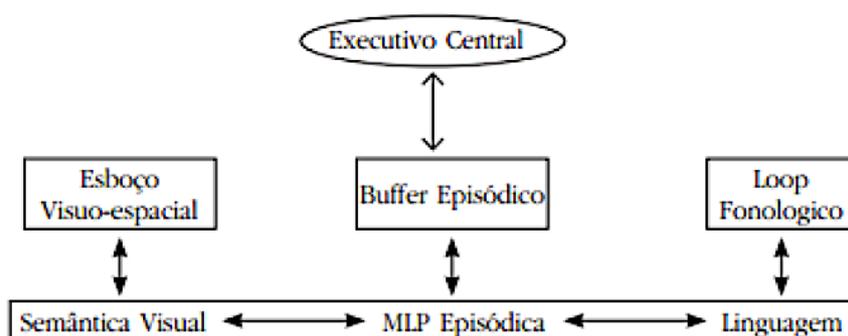
É graças ao esboço espaço-visual que o indivíduo é capaz de se localizar e planejar seus movimentos através da constante atualização das informações mandadas ao cérebro através da percepção de novas informações espaço-visuais.

O retentor episódico é um componente de armazenamento temporário e com capacidade limitada, acessível à consciência e capaz de trocar informações com a memória de longo prazo episódica e com a memória de longo prazo semântica na construção das representações mentais dadas por novas informações.

O papel do retentor episódico é, segundo Baddeley (2007, 2017) gerenciar uma grande quantidade de informação que pode ultrapassar a capacidade da alça fonológica e/ou do esboço espaço-visual sem depender do executivo central.

Assim, podemos entender que, mesmo sendo um dos componentes da memória de trabalho, o retentor episódico é o único componente capaz de desempenhar função tão importante quanto o executivo central, mas não faz o mesmo papel que este componente.

**Figura 04:** Modelo atual de memória de trabalho



(Fonte: Baddeley, 2017)

<sup>126</sup> É necessário entender que todas as informações dadas tanto pela alça fonológica quanto pelo esboço espaço-visual são informações temporárias e com representação mental na consciência dos indivíduos.

O interessante sobre a dependência dos componentes da memória de trabalho está em como cada um deles vai se desenvolvendo, pois há evidências de que a alça fonológica e o esboço espaço-visual se formam de maneira independente, mas, com o amadurecimento do executivo central, os dois componentes passam a ter um alto grau de interdependência, ficando a cabo do executivo central mediar a comunicação entre eles (BADDELEY, 2017).

O fato é que a memória de trabalho desempenha papel crucial durante o processamento de atividades cognitivas complexas, tais como a aprendizagem, o raciocínio e a compreensão da linguagem e, quaisquer falhas nesse sistema podem gerar prejuízos no processo de aprendizagem, na leitura e compreensão de textos e na resolução de problemas matemáticos (ALLOWAY, 2006).

No que diz respeito à compreensão da linguagem, há evidências de que limitações associadas ao processamento, ao armazenamento ou à manipulação de informações pela memória de trabalho podem gerar incapacidades de ordem oral, leitura ou escrita (CIASCA, 2003).

Segundo Souza e Sisto (2011), as pessoas com dificuldades do processamento de informações da linguagem na memória de trabalho<sup>127</sup> têm dificuldades em organizar de forma adequada as informações recebidas, assim como podem apresentar problemas para evocar de outros módulos da memória as informações necessárias para a execução de uma tarefa.

Há evidências também de que problemas associados à memória de trabalho podem gerar dificuldades de compreensão textual, uma vez que ler envolve a capacidade de decodificar palavras e dar coerência às informações trazidas por cada texto e integrá-las aos conhecimentos já armazenados na memória de longo prazo, dependendo, neste caso, da alça fonológica.

As limitações da memória de trabalho podem estar relacionadas a prejuízos no processamento da linguagem (DANEMAN; CARPENTER, 1980; DE JONG, 2006; KEMERER, 2014), já que, para executar tarefas de compreensão leitura, segmentação sintática e construção de inferências, pois a integração das informações na memória de trabalho necessita de muitos recursos de cada um dos componentes e, assim, caso haja ineficiência por parte de qualquer um deles, o funcionamento da memória de trabalho fica comprometido e os recursos

---

<sup>127</sup> As dificuldades de aprendizagem com relação a problemas na memória de trabalho estão associadas aos déficits cognitivos correspondentes à esquizofrenia e TDHA e pode ter sua atividade regulada pela ação de dopamina sobre o córtex pré-frontal (BADDELEY, KOPELMAN e WILSON, 2005; LAPOINTE, 2012; KEREMER, 2014; AXMACHER e RASH, 2017).

disponíveis podem não dar conta de fazer com que a tarefa seja executada como deveria.

A grande questão é que a memória de trabalho representa um modelo que funciona como interface entre o processamento e armazenamento temporário de informações que podem ser guardadas na memória de longo prazo ou descartadas, sendo ela a responsável por todas as atividades cognitivas complexas e que demandam milésimos de segundos para serem executadas e processadas e, então, tornarem-se conhecimento adquirido.

Para que as informações se tornem conhecimento adquirido, é necessário que, por meio de repetição, elas sejam codificadas e armazenadas na MLP, que pode enviar para a MCP as informações sempre que necessárias<sup>128</sup>.

É papel da memória de longo prazo servir como arquivo de informações consolidadas, podendo armazená-las por longos períodos de tempo, que vão desde minutos e semanas a décadas, podendo ser exemplos de conhecimentos armazenados na MLP os conhecimentos adquiridos na escola durante a infância e lembranças de um evento ocorrido há bastante tempo.

Torna-se importante frisar que, diferentemente da memória de curto prazo, a capacidade de armazenamento de informações na MLP é ilimitado e é ela que contém as informações que dispomos de maneira mais ou menos permanente (WILSON, 2009). Além disso, é sabido que o conhecimento armazenado na MLP afeta a maneira como os indivíduos percebem o mundo e tomam suas decisões<sup>129</sup>.

Segundo Baddeley, Kapelman e Wilson (2004), as informações são armazenadas na memória de trabalho por meio de uma grande rede de esquemas que são interligados entre si<sup>130</sup>.

---

<sup>128</sup> As informações já conhecidas pelo indivíduo são recuperadas por meio do acesso da memória de curto prazo à memória de longo prazo a fim de que não seja necessário armazenar a mesma informação mais de uma vez, por isso há a possibilidade de descarte de informação nova que é irrelevante para o conhecimento do indivíduo pela fase de percepção do estímulo (BADDELEY, 2017).

<sup>129</sup> Como podemos ver na figura 2, a memória de longo prazo é responsável pelas operações mentais de armazenamento, esquecimento e recuperação de informações.

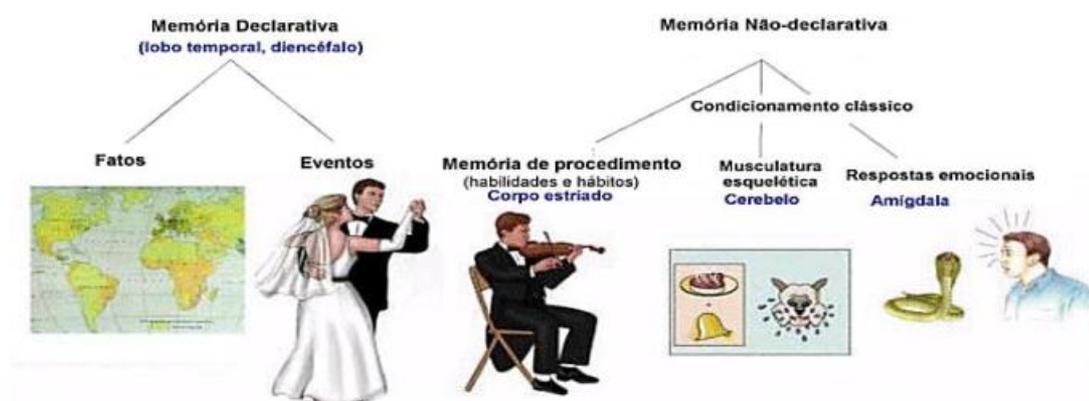
<sup>130</sup> Para Baddeley (2000) não há necessidade de distinguir tipos de memória dentro da memória de trabalho, pois cada módulo é intimamente dependente do outro e não independentes entre si, já que há inúmeras tarefas que levam a ser perceptível o acesso e uso dos dois módulos ao mesmo tempo, como quando se pergunta a uma pessoa o que ela tomou no café da manhã, fazendo com que ela evoque o que comeu – memória episódica – sendo influenciada pelo que geralmente faz parte do café da manhã das pessoas – memória semântica.

Estes esquemas, por sua vez, são os módulos fundamentais para a manutenção do conhecimento e são responsáveis por construir a cognição do indivíduo, além de reduzir a sobrecarga da memória de trabalho<sup>131</sup>.

É conhecido na literatura da área que há dois tipos de MLP: a memória declarativa e a memória não declarativa. Segundo Wilson (2009), a memória declarativa armazena e processa fatos e eventos que podem ser evocados por meio de palavras, enquanto a memória não declarativa é a responsável pelas ações motoras envolvendo procedimentos e habilidades que o indivíduo adquiriu ao longo de sua vida.

A memória declarativa é dividida em dois componentes: a memória episódica<sup>132</sup> e a memória semântica, sendo que as informações contidas nesta memória derivam das informações contidas naquela, fazendo com que os indivíduos sejam capazes de apreender novos fatos e conceitos ao longo de suas vidas.

**Figura 05:** Subdivisões da Memória de Longo Prazo



(Fonte: Bear; Connors; Paradiso, 1996)

Segundo Wilson (2009), a memória episódica leva o indivíduo a ser capaz de se lembrar de incidentes e episódios específicos do passado. Além disso, a autora frisa que muitas pessoas com comprometimento da neste tipo de memória terão problemas com tarefas que envolvam a memória explícita, pois ela é vulnerável aos mais diversos quadros neurológicos.

<sup>131</sup> Segundo Baddeley, Kapelman e Wilson (2004), os esquemas interacionais que existem na memória de longo prazo constroem uma rede semântica que serve para espalhar a informação para os módulos interligados quando uma nova informação é ativada e assim ser capaz de processá-la, armazená-la ou esquecê-la, em prol da não sobrecarga da memória de trabalho.

<sup>132</sup> A memória episódica também é conhecida como *memória autobiográfica*.

É ainda por conta da memória episódica que o indivíduo é capaz de recordar eventos no campo espaço-temporal da mente, sendo estas lembranças sensíveis aos contextos de uso. Além disso, é a memória episódica que ativa todas as ações e processa o conhecimento tácito<sup>133</sup>.

Já a memória semântica, segundo Wilson (2009), é a memória que armazena e processa os conhecimentos gerais sobre o mundo, tais como fatos, significados de palavras, aparência de objetos ou cor das coisas e é usada quando o indivíduo é questionado, por exemplo, sobre qual é a cor de uma banana ou qual é a capital da França.

No que diz respeito à memória não-declarativa, Wilson (2009) aponta para quatro subtipos de memórias neste módulo: (i) memórias adquiridas e ativadas por *priming*, (ii) memória procedural, (iii) memória associativa e (iv) memória não associativa.

Com relação a cada um destes tipos de memória, podemos dizer que (i) está relacionada ao fato de contermos informações sobre um objeto antes mesmo de entrar em contato direto com ele, o que acontece, por exemplo, quando alguém dá dicas sobre um objeto desconhecido para que a outra possa adivinhar, sendo necessário recorrer às impressões que se tem por algo<sup>134</sup>, (ii) está relacionada às habilidades e hábitos do indivíduo, como ser capaz de dar um nós nos cadarços do tênis, (iii) e (iv) estão intimamente relacionadas aos outputs do indivíduo, sendo a memória associativa relacionada aos comportamentos automáticos pelo simples fato de já ter experimentado uma vivência, enquanto a memória não associativa está relacionada ao fato de o indivíduo ter determinado comportamento sem nem mesmo se dar conta, tendo-o apreendido, provavelmente, por meio de tarefas de repetição.

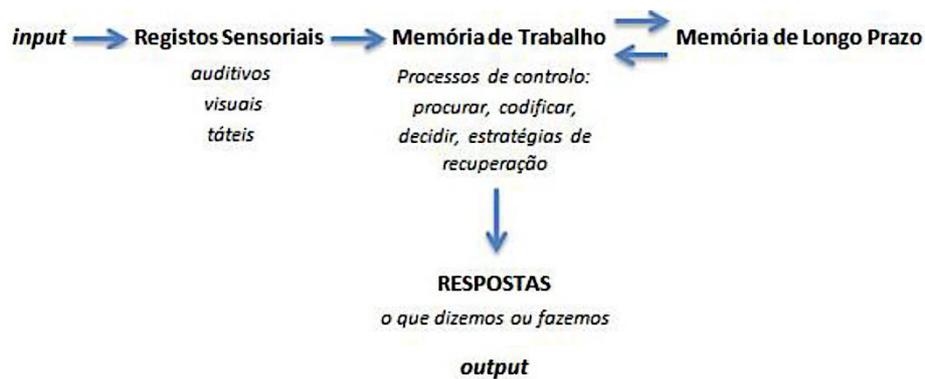
De acordo com Wilson (2009), sabe-se que, tal qual a memória declarativa, problemas neurológicos diversos podem afetar diretamente o funcionamento da memória não-declarativa e, por isso, diversos estudos que têm como foco a reabilitação desta memória têm sido desenvolvidos, mas sem grandes avanços.

---

<sup>133</sup> O conhecimento tácito é adquirido ao longo da vida, pela experiência e completamente dependente da memória (DE JONG, 2006).

<sup>134</sup> Daí a ideia de uma memória pré-ativada (*priming*).

**Figura 06:** O modelo clássico de memória



(Fonte: Atkinson; Shiffrin, 1968)

Assim, levando em consideração as informações acima sobre como uma informação é armazenada e processada pela memória desde a sua entrada enquanto *input* e sua saída como *output* concretizado por meio de linguagem ou movimentos, acreditamos ser capazes de permear o conceito de memória de trabalho com as tarefas linguísticas a serem desenvolvidas nesta pesquisa a fim de explicar como são distribuídos os nomes repetidos, os pronomes plenos e os pronomes nulos em função anafórica em Português Brasileiro.

Contudo, deixamos de tocar em questões mais clínicas sobre os mecanismos cerebrais envolvidos durante a percepção, processamento e armazenamento da informação na memória e as possíveis causas de perda de memória, por não ser objeto de nosso trabalho, mas podendo ser encontrados conceitos e resultados de pesquisas sobre o assunto em Baddeley, Kapelman e Wilson (2005), Wilson (2009), Izquierdo (2011), Lapointe (2012), Baddeley, Eysenck e Anderson (2014), Keremer (2014), Ward (2015), Groome e Eysenck (2016) e Axmacher e Rash (2017).

Vimos, ao longo deste capítulo, diversos trabalhos que buscam dar explicações acerca do processamento anafórico. Além disso, vimos também como a memória é dividida em componentes com funções específicas, o que tem levado diversos pesquisadores a hipotetizar de que maneira descartamos, guardamos e/ou processamos. Buscando ter mais propriedade sobre as explicações já encontradas tanto para o uso quanto para o processamento de expressões anafóricas, revisitaremos, no capítulo seguinte, os principais trabalhos que tratam deste tema em português brasileiro.

### 3 OS ESTUDOS SOBRE A RETOMADA ANAFÓRICA EM PORTUGUÊS

Em Português Brasileiro, diversos estudos têm sido desenvolvidos, a fim de buscar explicações sobre como as formas anafóricas são utilizadas ou processadas nesta língua. Neste capítulo, revisitaremos alguns dos principais trabalhos que tem a retomada anafórica como objeto, começando por trabalhos ligados à observação de uso e variação em dados de *corpora* linguísticos e, posteriormente, os trabalhos ligados ao processamento das anáforas durante tarefas de leitura.

#### 3.1 Breve panorama sobre a retomada anafórica em dados de *corpora* linguísticos

Nesta seção, trataremos sobre alguns trabalhos sobre a retomada anafórica feitos a partir da coleta em dados de *corpora* linguísticos, e, para tal, apresentaremos, inicialmente, os trabalhos ligados à posição sintática de sujeito e, posteriormente, aqueles que tratam da posição de objeto.

##### 3.1.1 A variação de uso dos pronomes na posição de sujeito

A língua portuguesa, como vimos anteriormente, é uma língua *pro-drop*, ou seja, admite o preenchimento da posição de sujeito ou de objeto por uma categoria vazia.

Por conta desta configuração, estudos como os de Kato e Tarallo (1986), Kato (1989, 1998, 2000), Barbosa, Duarte e Kato (2005), Duarte e Kato (2014), Duarte e Silva (2019) têm buscado dar explicações sobre possíveis fatores de diversas ordens linguísticas que possam ter influência sobre a escolha do preenchimento da posição de sujeito por meio de um pronome lexical ou por uma categoria vazia.

Levando em consideração possibilidades de motivação para o uso de uma forma ou de outra, outros tantos estudos (TARALLO, 1987; PAREDES SILVA, 1988; DUARTE, 1996, 2000, 2003, 2018; CAVALCANTE E DUARTE, 2008) têm buscado dar evidências sobre como o preenchimento ou não da posição de sujeito tem ocorrido, tanto em Português Brasileiro quanto em Português Europeu.

Estes estudos têm apontado que, com o passar dos séculos, sobretudo tendo como base documentações escritas e faladas dos três últimos séculos, a configuração do PB tem se tornado distante daquela do PE, havendo, inclusive, pesquisadores que advogam em favor de duas gramáticas distintas para o Português (NEGRÃO, 1999).

Dentre essas tantas variações, uma bastante documentada em pesquisas, sobretudo nos últimos trinta anos, é a possível perda do parâmetro *pro-drop* em PB (DUARTE, 1995) e uma tendência ao preenchimento da posição de sujeito através de pronome lexical.

De acordo com Duarte (1995), a possibilidade de reconhecimento do sujeito sintático em PB se dá, assim como em PE, por meio do uso de um dos pronomes pessoais do caso reto (pronome lexical) ou por meio de morfemas verbais que são específicos da pessoa do discurso.

Com o passar do tempo, de acordo com a autora, sobretudo em *corpora* de fala, é possível perceber que os falantes de todo o território brasileiro, algumas regiões com mais tendência do que outras, tem optado pelo preenchimento do sujeito em vez do apagamento, o que tem levado a uma possível reconfiguração das formas pronominais da posição de sujeito, bem como do paradigma verbal.

Essa possibilidade se deu graças a um massivo processo de variação na frequência de uso dos pronomes lexicais, sobretudo na fala, e, de acordo com Duarte (1995), à perda do princípio Evite Pronome.

Outro fator relevante para a mudança no uso das formas pronominais se deu, dentre outros fatores, por conta da gramaticalização da forma nominal *a gente* (MENON, 1996) que levou, sobretudo na fala dos mais jovens, a configurar a primeira pessoa do plural conjugando o verbo na terceira pessoa do singular em concorrência com a forma canônica de conjugação do pronome *nós*.

Para muitos estudiosos da Sociolinguística (CAVALCANTE; DUARTE, 2008; DUARTE, 1995, 2000, 2003, 2018; HOLMBERG, NAYADU E SHEEHAN, 2009), quando levado em consideração dados de *corpus* de fala e de *corpus* de escrita, o aumento na ocorrência de sujeitos plenos é diretamente relacionado a (i) o aparecimento de novas formas pronominais e (ii) a rápida e crescente simplificação que a morfologia verbal do PB tem sofrido. Outros fatores como a (i) ordem sentencial sujeito-verbo-objeto (SVO) ou objeto-verbo-sujeito (OVS), (ii) traço de animacidade do referente e (iii) tipo de sentença, finita ou infinita, por

exemplo, tem sido observado como fatores diretamente ligados ao uso que tem influenciado na escolha do preenchimento do sujeito por meio de pronomes lexicais em vez de nulos.

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 uma significativa redução ou simplificação nos paradigmas flexionais. (DUARTE, 1995, p.107).

Levando em consideração a redução do paradigma verbal, podemos observar as reconfigurações que o verbo sofreu ao longo do século XIX e a primeira parte do século XX com relação à segunda parte deste mesmo século, conforme mostrado por Kato e Duarte (2014, p. 03).

**Quadro 05:** Os pronomes nominais e o paradigma verbal em PB

Pessoa	Pronomes	Século XIX	Século XX/1	Século XX/2
1ps	eu	<b>estudo</b>	<b>estudo</b>	<b>estudo</b>
1pp	<u>nós</u> a gente	<b>estudamos</b> .....	<b>estudamos</b> <b>estuda</b>	<u>estudamos</u> <b>estuda</b>
2ps	tu você	<b>estudas</b> <b>estuda</b>	<b>estudas</b> <b>estuda</b>	<b>estuda(s)</b> <b>estuda</b>
2pp	vós vocês	<b>estudais</b> <b>estudam</b>	..... <b>estudam</b>	..... <b>estuda(m)</b>
3ps	ele, ela	<b>estuda</b>	<b>estuda</b>	<b>estuda</b>
3pp	eles, elas	<b>estudam</b>	<b>estudam</b>	<b>estuda(m)</b>

É possível observar, no quadro acima, que, no século XIX, no Brasil, os verbos eram conjugados, tanto na fala quanto na escrita, seguindo a conjugação canônica da língua portuguesa, tendo cada verbo um morfema de pessoa específico.

Já na primeira parte do século XX, Duarte e Kato (2014) apontam que a forma pronominal a gente passou a ser frequente na fala dos brasileiros, havendo dados iniciais deste uso na cidade do Rio de Janeiro e, posteriormente, evoluindo para uso em diversas regiões do país (LOPES, 2003), o que levou a um novo uso do verbo da terceira pessoa do singular.

Junto a esta nova forma de uso para fazer referência à primeira pessoa do plural, começou a ser usada menos frequentemente a segunda pessoa do plural *vós*, sendo substituída pelo uso da forma *vocês*, que, por sua vez, no singular, já concorria com a forma pronominal *tu*, configurando em variação de

uso entre a conjugação canônica da segunda pessoa do plural terminada em *s* e o verbo da terceira pessoa do singular (RUMEU, 2013).

Todas essas mudanças, que foram sendo configuradas a partir de processos de variação no Português Brasileiro, fizeram com que Kato e Duarte (2014) advogassem que, no estágio atual de nossa língua, o quadro morfológico das desinências número-pessoais do verbo tenha sido reduzido. Isto, de acordo com as autoras, faz com que, cada vez mais, os falantes necessitem marcar a posição de sujeito com uma das formas pronominais, já que, sem a desinência verbal específica, trazer à mente sobre qual sujeito se fala, tem se tornando uma tarefa dúbia com mais de uma forma verbal para duas ou mais formas que funcionam como pronomes nominais.

É interessante frisar que as mudanças na estratégia de pronominalização de sujeito nulo para sujeito lexical em Português Brasileiro por conta da redução do sistema de flexão verbal pode ser percebida tanto na fala quanto na escrita, o que aponta para uma possível estabilização de mudança gramatical (BARBOSA, DUARTE E KATO, 2005; DUARTE, 2007).

Contudo, é importante destacar que, apesar de as pesquisas supracitadas apontarem para um processo de mudança gramatical, ainda é viável, em Português Brasileiro, ser feito o uso dos sujeitos nulos, como, por exemplo, em construções nas quais os predicados exprimem fenômenos naturais; em casos de possíveis referências indeterminadas ou genéricas; em contextos com sujeitos correferentes; e em respostas afirmativas ou negativas (HOLMBERG, NAYUDU E SHEEHAN, 2009).

O conjunto dos estudos sociolinguístico em PB sobre a retomada anafórica na posição de sujeito tem levado a compreender que existem contextos que parecem favorecer a retomada por um pronome lexical ou por um pronome nulo. Os pronomes lexicais foram encontrados em contextos como (a) orações com referentes acessíveis, (b) contextos não enfáticos ou não contrativos, (c) orações coordenadas não iniciais e (d) uso da terceira pessoa do singular ou do plural (DUARTE, 1996).

Curiosamente, parece que a terceira pessoa do singular e do plural foram os únicos pronomes que não sofreram variação no uso, sendo sua forma preferencialmente marcada, tendo a conjugação verbal para estas pessoas sem mudança em curso ou estabilizada (OLIVEIRA, 2005).

Há também a ressalva de que, segundo Nicolau (1995) e Cabana (2004), a preferencial pelo uso de pronome lexical em vez de pronome nulo nos contextos expostos acima, parece não afetar todos os dialetos no território brasileiro, uma vez que o dialeto mineiro pode ser considerado uma exceção a esta mudança em curso.

É cabal perceber que as pesquisas sobre a retomada anafórica na posição de sujeito tem levado pesquisadores da área de sociolinguística a considerar, além de questões sintáticas, questões semânticas como a animacidade do referente, bem como o sexo, a idade, a escolaridade e a região na qual o texto, falado ou escrito, é produzido e, assim, buscar explicações para como os pronomes são retomados na posição de sujeito.

Porém, não somente na posição de sujeito a variação do pronome nulo para o pronome pleno tem sido percebida, pois, na posição de objeto, pesquisas como as de Corrêa (1991), Nunes (1996) e Cyrino, Duarte e Kato (2000) tem mostrado mudanças no padrão de preenchimento, conforme veremos na seção a seguir.

### **3.1.2 A variação de uso dos pronomes na posição de objeto**

Os estudos sobre como são retomados os referentes na posição de objeto em Português Brasileiro remetem ao trabalho pioneiro de Omena (1978). A autora analisou a fala de indivíduos de baixa classe social em fase alfabetização.

Omena (1978) observou como o pronome pessoal da terceira pessoal acusativa era retomado pelos indivíduos, já que poderia ser realizado (i) pela forma elíptica, (ii) pelos pronomes ele(s)/elas(s) ou (iii) pelos pronomes objetos o(s)/a(s).

Analisando se o pronome tinha ou não dupla função, a posição do objeto, a classe de palavra do antecedente, o traço semântico do antecedente, a distância entre o antecedente e o conseqüente, a função sintática do antecedente, a profundidade entre o antecedente e o conseqüente, além da presença de um ou mais candidatos ao papel de antecedente, Omena (1978) encontrou resultados que apontaram para o fato de que o traço semântico do antecedente, a função sintática do antecedente e o pronome com ou sem dupla função tiveram influenciam no tipo de retomada anafórica na posição de sujeito.

De acordo com Omena (1978), os antecedentes menos animados favorecem o apagamento do objeto, assim com a função sintática paralela também. Além disso, a autora concluiu que os pronomes que têm apenas uma função sintática favorecem o apagamento do objeto. Por fim, Omena (1978) afirmou que, em sua amostra de fala, não encontrou ocorrência do pronome clítico na posição de objeto, o que pode ser um reflexo da baixa escolaridade dos informantes da pesquisa.

Duarte (1986, 1989) buscou indícios de possíveis influências sobre a variação do objeto nulo na fala de paulista a partir da linguagem coletada a partir da fala propagada pela televisão, tendo controlado fatores linguísticos de cunho morfológico, sintático e semântico que poderiam levar à variação.

A autora encontrou resultados que apontam para o fato de que os pronomes clíticos tendem a aparecer antes de um verbo na forma verbal simples do modo indicativo, enquanto os objetos nulos ocorrem em qualquer forma verbal, mas de maneira menos frequente do que os SNs anafóricos em construções no gerúndio.

Duarte (1986, 1989) concluiu que o traço de animacidade do antecedente tem forte influência sobre a forma a ser usada na posição de objeto já que os pronomes plenos e os clíticos tendem a retomar antecedentes mais animados, enquanto os pronomes nulos e os nomes tendem a retomar antecedentes menos animados.

Cyrino (1997) analisou, através de textos dos séculos XVI ao XX, a realização da retomada na posição de objeto direto pelas formas plenas ou nulas em Português Brasileiro, levando em consideração o tipo de antecedente, o modo verbal, o tipo de oração em que o objeto é retomado e o traço de animacidade e de especificidade do antecedente.

De acordo com a autora, o objeto nulo com antecedente mais específico surgiu no PB por volta do século XIX e em contextos nos quais o antecedente era menos animado, o que parece ter favorecido à retomada por objeto nulo, dando ao traço de especificidade do antecedente o papel de possível fator que influencia no tipo de retomada feita na posição de objeto.

Oliveira (2007) analisou textos produzidos por crianças em anos escolares iniciais a fim de verificar se era possível encontrar evidências de uma mudança paramétrica na posição na posição direto, levando em conta as possibilidades

de retomada por objeto direto nulo, pronome tônico, SN anafórico pleno ou clítico acusativo de terceira pessoa.

Os fatores que condicionaram a variação, de acordo com Oliveira (2007) foram a animacidade do antecedente, especificidade e referencialidade do referente, o uso de tempos simples ou compostos, posição de próclise ou ênclise, bem como a escolaridade da criança produtora do texto.

Segundo a autora, o objeto nulo foi muito usado pelas crianças em textos nos quais o referente era menos animado ou em orações com tempo verbal simples; o pronome tônico ele/ela foi a segunda opção de uso das crianças, mas sempre retomando seres animados e/ou concretos, enquanto o uso de nomes anafóricos foi favorecido tanto quando o antecedente era menos animado quanto quando o tempo verbal era simples. Para Oliveira (2007), a baixa ocorrência dos clíticos na posição de objeto direto sugere estes pronomes são adquiridos ao longo dos anos de vida escolar, o que não favorecia o grupo de indivíduos produtores do texto que compuseram o *corpus* de pesquisa.

Costa (2011) analisou a retomada por pronome nulo na posição de objeto em peças teatrais escritas nos séculos XIX e XX em Português Europeu e em Português Brasileiro, levando em consideração a variante do português, o século de nascimento do autor, o tipo de oração, a pessoa do discurso, a animacidade do referente, a realização do objeto, a ordem do objeto, os elementos entre o verbo e o objeto e o estatuto do referente.

De acordo com Costa (2011), tiveram influência sobre a escolha para a forma a ser usada na retomada do objeto fatores de ordens linguísticas e sociais, como: a animacidade do referente, o século de nascimento do autor da peça, o estatuto do referente e a pessoa do discurso.

Costa (2011) afirma que o traço menos animado do referente favorece o aparecimento do objeto nulo, enquanto os referentes com traços mais animados favorecem o surgimento de objeto retomados por pronomes lexicais.

Já no que diz respeito ao século de nascimento do autor, Costa (2011) comprovou que, quanto mais o nascimento foi próximo do século XX, maior foi a tendência do surgimento do objeto nulo. Além disso, Costa (2011) mostrou que o objeto nulo ocorre na terceira pessoa por conta do desaparecimento de seus clíticos em PB, enquanto a omissão dos objetos de primeira e de segunda pessoa não são comuns em nossa língua.

Pereira (2011) analisou as formas de retomada do objeto acusativo em textos produzidos por alunos de quinta a oitava série, observando possíveis influências da pessoa do discurso, da ordem do anafórico, do modo verbal, da animacidade do objeto, além do sexo, idade e escolaridade dos indivíduos produtores do texto.

A autora observou possíveis preenchimentos do objeto pelas formas nulas, oblíquas, SN *o mesmo*, pronome demonstrativo e pronome reto, tendo encontrado resultados que apontam para o fato de os indivíduos do sexo masculino preferirem o uso do objeto nulo, além de que, em todas as ocorrências de objeto nulo, tanto nas ocorrências feitas por homens quanto por mulheres, haver tendência de retomada de referente menos animado.

Ao analisar os resultados, excluindo a forma nula, Pereira (2011) observou que o uso do pronome reto está diretamente ligado à pessoa do discurso e ao sexo e escolaridade do sujeito produtor do texto, além de ter ficado clara a tendência de a terceira pessoa favorecer o aparecimento do pronome pleno na posição de objeto acusativo.

Por fim, ao considerar as retomadas plenas e nulas, Pereira (2011) afirma que somente a animacidade do antecedente tem influência na escolha da forma a preencher a posição de objeto acusativa, mas que o uso destas formas aponta para uma possível distribuição complementar.

Pinto (2015) analisou a retomada na posição de objeto direto acusativo em *corpus* de fala, considerando fatores externos e internos da língua, a fim de explicar se o pronome clítico, o pronome reto, o SN e o objeto nulo são variantes entre si.

A autora percebeu que fatores como o sexo, a idade e a escolaridade do falante têm influência sobre a preferência de uso das formas nulas na posição de objeto, sobretudo quando o referente é sentencial, além de ser usado mais por sujeitos jovens e menos escolarizados do que por sujeito mais velhos e mais escolarizados.

Para Pinto (2015), outros fatores que tiveram influência sobre a escolha da retomada na posição de objeto acusativo foram a função sintática do constituinte retomado e a transitividade verbal e a estrutura projetada pelo verbo, já que o objeto nulo ocorreu com mais frequência quando estava em uma estrutura complexa, bem como quando o seu referente tinha outra função sintática e era menos animado.

Sousa (2017) observou a realização do acusativo anafórico na posição de objeto a partir de *corpus* escrito composto por peças teatrais dos séculos XIX e XX, a fim de verificar os contextos que favoreceram o surgimento do objeto nulo. Neste estudo, o autor levou em consideração fatores como o a faixa temporal de produção dos textos para encontrar evidências de possível mudança, o traço de animacidade do referente e o modo verbal, bem como a estrutura das orações.

O autor afirma que o fator temporal aponta para um processo de longa duração tanto no Português Europeu quanto no Português Brasileiro, o que apontou para a realização do objeto nulo com o passar dos anos nos textos do *corpus* analisado.

De acordo com Sousa (2017), um fator que favoreceu o aparecimento de objetos nulos, em PB, foi a combinação dos traços semânticos do antecedente, pois quanto menos animado é o referente, maior é o número de ocorrências do objeto nulo, tendo sido feita a ressalva de que mesmo quando o referente é mais animado, há retomadas por objetos nulos nos textos do *corpus*.

No que tange ao modo verbal, Sousa (2017) afirma que o modo imperativo favorece o aparecimento de objetos nulos em PE, bem como a forma desenvolvida da oração, o padrão estrutural e a função do antecedente, além de destacar que a frequência de uso dos objetos nulos em PB é muito maior do em PE.

Sousa (2017) afirmou ainda ter encontrado dados que mostraram o uso dos pronomes nominativos na posição de objeto acusativo em PB quando a combinação dos traços mais ou menos humano do antecedente em estrutura canônica de SVO, sem ter encontrado esse tipo de retomada em PE.

Othero e Spinelli (2017) investigaram o uso de retomadas por meio de pronomes plenos e nulos na posição de objeto direto em dados de *corpus* de fala, analisando traços de animacidade, especificidade e gênero semântico do antecedente, a fim de explicar as possíveis motivações na escolha de uso de uma forma ou de outra para a retomada anafórica de terceira pessoa. As autoras também analisaram a retomada anafórica para a primeira e para a segunda pessoa do discurso, apesar de não ser este o objetivo principal de seu trabalho.

De acordo com os resultados encontrados, Othero e Spinelli (2017) afirmam que, de modo geral, os clíticos não estão desaparecendo em PB, pois a baixa frequência de uso destes pronomes plenos a preencher a posição de objeto direto ocorre somente na terceira pessoa, não nas demais. As autoras

acreditam que, pelo fato de a primeira e a segunda pessoa serem pessoas do discurso, os traços de seres mais animados, mais específico e gênero favorecem a marcação pelo pronome e não o seu apagamento.

Já para a retomada quando o referente é a terceira pessoa, Othero e Spinelli (2017) apontam que o único fator que parece ter influência direta na escolha da forma plena ou nula para a posição de objeto direto é o gênero semântico do antecedente. As autoras afirmam que se o antecedente for menos marcado pelo gênero semântico, mais tendência há de retomá-lo por uma categoria vazia, enquanto se o antecedente for mais marcado pelo gênero, mais tendência há para o uso de um pronome pleno.

Vemos, portanto, que diversos estudos que usaram dados de *corpus* escrito e falado têm buscado dar evidências sobre o uso de pronomes plenos e de pronomes nulos tanto na posição de sujeito quanto de objeto, apontando para a influência tanto de fatores externos quanto de fatores internos da língua.

A seguir, revisitaremos os trabalhos que analisam não a produção da língua, ou seja, textos prontos, mas o processo em si, buscando ter explicações de como compreendemos a retomada anafórica.

### **3.2 Panorama sobre a retomada anafórica em pesquisas experimentais**

As pesquisas sobre o processamento correferencial em Português Brasileiro (PB) e em Português Europeu (PE) podem ser, *grosso modo*, divididas em dois grupos: (i) os estudos em que são focados fatores sintáticos do processamento e (ii) os estudos em que são focados fatores semântico-pragmáticos do processamento do par anáfora-antecedente.

Contudo, algumas pesquisas trazem uma mescla de fatores sintático-semânticos e, por conta disso, preferimos revisitá-las em ordem cronológica, sem distinguir por meio de critérios de Sintaxe ou de Semântica, para, então, ter um bom aparato teórico sobre o delineamento do quadro das pesquisas desenvolvidas no Brasil sobre este tema tão caro a esta tese.

Em pesquisas seminais sobre o tema, Leitão (2005a, 2005b) e Leitão e Maia (2005) buscaram explicar, por meio de experimentos em que foi usada a técnica da leitura automonitorada, como eram processados os pronomes, os sintagmas nominais mais gerais (hiperônimos) e os sintagmas nominais mais específicos (hipônimos) levando em consideração o tempo de leitura da anáfora

na posição de objeto de sentenças justapostas. Nestas pesquisas, os autores afirmam que seus estudos têm o escopo localizado na interface sintaxe-semântica.

Leitão (2005a, 2005b) observou em seus estudos a influência da distância entre o referente e a retomada anafórica na posição de objeto, a ligação sintática existente entre o par anáfora-antecedente, a escolha dos itens lexicais na região da anáfora e a concordância de gênero entre o referente e a retomada anafórica.

A fim de observar se os pronomes seriam lidos mais facilmente do que os sintagmas nominais mais gerais e se estes, por sua vez, seriam lidos mais rápidos do que os sintagmas nominais mais específicos, Leitão (2005a)<sup>135</sup> desenvolveu uma série de cinco experimentos, tendo realizado testes de compatibilidade semântica para o desenvolvimento dos itens experimentais a serem registrados por meio da técnica da leitura automonitorada.

No primeiro experimento, Leitão (2005a) buscava encontrar indícios de que tanto os pronomes lexicais<sup>136</sup> quanto os pronomes *pro*<sup>137</sup> eram capazes de reativar, na memória de trabalho, antecedentes correferenciais na posição de objeto sentenças como exemplificadas em (56) (LEITÃO, 2005a, p.93).

(56)

*Antecedente + animado com retomada por pronome*

a. Os vizinhos entregaram **o ladrão** à polícia de manhã, mas depois retiram **ele/Ø** da investigação.

*Antecedente + animado sem retomada anafórica*

b. Os vizinhos entregaram **o ladrão** à polícia de manhã, mas a investigação terminou logo.

*Antecedente - animado com retomada por pronome*

c. Os vizinhos entregaram **o punhal** à polícia de manhã, mas depois retiram **ele/Ø** da investigação.

*Antecedente - animado sem retomada anafórica*

d. Os vizinhos entregaram **o punhal** à polícia de manhã, mas a investigação terminou logo.

Segundo Leitão (2005a), os resultados do primeiro experimento apontam que os pronomes lexicais e os pronomes *pro* são formas psicologicamente

<sup>135</sup> Leitão (2005a) trabalha com um leque de teorias tanto de cunho sintático, como a Teoria da Centralização (GORDON; GROSZ; GUILLIOM, 1993), a Teoria da Proeminência Sintática (CHAMBER; SMITH, 1998), a Teoria do Paralelismo Sintático (GORDON; HENDRICK, 1997; 1998), quanto com teorias de cunho semântico, como a Teoria da Animacidade (CORREA; ALMEIDA; PORTO, 2002; LEITÃO; MAIA, 2002), a teoria da Hipótese da Carga Informacional (ALMOR, 1999), além da Teoria do Construal (FRAZIER; CLIFTON, 1996) que trabalha no escopo da interface sintaxe-semântica.

<sup>136</sup> Denominação utilizada pelo autor para se referir aos pronomes plenos.

<sup>137</sup> Denominação utilizada pelo autor para se referir aos pronomes nulos.

anafóricas capazes de estabelecer a correferência na posição de objeto direto. Além disso, Leitão (2005a) afirma que os traços de animacidade não desempenham efeitos significativos durante o processamento da anáfora na posição de objeto.

Já no segundo experimento, Leitão (2005) investigou se os pronomes lexicais eram lidos mais rapidamente do que os nomes, contendo a mesma quantidade de caracteres na posição de objeto, conforme exemplificado em (57) (LEITÃO, 2005a, p.102).

- (57) a. Os vizinhos entregaram **Ivo** na polícia mas depois absolveram **ele** no júri.  
b. Os vizinhos entregaram **Ivo** na polícia mas depois absolveram **Ivo** no júri.

Os resultados do segundo experimento, de acordo com Leitão (2005a), apontam que os pronomes lexicais são lidos mais rapidamente do que os nomes, o que corrobora a Teoria da Centralização (GORDON; GROSZ; GUILLIOM, 1993), mesmo estando na posição de objeto, o que, segundo o autor, corrobora a teoria da Hipótese da Carga Informacional (ALMOR, 1999).

Contudo, segundo Leitão (2005a), estes resultados vão de encontro ao conceito de centro anafórico, conforme proposto por Gordon, Grosz e Guilliom (1993), já que estes autores preveram a facilitação do processamento por meio dos pronomes quando comparados ao processamento dos nomes na posição mais proeminente da sentença, ou seja, na posição de sujeito e não na posição de objeto.

Entretanto, Leitão (2005a) recorre a uma das predições da teoria da Hipótese da Carga Informacional (ALMOR, 1999) em que se afirma que há um problema na formulação da Teoria da Centralização (GORDON; GROSZ; GUILLIOM, 1993), pois os pronomes tendem a ser processados mais facilmente do que os nomes tanto quando estão na posição de foco discursivo como quando não estão, ou seja, tanto na posição de sujeito quanto de objeto da sentença.

No terceiro experimento, Leitão (2005a) observou se sintagmas nominais menos específicos (hiperônimos) seriam processados mais facilmente do que sintagmas nominais mais específicos (hipônimos) o que, segundo ele, serviria como indício para invalidar a Teoria da Centralização (GORDON; GROSZ; GUILLIOM, 1993) e corroborar com os conceitos da teoria da Hipótese da Carga

Informacional (ALMOR, 1999). Exemplificamos um item experimental do terceiro experimento em (58) (LEITÃO, 2005a, p.109).

- (58) a. Os vizinhos adquiriram **um carro** na loja mas depois venderam **o veículo** no feirão.  
 b. Os vizinhos adquiriram **um carro** na loja mas depois venderam **o chevete** no feirão.

De acordo com Leitão (2005a), os resultados do terceiro experimento corroboram, de fato, com a teoria da Hipótese da Carga Informacional (ALMOR, 1999), já que os sintagmas nominais mais específicos (hipônimos) foram processados mais facilmente do que os sintagmas nominais mais gerais (hiperônimos) na posição de objeto da sentença. Além disso, Leitão (2005a) defende que estes achados “põem em xeque a abrangência explicativa da Teoria da Centralização” (LEITÃO, 2005a, p. 141), mas não explica de que maneira estes resultados são conflitantes com esta teoria de acordo com sua perspectiva.

O quarto experimento foi criado, a fim de observar de que maneira os traços de animacidade podem gerar efeitos durante o processamento correferencial.

Para a criação deste experimento, Leitão (2005a) analisou se o antecedente com traço menos animado tende a favorecer o pronome lexical como forma anafórica na posição de objeto, como exemplificado em (59) (LEITÃO, 2005a, p.116-117).

- (59) *Antecedente + animado com retomada ambígua com pronome lexical/pro*  
 a. Paulo entregou **Marcos** para a polícia, mas depois retiraram **ele/Ø** da investigação.  
*Antecedente – animado com retomada ambígua com pronome lexical/pro*  
 b. Paulo entregou **o revólver** para a polícia, mas depois tiraram **ele/Ø** da investigação.

Os resultados do quarto experimento, de acordo com Leitão (2005a), não mostraram efeito significativo da animacidade do antecedente durante o processamento anafórico o que, de acordo com o autor, já era previsto, uma vez que o pronome lexical estabelece elo mais forte com o traço + animacidade, enquanto *pro* não.

No quinto experimento, Leitão (2005a) investigou se o efeito do Paralelismo Sintático (CHAMBERS; SMITH, 1998) também atua no processamento da anáfora que está na posição de objeto da sentença. Além

disso, Leitão (2005a) observou se o efeito do Paralelismo Sintático interage com fatores de animacidade, a fim de observar se tanto os fatores sintáticos quanto os fatores semânticos influenciam no processamento correferencial. Mostramos um exemplo de item deste experimento em (60) (LEITÃO, 2005a, p.128-129).

(60)

*Antecedente – animado/ + animado na posição de sujeito com retomada pronominal na posição de objeto*

- a. **A câmera** registrou os movimentos do ladrão e depois analisou **ela** no laboratório.
- b. **A Mônica** registrou o caso na delegacia e depois o psicólogo analisou **ela** no consultório.

*Antecedente – animado/ + animado na posição de objeto com retomada pronominal na posição de objeto*

- a. O consumidor registrou **a câmera** no seguro de roubos e furtos e depois o perito analisou **ela** na vistoria.
- b. O pesquisador registrou **a Mônica** no cadastro de bolsista do país e depois o conselho analisou **ela** na reunião.

De acordo com Leitão (2005a), os resultados do quinto experimento mostraram que não houve efeito significativo nem para o efeito do Paralelismo Sintático, nem para os efeitos de animacidade quando a anáfora está na posição de objeto de sentença justaposta.

A fim de encontrar algum indício de quais fatores sintáticos e semânticos podem influenciar o processamento correferencial, Leitão (2005a) observou, a partir dos resultados do quinto experimento que, se for considerada a transitividade verbal como uma variável independente, encontra-se um efeito marginal que aponta para o processamento da anáfora pronominal como menos custoso quando a anáfora é complemento de verbos bitransitivos do que quando ela é complemento de verbos transitivos.

A partir desta interpretação dos resultados, Leitão (2005a) sugere que a grade argumental e a posição sintática influenciam diretamente no processamento correferencial, além de afirmar, contundentemente, que o processamento dos pronomes é mais fácil do que o processamento de nomes na posição de objeto em sentenças justaposta.

Leitão, Peixoto e Santos (2008) investigaram o processamento da correferência no nível intrassentencial<sup>138</sup>, levando em consideração um referente

---

<sup>138</sup> O nível intrassentencial diz respeito ao escopo da sentença, já o nível interssentencial, ao estudo entre períodos ou parágrafos.

não disponível anteriormente, a fim de observar como atua o princípio B<sup>139</sup> da Teoria da Ligação (CHOMSKY, 1981)<sup>140</sup>. Além disso, Leitão, Peixoto e Santos (2008) observaram se os traços de animacidade podem influenciar o leitor a fazer uma tentativa de vinculação entre o pronome na posição de objeto e o referente na posição de sujeito.

Para chegar aos resultados desejados, Leitão, Peixoto e Santos (2008) desenvolveram dois experimentos utilizando a técnica da leitura automonitorada, sendo que o primeiro experimento visava provar que o tempo de leitura do pronome é semelhante em todas as condições, apontando para o bloqueio ou atraso efetivo da ligação entre anáfora e antecedente. Além disso, Leitão, Peixoto e Santos (2008) acreditavam que o tempo de leitura dos advérbios variaria entre as sentenças por haver traços semânticos entre o sujeito e os advérbios. Exemplificamos o item experimental do experimento um, de acordo com Leitão, Peixoto e Santos (2008, p.54), em (61).

- (61)
- a. Tião atropelou **ele** imprudentemente na estrada de Cabedelo.
  - b. Talita atropelou **ele** imprudentemente na estrada de Cabedelo.
  - c. As motoristas atropelaram **ele** imprudentemente na estrada de Cabedelo.
  - d. As carretas atropelaram **ele** imprudentemente na estrada de Cabedelo.

Os resultados do primeiro experimento apontam, segundo Leitão, Peixoto e Santos (2008), para o processamento correferencial sendo executado em dois estágios. No primeiro estágio, há bloqueio de vinculação entre o sujeito e o pronome na posição de objeto, como previsto pela Teoria da Ligação (CHOMSKY, 1981). Já no segundo estágio, o leitor busca fazer as ligações do par anáfora-antecedente levando em consideração os traços de animacidade, gênero e número.

No segundo experimento, exemplificado em (62), Leitão, Peixoto e Santos (2008, p.58) observaram se, a introdução de um preâmbulo no qual há um antecedente disponível estruturalmente e que concorda em gênero, número e animacidade leva ainda ao bloqueio previsto pelo Princípio B.

---

<sup>139</sup> De acordo com o princípio B da Teoria da Ligação (CHOMSKY, 1981), há atraso ou bloqueio da correferência entre o sujeito nominal e o pronome a ele referente na posição de objeto quando no nível intrassentencial.

<sup>140</sup> Leitão, Peixoto e Santos (2008) citam como teorias que fundamentam este trabalho a Teoria da Ligação (CHOMSKY, 1981) e os estudos posteriores acerca desta teoria com as reformulações propostas por Kenisson (2003). Para tratar de animacidade, os autores não citam teorias ou autores que tratam do tema.

(62)

- a. Carlos atravessou a rua correndo. Tião atropelou **ele** imprudentemente na estrada de Cabedelo.
- b. Carlos atravessou a rua correndo. Talita atropelou **ele** imprudentemente na estrada de Cabedelo.
- c. Carlos atravessou a rua correndo. As motoristas atropelaram **ele** imprudentemente na estrada de Cabedelo.
- d. Carlos atravessou a rua correndo. As carretas atropelaram **ele** imprudentemente na estrada de Cabedelo.

De acordo com Leitão, Peixoto e Santos (2008), os resultados do segundo experimento apontam que o tempo de leitura do pronome foi mais lento quando comparado ao tempo de leitura do mesmo pronome no primeiro experimento. Segundo Leitão, Peixoto e Santos (2008) isso poderia ser um indício de que, durante o processamento correferencial, mesmo havendo um possível antecedente da anáfora, há bloqueio efetivo da ligação, conforme previsto pelo Princípio B.

Utilizando a técnica da leitura automonitorada, Leitão e Simões (2008) investigaram a influência da distância semântica no estabelecimento da correferência com base no tempo de leitura do par anáfora-antecedente, tendo como formas anafóricas os pronomes plenos e os nomes repetidos.

Leitão e Simões (2008) acreditavam que o tempo de leitura aumentaria durante o processamento de pronomes plenos e nomes repetidos, o que indicaria que a anáfora seria mais difícil de ser processada quando distanciada semanticamente de seu referente<sup>141</sup>. Mostramos em (63), o exemplo de um item experimental usado por Leitão e Simões (2011, p.266-267).

(63)

Distância curta (10 a 14 sílabas)

- a. Rui plantou uma árvore na frente de casa. **Ele** gosta de jardinagem.
- b. Rui plantou uma árvore na frente de casa. **Rui** gosta de jardinagem.

Distância média (24 a 38 sílabas)

- a. Rui plantou uma árvore na frente de casa e a árvore cresceu rapidamente. **Ele** gosta de jardinagem.
- b. Rui plantou uma árvore na frente de casa e a árvore cresceu rapidamente. **Rui** gosta de jardinagem.

Distância longa (34 a 38 sílabas)

- a. Rui plantou uma árvore na frente de casa e a árvore cresceu rapidamente e criou uma enorme sombra. **Ele** gosta de jardinagem.

---

<sup>141</sup> Os autores trabalham com as predições da Teoria da Centralização (GORDON, GROSZ e GUILLIOM, 1993) e da teoria da Hipótese da Carga Informacional (ALMOR, 1999).

b. Rui plantou uma árvore na frente de casa e a árvore cresceu rapidamente e criou uma enorme sombra. **Rui** gosta de jardinagem.

Os resultados do experimento desenvolvido por Leitão e Simões (2011) levaram os autores a afirmarem que a distância não tem efeito significativo durante o processamento correferencial e que, independentemente do tamanho da distância entre a anáfora e seu antecedente, o pronome sempre é mais fácil de ser processado do que o nome.

Queiroz e Leitão (2008) e Queiroz (2009) comparam o tempo de processamento do pronome pleno, do nome repetido e de sintagmas nominais como hiperônimos e como hipônimos como possíveis formas anafóricas de antecedentes no nível interssentencial<sup>142</sup>.

Queiroz e Leitão (2008) e Queiroz (2009) desenvolveram três experimentos usando a técnica da leitura automonitorada para (i) observar se os pronomes sempre seriam mais fáceis de processar, (ii) observar se os hiperônimos seriam mais fáceis de processar do que os hipônimos e (iii) capturar um possível efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) na posição de objeto<sup>143</sup>.

Em (64), há o exemplo de um item usado por Queiroz (2009, p.34) no primeiro experimento<sup>144</sup>.

- (64) a. Bia leu o artigo de física e depois **ela** atentamente respondeu o questionário.  
b. Bia leu o artigo de física e depois **Bia** atentamente respondeu o questionário.

Os resultados do primeiro experimento, segundo Queiroz e Leitão (2008) e Queiroz (2009), apontam que os pronomes são lidos mais rapidamente na posição de objeto, corroborando as premissas da Penalidade do Nome Repetido<sup>145</sup>, conforme previsto por Leitão (2005a).

<sup>142</sup> Queiroz e Leitão (2008) e Queiroz (2009) trabalham com as predições da Teoria da Centralização (GORDON, GROSZ e GUILLIOM, 1993) e da teoria da Hipótese da Carga Informacional (ALMOR, 1999).

<sup>143</sup> Queiroz e Leitão (2008) e Queiroz (2009) tratam da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e da teoria da Proeminência Sintática (CHAMBERS; SMITH, 1998) como parte da Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995).

<sup>144</sup> Os itens experimentais usados no primeiro e no segundo experimento estão descritos tanto no artigo de Queiroz e Leitão (2008) quanto na dissertação de Mestrado de Queiroz (2009). A fim de simplificar a leitura, preferimos colocar a paginação dos exemplos de acordo com a dissertação de Queiroz (2009).

<sup>145</sup> É necessário ater-se ao fato de que, quando propostas as premissas do efeito denominado de Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), esperava-se encontrá-lo somente na posição mais proeminente da sentença, ou seja, na posição de sujeito. As interpretações de Almor (1999) quando da formulação da teoria da Hipótese da Carga Informacional propõem que as premissas desta penalidade precisam ser revistas, pois quando

Em (65), há o exemplo de um item usado por Queiroz (2009, p.39) no segundo experimento.

- (65) a. O felino fugiu da jaula no zoológico e depois **o bicho** violentamente atacou os visitantes estrangeiros.  
 b. O felino fugiu da jaula no zoológico e depois **o tigre** violentamente atacou os visitantes estrangeiros.

No segundo experimento, de acordo com Queiroz e Leitão (2008) e Queiroz (2009), os resultados apontam para a corroboração dos resultados encontrados por Almor (1999), pois os hiperônimos são processados mais facilmente do que os hipônimos, de acordo com os autores.

Mostramos, em (66), um item experimental usado por Queiroz (2009, p.41) no terceiro experimento.

- (66) a. A briga no quintal estava em pleno andamento.  
 b. Ana insultou Edu com palavras agressivas, depois Ana/ela golpeou Edu no nariz com muita força.
- a. A briga no quintal estava em pleno andamento.  
 b. Ana insultou Edu com palavras agressivas, depois Ana/ela golpeou ele/ele no nariz com muita força.

De acordo com Queiroz (2009), os resultados do terceiro experimento apontam que a teoria da Proeminência Sintática (CHAMBERS; SMITH, 1998) é corroborada, pois, independentemente da expressão anafórica usada – nome repetido ou pronome pleno – a retomada na posição de sujeito é sempre mais facilmente processada do que quando a retomada ocorre na posição de objeto da sentença.

Além disso, segundo a interpretação Queiroz (2009), os resultados do terceiro experimento apontam para evidências de que o efeito de penalização para o uso do nome repetido ocorre somente na posição de sujeito. Contudo, Queiroz (2009) justifica que, antes de aceitar tais resultados como concretos<sup>146</sup>, é necessário levar em consideração os traços semânticos das entidades envolvidas no processamento correferencial.

---

comparado o tempo de leitura de nomes repetidos versus pronomes plenos na posição de objeto, o pronome sempre é mais fácil de ser processado, o que aponta para a concretização da Penalidade do Nome Repetido não somente na posição de sujeito, mas também na posição de objeto. Contudo, a autora não faz menção alguma a esta interpretação ao longo do trabalho.

<sup>146</sup> Estes resultados iriam contra os achados da própria autora no primeiro experimento. Para mais informações, ver Queiroz (2009).

Maia e Cunha-Lima (2011) investigaram o comportamento dos nomes repetidos, pronomes plenos e pronomes nulos como possíveis expressões anafóricas durante o processamento correferencial, a fim de verificar a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) e não da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) na posição de sujeito no nível interssentencial<sup>147</sup>.

A fim de cumprirem com seu objetivo, Maia e Cunha-Lima (2011, p.07), conforme exemplificado em (67), criaram um experimento usando a técnica da leitura automonitorada para comprovarem se os nomes repetidos seriam, de fato, as formas anafóricas mais penalizadas em qualquer contexto quando na posição de sujeito da sentença. Além disso, Maia e Cunha-Lima (2011) esperavam que não houvesse diferença significativa das anáforas na posição de objeto.

- (67) a. Bruna namora Vítor há mais de dois anos e meio.  
 b. **Bruna/Ela/Ø** conheceu Vítor em uma viagem a França.  
 c. Bruna conheceu **Vítor/ele/Ø** em uma viagem a França.  
 d. É possível encontrar um grande amor em outro país.

Maia e Cunha-Lima (2011) encontraram padrões de leitura que mostravam a facilitação de leitura dos pronomes plenos, mostrando a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA) na posição de sujeito, mas nenhum possível efeito de penalização para nomes ou pronomes na posição de objeto, conforme previsto por Almor (1999) e Leitão (2005a).

De acordo com a interpretação de Maia e Cunha-Lima (2011), este resultado inesperado aponta para um possível padrão de processamento em Português Brasileiro semelhante ao que acontece durante processamento correferencial em Espanhol Argentino, sendo os pronomes nulos mais facilmente processados do que os pronomes plenos em contextos de retomada na posição de sujeito. Além disso, em todos os contextos, segundo Maia e Cunha-Lima (2011), o nome repetido é sempre penalizado.

---

<sup>147</sup> Maia e Cunha-Lima (2011) trabalharam com a Teoria da Centralização (GORDON, GROSZ e GUILLION, 1993) e as premissas da Penalidade do Nome Repetido (GROSZ, JOSHI, WEINSTEIN, 1995) e da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) para o desenvolvimento da pesquisa.

Morgado (2011)<sup>148</sup> analisou o processamento de pronomes plenos e de pronomes nulos levando em consideração o papel temático de seu referente durante o processamento correferencial em dois experimentos usando a técnica da leitura automonitorada *on-line* para a leitura dos itens experimentais e *off-line*<sup>149</sup> para a resolução de questionários.

No primeiro experimento, Morgado (2011, p.73), como exemplificado em (68), buscava testar se quando há dois antecedentes com os mesmos traços semânticos, criando um efeito de ambiguidade, o processamento correferencial sofre influência do tipo de verbo enquanto agentivo e não agentivo.

Segundo Morgado (2011), era esperado que o pronome nulo fosse sempre mais facilmente processado na posição de sujeito, independentemente do tipo de verbo usado após a retomada anafórica.

(68) a. Após o acidente, a Ângela<sub>i</sub> ajudou a Rosa<sub>j</sub> na ambulância. Dias depois, Ø/ela<sub>i/j</sub> levantou os exames no hospital.

b. Após o acidente, a Ângela<sub>i</sub> foi ajudada pela Rosa<sub>j</sub> no hospital. Dias depois, Ø/ela<sub>i/j</sub> levantou os exames no hospital.

De acordo com Morgado (2011), os resultados do primeiro experimento apontam que, no nível interssentencial, as pistas sintáticas são mais fortes do que as pistas semânticas, levando em consideração o fato de as retomadas da entidade na posição de sujeito serem facilmente processada, conferindo à entidade na posição de sujeito o estatuto de antecedente preferencial, conforme previsto pela Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995).

Além disso, tomando como base a Teoria da Acessibilidade (ARIEL, 1991), Morgado (2011) afirma que o sujeito nem sempre é a forma mais proeminente da sentença e, por conta disso, é necessário que sua retomada ocorra por meio de um pronome pleno. Isto leva Morgado (2011) a afirmar que o Princípio Evite Pronome (CHOMSKY, 1981) não é produtivo nestes contextos, sendo parcialmente corroborada a Hipótese da Posição do Antecedente (CARMINATI, 2002).

<sup>148</sup> Morgado (2011) faz uso de informações de um vasto leque de teorias, por ela revistas como (i) teorias do processamento ao nível do discurso – Teoria da Acessibilidade (ARIEL, 1991), Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN) e teoria da Hipótese da Carga Informacional (ALMOR, 1999), e (ii) teorias do processamento ao nível sentencial – teoria da Hipótese da Posição do Antecedente (CARMINATI, 2002), além de levar em conta informações previstas pela Teoria da Ligação (CHOMSKY, 1981), pela teoria dos Papéis Temáticos (FILMORE, 1968) e pela teoria das Construções Passivas (CHOMSKY, 1981; JAEGLI, 1986).

<sup>149</sup> Para mais informações, ver Morgado (2011).

Já no segundo experimento, como exemplificado em (69), Morgado (2011, p.89) buscava observar se o efeito de ambiguidade interfere durante o processamento no nível intrassentencial de sentenças subordinadas concessivas, utilizando verbos agentivos.

- (69) a. O Gonçalo socorreu o Raul após o acidente, se bem que Ø/ele não tivesse pronunciado uma única palavra.  
b. O Gonçalo foi socorrido pelo Raul após o acidente, se bem que Ø/ele não tivesse dito uma única palavra.

De acordo com Morgado (2011), os resultados do segundo experimento apontam para os resultados já encontrados no primeiro experimento, pois as pistas sintáticas são mais fortes do que as pistas semânticas, ou seja, a ambiguidade não interfere significativamente na resolução do par anáfora-antecedente nestes contextos.

Já ao analisar os tempos de processamento dos pronomes plenos e dos pronomes nulos, Morgado (2011) foi levada a afirmar que há comportamentos distintos destes pronomes, sendo o pronome nulo a forma preferencial para retomar um antecedente mais proeminente sintaticamente e o pronome pleno a forma preferencial para retomar um antecedente menos proeminente na sentença.

Além disso, Morgado (2011) afirma que a Teoria da Acessibilidade (ARIEL, 1991), a teoria da Hipótese da Posição do Antecedente (CARMINATI, 2002) e o Princípio Evite Pronome (CHOMSKY, 1981) são corroborados pelos resultados deste experimento, assim como os efeitos significativos do tempo de leitura dos pronomes plenos e dos pronomes nulos desaparecem em sentenças passivas.

Leitão, Ribeiro e Maia (2012)<sup>150</sup> verificaram como é o comportamento de pronomes plenos e nomes repetidos durante o processamento correferencial, a fim de observar se a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) ocorre em Português Brasileiro.

Para tal, Leitão, Ribeiro e Maia (2012) desenvolveram dois experimentos utilizando a técnica do rastreamento ocular para (i) observar se o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) ocorre quando processada a retomada anafórica na posição de sujeito de sentenças

---

<sup>150</sup> Leitão, Ribeiro e Maia (2012) trabalharam com a Penalidade do Nome Repetido (GORDON, GROSZ e GUILLIOM, 1993) e a teoria da Hipótese da Carga Informacional (ALMOR, 1999).

coordenadas e (ii) observar o mesmo tipo de efeito quando processada a retomada anafórica na posição de objeto direto de sentenças coordenadas.

Mostramos, em (70), o exemplo de um item usado no primeiro experimento por Leitão, Ribeiro e Maia (2012, p.45).

- (70) a. Eva<sub>i</sub> comprou a tela na galeria e depois ela<sub>j</sub> vantajosamente vendeu no leilão.  
b. Eva<sub>i</sub> comprou a tela na galeria e depois Eva<sub>i</sub> vantajosamente vendeu no leilão.

Leitão, Ribeiro e Maia (2012) afirmam que os resultados do primeiro experimento corroboram os resultados de Queiroz e Leitão (2008) ao apontar para o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) na posição de sujeito, pelo menos em sentenças em que há paralelismo sintático<sup>151</sup>.

Em (71), exemplificamos um item do segundo experimento desenvolvido por Leitão, Ribeiro e Maia (2012, p.48).

- (71) a. Os colegas pintaram Léa<sub>i</sub> no camarim mas depois esqueceram ela<sub>i</sub> no palco.  
b. Os colegas pintaram Léa<sub>i</sub> no camarim mas depois esqueceram Léa<sub>i</sub> no palco

De acordo com Leitão, Ribeiro e Maia (2012), os resultados do segundo experimento corroboram a hipótese de que a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) ocorre não somente na posição de sujeito, mas também na posição de objeto, conforme previsto por Almor (1999).

Maia e Cunha-Lima (2012) resolveram reutilizar os itens experimentais que levaram aos resultados inesperados da ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) e não da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) encontrados por eles anteriormente<sup>152</sup> usando, desta vez, a técnica do rastreamento ocular, a fim de confirmar os resultados com o uso de outro paradigma experimental.

Com os resultados obtidos após a aplicação dos testes em laboratório, Maia e Cunha-Lima (2012) afirmaram que os nomes repetidos e os pronomes plenos têm comportamentos semelhantes, sem que os primeiros sejam penalizados pelos segundos, ou seja, não há tempo de processamento significativo durante a leitura que demonstre o efeito da Penalidade do Nome

<sup>151</sup> Em momento algum do trabalho é citada a concepção de paralelismo sintático adotada por Leitão, Ribeiro e Maia (2012), uma vez que não são citados os trabalhos de Chambers e Smith (1998) nem de Gordon e Kenisson (1998).

<sup>152</sup> Para mais detalhes, ver Maia e Cunha-Lima (2011).

Repetido (GORDON; GROSZ; GUILLIOM, 1993) em PB. Contudo, quando os pronomes nulos são analisados, o tempo de processamento é muito menor, ou seja, os pronomes nulos são mais facilmente processados quando retomam seu referente, penalizando os pronomes plenos, sendo os nulos as formas preferenciais para a retomada no contexto estudado, o que corrobora e reafirma os achados anteriores dos pesquisadores sobre a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008).

Luegi (2012)<sup>153</sup> desenvolveu uma série de quatro experimentos usando as mais diversas metodologias, a fim de analisar o processamento dos pronomes plenos e dos pronomes nulos no nível intrassentencial e, assim, explicar se o processamento correferencial seria resolvido por fatores que levam em consideração a função sintática do antecedente e sua posição estrutural ou se seria resultado de uma combinação destes dois fatores. Além disso, Luegi (2012) buscava evidências que indicassem estes efeitos atuando em diferentes fases do processamento correferencial.

Luegi (2012) esperava encontrar resultados que apontassem para os pronomes nulos como as formas mais facilmente processadas por serem as formas mais reduzidas em língua portuguesa, tornando-as, assim, as formas preferidas para retomar um antecedente mais proeminente na sentença.

Ademais, Luegi (2012) acreditava que em PE, a forma nula retomaria sempre o antecedente mais saliente e a forma plena o antecedente menos saliente. Já em PB, a forma nula deveria ser correferente um antecedente que a c-comande ou, de maneira mais restritiva, a forma nula deveria ser identificada pelo SN mais próximo que a c-comande. Já a forma plena é vislumbrada por conta da perda de produtividade do Princípio Evitar Pronome em Português (CHOMSKY, 1981) do Brasil e, por isso, retomará mais antecedentes sujeito que a mesma forma em Português Europeu.

O primeiro experimento, exemplificado em (72), desenvolvido por Luegi (2012, p.125), por meio da técnica a leitura automonitorada, investigava o comportamento dos pronomes plenos e nulos na posição sujeito e na posição

---

<sup>153</sup> Para o desenvolvimento das pesquisas vinculadas a este trabalho, Luegi (2012) revisitou uma vasta gama de teorias que pudessem ajuda-la a explicar o processamento correferencial, a saber: a Teoria da Ligação (CHOMSKY, 1981), a Teoria da Reflexividade (REINHART, 1983; REULAND, 1993; GRÓDZINSKY e REINHART, 1993), a Teoria da Acessibilidade (ARIEL, 1991), a teoria da Hipótese da Carga Informacional (ALMOR, 1999), a Teoria da Centralização (GROSZ, JOSHI e WEINSTEIN, 1995), a teoria do Paralelismo Estrutural (CARAMAZA *et al.*, 1976), a teoria da Hipótese da Posição do Antecedente (CARMINATI, 2002) e a teoria da Hipótese da Vantagem da Primeira Referência (GERNSBACHER e HARGREAVES, 1988)

de objeto, a fim de analisar os efeitos da função sintática em sentenças na ordem canônica do PE e do PB.

- (72) a. A Sara<sub>i</sub> escreveu para o Emanuel<sub>j</sub> em maio quando Ø foi jubilada<sub>i/o<sub>j</sub></sub> pela faculdade no ano passado.  
 b. Para o Emanuel escreveu a Sara quando Ø foi jubilada<sub>i/o<sub>j</sub></sub> pela faculdade no ano passado.

Os resultados do primeiro experimento, segundo Luegi (2012), não permitiram a confirmação de nenhuma hipótese, uma vez que não houve efeito significativo para as variáveis testadas. Contudo, Luegi (2012) afirma que há indícios de que a função sintática pode ser fator decisivo durante o processamento correferencial, já que foi revelado, por meio deste experimento, que o sujeito, estando antes do verbo, é o antecedente mais saliente, sendo, preferencialmente, retomado por pronome nulo.

Entretanto, segundo Luegi (2012), quando o sujeito está depois do verbo, o pronome nulo é sensível tanto à função sintática do antecedente quanto à sua posição estrutural, apontando para uma combinação destes dois fatores influenciando o processamento correferencial em PE.

Para além disso, Luegi (2012) afirma que os falantes nativos de PE leem mais rápido do que os falantes nativos do PB, sobretudo quando ocorre o pronome nulo, o que pode ser um indicativo da dificuldade que estes falantes podem ter de processar construções com formas nulas, uma vez que estas formas são menos frequentes em sua variedade.

O segundo experimento desenvolvido por Luegi (2012, p.143), exemplificado em (73), foi aplicado usando a técnica do Paradigma do Mundo Visual, a fim de detalhar melhor os resultados encontrados no primeiro experimento.

- (73) a. O bombeiro<sub>i</sub> perguntou pelo militar<sub>j</sub> no quartel quando ele/Ø recebeu a medalha de condecoração.  
 b. Pelo militar<sub>i</sub> perguntou o bombeiro<sub>j</sub> quando ele/Ø recebeu a medalha de condecoração.

Os resultados do segundo experimento indicam, de acordo com Luegi (2012), que tanto a função sintática quanto a posição estrutural dos antecedentes são consideradas durante o processamento correferencial, sendo que quando as duas informações são combinadas positivamente, ou seja, quando a entidade é ao mesmo tempo o sujeito da sentença e se encontra em

uma posição estrutural mais alta, ela é a entidade mais saliente e, preferencialmente retomada por uma expressão anafórica nula. Por outro lado, segundo Luegi (2012), quando estes fatores se combinam negativamente, ou seja, quando a entidade tem função sintática pouco relevante e se encontra em uma posição estrutural mais baixa, como no caso do oblíquo pós-verbal, ela se torna uma entidade pouco saliente e, preferivelmente, retomada por uma expressão anafórica plena.

Contudo, Luegi (2012) afirma que quando estes fatores se opõem e a entidade com função sintática mais saliente não está na posição estrutural mais alta, então as duas entidades são consideradas como antecedente para a expressão anafórica.

Para o terceiro experimento, Luegi (2012) aplicou questionários com os itens experimentais do segundo experimento para observar se os resultados seriam confirmados durante um experimento *off-line*.

De acordo com Luegi (2012), os resultados apontam, categoricamente, que, tanto em PE quanto em PB, há preferência pela retomada do sujeito de uma sentença precedente por pronome nulo, independente da sua posição estrutural – antes ou depois do verbo.

O quarto experimento desenvolvido por Luegi (2012, p.172), exemplificado em (74), foi aplicado por meio da técnica do rastreamento ocular com o objetivo de dar resultados mais robustos e, assim, complementar a discussão proposta em seu trabalho.

- (74) a. A Carla<sub>i</sub> falou com o Márcio<sub>j</sub> na secretaria quando Ø<sub>ij</sub>/ele<sub>i</sub>/ela<sub>j</sub> foi retirado/a da investigação do assalto.  
b. Com o Márcio<sub>i</sub> falou a Carla<sub>j</sub> na secretaria quando Ø<sub>ij</sub>/ele<sub>i</sub>/ela<sub>j</sub> foi retirado/a da investigação do assalto.

Os resultados do quarto experimento indicam, segundo Luegi (2012), que há interação da função sintática e da posição estrutural dos antecedentes para a construção da saliência do referente e mostra o padrão já encontrado de que os pronomes nulos retomam sujeitos, enquanto os pronomes plenos retomam objetos.

Luegi (2012), levando em consideração os resultados de todos os experimentos por ela desenvolvidos, afirma que a saliência é resultado da combinação de diferentes fatores que ocorrem em momentos distintos do processamento correferencial.

Além disso, Luegi (2012)<sup>154</sup> afirma que a ordem das orações, se SVO ou OVS, e a informação morfossintática são integradas e usadas na atribuição de saliência às entidades discursivas durante o processamento correferencial, levando a concluir que o peso destas informações ao longo do processamento é variável.

Maia (2013)<sup>155</sup> examinou como nomes repetidos, pronomes plenos e pronomes nulos são processados em Português Brasileiro e em Português Europeu em contextos de antecedentes salientes e não salientes.

Maia (2013) acreditava que, em PB, os sujeitos nulos seriam, independente da saliência sintática do seu referente, a forma mais difícil de ser processada por não apresentar traços de concordância verbal, além de estarem cedendo lugar aos pronomes plenos nesta variedade. Porém, Maia (2013) acreditava que, apesar de os nomes repetidos serem custosos durante o processamento, os pronomes nulos seriam mais custosos ainda quando exercendo a função anafórica na posição não saliente do que em posição saliente.

Já em PE, Maia (2013) acreditava que os pronomes nulos não seriam penalizados quando correferentes de antecedente salientes, configurando a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008), enquanto os pronomes plenos seriam mais facilmente processados do que os nomes repetidos, configurando a ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993).

Para o desenvolvimento desta pesquisa, Maia (2013) utilizou as técnicas de leitura automonitorada, rastreamento ocular e julgamento de aceitabilidade para os experimentos que envolviam a variedade do PB e de leitura conduzida para o experimento que envolvia a variedade do PE.

O primeiro experimento de Maia (2013, p.37), exemplificado em (75), era uma replicação do experimento desenvolvido por Gerlomini-Lezama (2008) em Espanhol Argentino, sendo traduzido e adaptado por ele em parceria com Maia e, assim, obter resultados sobre os efeitos de penalização das formas anafóricas.

---

<sup>154</sup> Em suas considerações finais, Luegi (2012) alerta que o uso de diferentes paradigmas experimentais é necessário para que haja uma análise transversal dos resultados, a fim de que se possa verificar o que acontece em cada estágio do processamento.

<sup>155</sup> Maia (2013) trabalhou com as predições da Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995) e da teoria da Hipótese da Carga Informacional (ALMOR, 1999).

- (75) a. **Maria** se encontrou com João. **Maria/Ela/ Ø** achou ele triste.  
 b. João se encontrou com **Maria**. **Maria/ Ela/ Ø** achou ele triste.

Os resultados do primeiro experimento apontam, segundo Maia (2013) para a ocorrência da Penalidade do Pronome Pleno, conforme previsto por Gerlomini-Lezama (2008), mas não para a Penalidade do Nome Repetido em PB na posição de sujeito, conforme previsto por Gordon, Grosz e Guilliom (1993). Já na posição de objeto, segundo Maia (2013), não há penalização para nenhuma das duas formas pronominais, sendo o nome repetido a forma mais fácil de processar na posição menos saliente.

No segundo experimento, exemplificado em (76), Maia (2013, p.46) manipulou os itens do primeiro experimento para averiguar o comportamento do pronome clítico *o* na posição de objeto ao invés do pronome tônico *ele*.

- (76) a. **Maria** se encontrou com João. **Maria/Ela/ Ø** o achou(-o) triste.  
 b. João se encontrou com **Maria**. **Maria/ Ela/Ø** o achou(-o) triste.

Os resultados do segundo experimento, de acordo com Maia (2013), confirmam os resultados do primeiro experimento com a ressalva de que o uso do clítico no lugar do pronome tônico não gerou novo padrão de processamento para os nomes repetidos em PB, pois as diferenças significativas nos tempos de leitura dos itens experimentais do primeiro experimento quando comparados aos tempos de leitura dos itens experimentais do segundo experimento foram idênticas entre si.

No terceiro experimento, exemplificado em (77), Maia (2013, p.54) testou o comportamento dos pronomes plenos e dos pronomes nulos quando não há manipulação da saliência do antecedente.

- (77) a. Paula deixou Miguel no aeroporto internacional. **Paula/Ela/Ø** o abraçou(-o) pela última vez na vida.  
 b. Miguel deixou Paula no aeroporto internacional. **Paula/Ela/Ø** o abraçou(-o) pela última vez na vida.

Os resultados do terceiro experimento, segundo Maia (2013), apontam que, quando não há manipulação da saliência do antecedente, os pronomes plenos e os nomes repetidos apresentam comportamentos semelhantes entre si, enquanto os pronomes nulos são as formas mais facilmente processadas,

apontando novamente para a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) em PB.

Por mais uma vez encontrar o efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) e a não ocorrência da Penalidade do Nome Repetido (GORSON; GROSZ; GUILLIOM, 1993), Maia (2013) aponta que fatores como o controle do tamanho dos itens experimentais em quantidade de caracteres, o contexto pesquisado ser intra ou interssentencial e demais condições de natureza semântica podem ser capazes de interferir no processamento de nomes repetidos, pronomes plenos e pronomes nulos, propiciando a ocorrência de uma penalidade em detrimento da outra.

No quarto experimento, Maia (2013) aplicou um teste de julgamento de aceitabilidade das sentenças usadas no terceiro experimento, a fim de observar se os resultados seriam os mesmos e, assim, confirmar suas conclusões do experimento anterior.

Os resultados do quarto experimento, de acordo com Maia (2013) mostram que, em contextos nos quais o antecedente está na posição de sujeito, os pronomes nulos são penalizados, o que vai contra os resultados do experimento *on-line*. Já nos casos em que os antecedentes estão na posição de objeto, tanto os resultados dos experimentos *on-line* quanto do experimento *off-line* convergem para demonstrar a baixa preferência de uso e o maior custo de processamento dos sujeitos nulos em PB.

O quinto experimento desenvolvido por Maia (2013), exemplificado em (78), buscava evidenciar a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) em PB tanto quando o antecedente está em posição mais saliente como quando ele está em posição menos saliente.

(78) a. Mara conheceu Artur em um café na bela capital francesa. Mara/Ela/Ø usou vermelho no primeiro encontro.

b. Artur conheceu Mara em um café na bela capital francesa. Mara/Ela/Ø usou vermelho no primeiro encontro.

De acordo com Maia (2013), os resultados do quinto experimento apontam para a não ocorrência da Penalidade do Nome Repetido em PB nem quando o antecedente está em posição mais saliente, ou seja, sujeito da sentença, conforme previsto por Gordon, Grosz e Guilliom (1993), nem quando está em posição menos saliente, ou seja, objeto da sentença, conforme previsto por Almor (1999).

O sexto experimento desenvolvido por Maia (2013, p.78), exemplificado em (79), foi criado como uma replicação dos itens do terceiro experimento com manipulações necessárias para estar adequado à variante do Português Europeu. Contudo, os participantes deste experimento e de todos os outros experimentos desenvolvidos por Maia (2013) foram falantes nativos do Português Brasileiro.

(79)

a. A Paula deixou o Miguel no aeroporto internacional. A **Paula/Ela/Ø** abraçou-o pela última vez na vida.

b. O Miguel deixou a Paula no aeroporto internacional. A **Paula/Ela/Ø** abraçou-o pela última vez na vida.

De acordo com Maia (2013), os resultados do sexto experimento apontam para a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) em PE, sendo os pronomes nulos, assim como em PB, as formas anafóricas preferidas para estabelecer a correferência com antecedentes sintaticamente salientes. Contudo, Maia (2013) atenta para o fato de que, quando observada a preferência das retomadas para entidades menos salientes, não há penalização entre pronomes nulos e nomes repetidos em PE, enquanto há este tipo de penalização em PB, o que pode apontar para um possível efeito de comportamento distinto dos pronomes nestas variantes do Português.

Teixeira (2013)<sup>156</sup> e Teixeira e Soares (2014) investigaram, utilizando a técnica do rastreamento ocular, como se comportam os SNs mais gerais (hiperônimos) e os SNs mais específicos (hipônimos) quando em relação de correferência com o termo antecedente em períodos complexos por coordenação e em períodos complexos por subordinação, a fim de demonstrar que a posição sintática é o principal fator para a resolução anafórica em PB.

Teixeira (2013) e Teixeira e Soares (2014) acreditavam que em períodos complexos, o sintagma nominal que ocupa a posição de sujeito da oração principal é o antecedente preferencialmente retomado para estabelecer a correferência com a primeira expressão referencial anafórica do período.

No primeiro experimento, como exemplificado em (80), Teixeira (2013, p.86) e Teixeira e Soares (2014) analisaram os tempos de leitura de períodos

---

<sup>156</sup> Teixeira (2013) trabalhou com as predições da Teoria da Acessibilidade (ARIEL, 1991), da Teoria da Centralização (GROSZ, JOSHI e WEINSTEIN, 1995), a teoria da Tipicidade do Termo Antecedente (GARROD e SANFORD, 1977; VAN GOMPEL, LIVERSEDGE e PEARSON, 2004) e da teoria da Hipótese da Posição do Antecedente (CARMINATI, 2002).

complexos por coordenação. As autoras acreditavam que as expressões anafóricas tendem a estabelecer correferência com os antecedentes na posição de sujeito da primeira oração coordenada, além de supor que não haveria custos adicionais no tempo de processamento quando o objeto direto da primeira oração passasse a ser o centro catafórico com potencial a ser retomado por uma expressão anafórica na segunda oração. Os custos adicionais surgiriam pelo fato de a hiponímia ser mais difíceis de processar do que a hiperonímia, além de acreditarem que o paralelismo estrutural das orações seria um fator relativamente fraco para o estabelecimento da correferência.

(80)

Orações coordenadas para analisar a retomada anafórica  
por hiponímia ou hiperonímia na posição de sujeito

**O macaco** subiu na árvore mais próxima e depois **o chipamzé** avidamente comeu os frutos maduros.

Orações coordenadas para analisar a retomada anafórica  
por hiponímia ou hiperonímia na posição de objeto

Os biólogos avistaram **um réptil** no rio mas depois assustaram **o animal** na margem.

O experimento foi analisado por meio de dois grupos: (i) participantes que leram as orações coordenadas com retomada anafórica por hiponímia na posição de sujeito, levaram Teixeira (2013) a afirmar que encontrou os mesmos resultados obtidos por Leitão (2005). E, (ii) participantes que leram as orações coordenadas com retomada anafórica por hiponímia na posição de objeto, tendo os resultados, segundo Teixeira (2013), apontado para a hiperonímia como sendo a forma mais fácil de processar na posição de objeto.

O segundo experimento, exemplificado em (81), desenvolvido por Teixeira (2013) e Teixeira, Fonseca e Soares (2014) buscava analisar como se dá o processamento de pronomes plenos e de pronomes nulos em contextos ambíguos, tendo sido analisadas as orações subordinadas temporais em ordem direta e em ordem inversa, para que, houvesse antecedente em alguns casos e, em outros, o referente estivesse depois da retomada.

Teixeira (2013) desenvolveu este experimento utilizando a técnica de rastreamento ocular, entretanto, houve um teste de sondagem para a produção de orações subordinadas, a fim de investigar qual seria o pronome escolhido pelos participantes para estabelecer a correferência com o sujeito e com o objeto de uma oração principal.

(81)

*Escolha da expressão anafórica entre pronome pleno e pronome nulo em contextos ambíguos.*

- a. **A Maria** conversava com **a Joana** enquanto **Ø/ela** cozinhava
- b. Enquanto **Ø/ela** cozinhava, **a Maria** conversava com **a Joana**.

Os resultados do experimento feito por meio da técnica do rastreamento ocular, de acordo com Teixeira (2013), apontam que os pronomes nulos estão em distribuição complementar em PB e que estes pronomes estabelecem correferência com o antecedente mais saliente.

Teixeira (2013) afirmar que o teste de sondagem indica que o pronome pleno pode estar sendo distribuído de maneira diferente nas quatro modalidades da língua, o que pode estar levando os usuários do PB a acessarem os pronomes plenos e os pronomes nulos de maneira gradual e distinta entre si, usando os nulos ligados a antecedentes mais acessíveis, enquanto os plenos são ligados a antecedentes menos acessíveis<sup>157</sup>.

Simões (2014)<sup>158</sup> investigou, utilizando a técnica da leitura automonitorada, se a resolução dos pronomes plenos e dos nomes repetidos durante o processamento correferencial seria influenciada por fatores de coesão e coerência por acreditar que sentenças nas quais há conjunção incoerente seriam mais difíceis de processar do que sentenças com conjunções coerentes.

Simões (2014) menciona que o objetivo de seu primeiro experimento, exemplificado em (82), era explicar a influência da coesão e da coerência no processamento correferencial com base no tempo de leitura dos pronomes em posição de sujeito em PB. Já o segundo experimento tem o mesmo objetivo, entretanto, é analisado o tempo de leitura dos nomes repetidos.

(82)

*Retomada na posição de sujeito com pronome pleno, analisando a presença ou ausência do conectivo em segmento textual congruente*

<sup>157</sup> De acordo com Teixeira (2013), os resultados de seus experimentos não levam a comprovar a distribuição complementar entre os pronomes plenos e os pronomes nulos em Português Brasileiro, mas a compreender que o pronome pleno apresenta comportamento flexível a depender do contexto e da posição estrutural da frase na qual o seu antecedente estiver.

<sup>158</sup> Simões (2014) cita em seu trabalho alguns estudos de Koch para tratar de coesão e coerência, assim como cita a teoria da Hipótese da Carga Informacional (ALMOR, 1999). No segundo experimento, a autora fala sobre a Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995), mas não faz menção a esta teoria ao longo da fundamentação teórica de seu trabalho.

Eva concluiu brilhantemente o texto, **portanto/Ø**, o trabalho ficou primoroso. Ela é uma excelente redatora.

*Retomada na posição de sujeito com pronome pleno,  
analisando a presença ou ausência do conectivo em segmento textual incongruente*

Eva concluiu brilhantemente o texto, **portanto/Ø**, o trabalho ficou medíocre. Ela é uma excelente redatora.

Os resultados do primeiro experimento, segundo Simões (2014), mostram que as sentenças com segmento textual congruente são mais facilmente lidas. Simões (2014) afirma que isso ocorre porque os fatores de coesão e coerência discursiva, quando combinados, geram efeitos sobre o papel dos conectivos, reforçando a expectativa relacionada à continuação da sentença.

Ainda de acordo com Simões (2014), foram analisados os tempos de leitura dos conectivos, não sendo encontrados tempos significativos de processamento para as conjunções, o que indica que este elemento não deve ser visto como facilitador do processamento intrassentencial.

Além disso, Simões (2014) afirma que as retomadas com pronome não sofrem nenhum tipo de efeito no tempo de leitura por haver um segmento incongruente na primeira sentença.

Em (83), há um exemplo de item experimental utilizado por Simões (2014, p.71) para a realização do segundo experimento.

(83)

*Retomada na posição de sujeito com nome repetido,  
analisando a presença ou ausência do conectivo em segmento textual congruente*

Eva concluiu brilhantemente o texto, **portanto/Ø**, o trabalho ficou primoroso. Eva é uma excelente redatora.

*Retomada na posição de sujeito com nome repetido,  
analisando a presença ou ausência do conectivo em segmento textual incongruente*

Eva concluiu brilhantemente o texto, **portanto/Ø**, o trabalho ficou medíocre. Eva é uma excelente redatora.

Os resultados do segundo experimento levaram Simões (2014) a afirmar que os resultados do primeiro experimento são confirmados no que diz respeito à sentença coerente ser processada mais facilmente do que a sentença incoerente. Entretanto, Simões (2014) afirma que a presença do conectivo

facilitou o processamento do último segmento da primeira sentença, o que pode gerar o efeito de dificuldade de processamento das sentenças incoerentes<sup>159</sup>.

Ao fazer uma comparação no tempo de leitura dos nomes repetidos e dos pronomes plenos, Simões (2014) afirma que o pronome pleno é mais facilmente processado do que o nome repetido nas sentenças incoerentes, podendo ser um indício do efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) nestes contextos.

Hora (2014)<sup>160</sup> analisou o comportamento dos pronomes plenos e dos pronomes nulos durante o processamento correferencial de estruturas complexas do PB usando a técnica da leitura automonitorada e de questionário.

Hora (2014) acreditava que tanto os pronomes plenos quanto os pronomes nulos tendem a retomar antecedentes que têm a mesma função sintática que eles e, acreditava ainda, que as duas formas pronominais podem ocorrer no mesmo contexto. Além disso, Hora (2014) acreditava que, na posição de sujeito, os tempos de leitura dos pronomes plenos e dos pronomes nulos seriam praticamente o mesmo<sup>161</sup>.

O primeiro experimento desenvolvido por Hora (2014), exemplificado em (84), visava demonstrar se o paralelismo sintático seria o fator mais relevante durante o processamento correferencial.

- (84) a. **Mário** beijou **Paula** quando **ele/ela** encontrou **ela/ele** no teatro.  
 b. **Mário** beijou **Paula** quando  $\emptyset$  encontrou **ele/ela** ali no teatro.  
 c. **Mário** beijou **Paula** quando **ele/ela** encontrou  $\emptyset$  ali no teatro.

Os resultados do primeiro experimento apontam que, segundo Hora (2014), as retomadas de sujeito são mais facilmente processadas do que as retomadas de objeto, independentemente do tipo de expressão anafórica usada como forma correferente, sobretudo quando a retomada de objeto acontece por meio de objeto nulo.

<sup>159</sup> Em suas considerações finais, Simões (2014) afirma que os resultados dos dois experimentos desenvolvidos por ela apontam que os recursos de coesão estão a serviço da coerência.

<sup>160</sup> Hora (2014) trabalhou com a Teoria da Ligação (CHOMSKY, 1981), com a Teoria da Acessibilidade (ARIEL, 1999), com a Teoria da Centralização (GONDON, GROSZ e GUILLION, 1993), com a teoria da Hipótese da Carga Informacional (ALMOR, 1999) e com a teoria da Hipótese da Posição do Antecedente (CARMINATI, 2002), além disso, a autora cita a teoria do Paralelismo Estrutural, mas não faz menção a nenhum pesquisador que tenha desenvolvido esta teoria.

<sup>161</sup> Para o desenvolvimento dos quatro experimentos, Hora (2014) manipulou os traços de gênero e de número. Para mais informações, ver Hora (2014).

Além disso, Hora (2014) afirma que a Hipótese do Termo Antecedente (CARMINATI, 2002) não é confirmada em PB e que os pronomes nulos e os pronomes plenos parecem exercer a mesma função nesta língua, ou seja, são formas que não estão em distribuição complementar.

A fim de observar se os resultados do primeiro experimento seriam consistentes, Hora (2014, p.74) desenvolveu o segundo experimento, como exemplificado em (85), por meio de questionário com sentenças nas quais os traços de número foram manipulados para observar como os participantes responderiam ao teste *off-line*.

- (85) a. **A médica** conversa com **as enfermeiras** no hospital embora **Ø/ela(s)** pare(m) pouco tempo durante o plantão.  
 b. Embora **a médica** converse com as enfermeiras no hospital, **Ø/ela(s)** para(m) pouco tempo durante o plantão.

Os resultados do segundo experimento confirmam os resultados do primeiro experimento, levando Hora (2014) a afirmar que os pronomes plenos e os pronomes nulos têm comportamento semelhante em PB, sobretudo quando são formas correferentes de um antecedente na posição de sujeito.

Hora (2014) afirmou também que os resultados dos dois experimentos juntos apontam para o fato de que, em PB, os pronomes plenos tendem a ser mais facilmente processados na posição de objeto, o que confirma que os pronomes nulos tendem a retomar antecedentes que aparecem na posição mais alta da estrutura sintática.

O terceiro experimento desenvolvido por Hora (2014), exemplificado em (86), teve manipulação do traço de número, a fim de analisar os resultados do segundo experimento usando a técnica da leitura automonitorada.

- (86) a. **O zelador** defendeu **os professores** no curso embora **Ø/ele(s)** fique(m) sempre muito mal com todos.  
 b. Embora **o zelador** defenda **os professores** no curso **Ø/ele(s)** fica(m) muito mal com todos.

De acordo com Hora (2014), os resultados do terceiro experimento apontam que tanto os pronomes nulos quanto os pronomes plenos têm a mesma função em PB, já que ambas retomam um antecedente na posição de sujeito sem preferência de uma em detrimento de outra.

O quarto experimento, exemplificado em (87), segundo Hora (2014, p.75), é uma replicação do terceiro experimento, sendo manipulada mais uma vez o

traço de número, levando à retomada da terceira pessoa do singular e não mais da terceira pessoa do plural, como no experimento anterior.

- (87) a. **O zelador** defendeu **a professora** no curso embora **Ø/ele(a)** fique muito mal com todos.  
 b. Embora **o zelador** defenda a professora no curso **Ø/ele(a)** fica muito mal com todos.

De acordo com Hora (2014), os resultados do quarto experimento são idênticos aos resultados do terceiro experimento, levando-a à afirmação de que os traços de número e de gênero são altamente relevantes para o estabelecimento da correferência.

A afirmação de Hora (2014) sobre os pronomes plenos e pronomes nulos terem a mesma função em PB dá a entender que ela não encontrou nenhum efeito de penalidade, apesar de não fazer menção a tal fato de maneira direta. Contudo, há de ser analisado de que maneira Hora (2014) interpretou os resultados dos seus experimentos, pois é comum, na literatura da área, encontrar dados que apontam para o fato de que em determinados contextos ocorre o efeito da Penalidade do Nome Repetido e (GORDON; GROSZ; JOSHI, 1993), e em outros contextos, o efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008).

A aceitação de resultados que apontem para o fato de não haver nenhuma penalização entre as formas, afirmando, ainda, que não há preferência de uso do pronome pleno ou do pronome nulo, como se a escolha fosse uma questão apenas estilística, pode ser bastante controversa e inesperada se tomamos as demais pesquisas como base.

Além disso, Hora (2014) descreve em seu trabalho experimentos que, em um primeiro momento, podem parecer repetitivos entre si, pois há manipulação do traço de gênero no primeiro experimento e no quarto experimento e há manipulação de traço de número no segundo e no terceiro experimento.

Contudo, o primeiro experimento tem a entidade introduzida no discurso por meio de nome próprio e, nos demais experimentos, todas as entidades são introduzidas por nomes comuns que tem como referente no mundo uma entidade animada, especificamente pessoas, que poderiam ser representadas por nomes próprios, sendo necessária, portanto, uma discussão acerca da diferença entre nomes comuns e próprios, o que pode ter gerado algum efeito durante o

processamento, já que os nomes próprios parecem ser mais bem definidos no universo linguístico do que nomes comuns.

E, com relação à retomada no centro anafórico, Hora (2014) não usou em momento algum a retomada por nome repetido, o que pode ter levado aos resultados inesperados de não ocorrência de nenhum tipo de penalidade para o pronome pleno ou para o pronome nulo, pois as formas pronominais foram comparadas entre si e não com a forma menos esperada neste contexto, ou seja, o nome repetido, conforme proposto tanto pelos autores da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) quanto pelo autor da Penalidade do Pronome Pleno (GERLOMINI-LEZMA, 2008).

Levando em consideração o conjunto de resultados dos quatro experimentos, Hora (2014) sugere que fatores como a ordem das orações, o tipo de expressão anafórica e a posição do antecedente desempenham papel significativo durante o processamento correferencial em PB.

Lima (2014)<sup>162</sup> analisou, por meio de dois experimentos utilizando a técnica da leitura automonitorada, se o foco e o paralelismo estrutural influenciam no processamento interssentencial de nomes repetidos e pronomes plenos, analisando comparativamente os tempos de leitura destas formas correferenciais.

Lima (2014) acreditava que o processamento seria mais rápido quando o antecedente estivesse na posição de foco estrutural, independentemente do tipo de retomada. Entretanto, comparando o tempo de leitura das duas formas anafóricas, Lima (2014) esperava que os pronomes plenos fossem mais fáceis de processar do que os nomes repetidos, apontando para a ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) no contexto estudado.

Além disso, Lima (2014) acreditava que o paralelismo estrutural seria um fator relevante no estabelecimento da correferência sendo capaz de se sobrepor a outros efeitos que possam atuar durante o processamento correferencial, como o efeito da proeminência sintática ou efeito de foco discursivo.

No primeiro experimento desenvolvido por Lima (2014, p.46), exemplificado em (88), foi analisado o tempo de leitura dos pronomes plenos e

---

<sup>162</sup> Lima (2014) trabalhou com a Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995), com a teoria da Hipótese da Carga Informacional (ALMOR, 1999) e com a teoria da Proeminência Sintática (CHAMBERS e SMITH, 1998)

dos nomes repetidos funcionando como expressões correferentes, a fim de observar se o foco estrutural facilitaria o processamento correferencial e se o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) seria encontrado nesse tipo de estrutura, podendo mostrar indícios de que esta penalidade varia de acordo ou não com a focalização do termo antecedente.

(88)

*Antecedente focalizado*

Foi **Ivo** quem substituiu Sheila na gincana. **Ele/Ivo** dependia do prêmio para viajar.

*Antecedente não focalizado*

Foi Sheila quem substituiu **Ivo** na gincana. Todos queriam **Ivo/ele** na gincana.

Segundo Lima (2014), os resultados do primeiro experimento indicam que o foco estrutural não teve efeito significativo durante o processamento das expressões anafóricas no contexto estudado neste trabalho. Entretanto, Lima (2014) sugere que este resultado pode ter sido gerado por uma possível atuação do paralelismo estrutural durante processamento dos itens experimentais. E, de acordo com Lima (2014), foi encontrado o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993).

No segundo experimento, exemplificado em (89), Lima (2014, p.52) analisou se há efeito de facilitação do processamento do foco estrutural em sentenças com o correferente em paralelo com relação ao seu antecedente.

(89)

*Antecedente focalizado*

Foi **Ivo** quem substituiu Sheila na gincana. Todos queriam **Ivo/ele** na competição.

*Antecedente não focalizado.*

Foi Sheila quem substituiu **Ivo** na gincana da escola no ano passado. **Ivo/Ele** dependia do prêmio para viajar.

Segundo Lima (2014), os resultados do segundo experimento mostram que o foco estrutural tem efeito significativo durante o processamento correferencial.

Lima (2014) explica que este resultado já era esperado pelo fato de as sentenças estarem clivadas, fazendo com que o foco estivesse altamente ativo na memória de trabalho e, por isso, fosse recuperado mais facilmente.

Além disso, Lima (2014) afirma que o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) foi encontrado tanto em sentenças com o antecedente focalizado quanto em sentenças com o antecedente não focalizado.

Lima (2015)<sup>163</sup> analisou, utilizando a técnica do rastreamento ocular, como acontece o processamento de nomes repetidos, pronomes plenos e pronomes nulos na posição de sujeito simples e de sujeito composto com distanciamento sintático, a fim de observar se a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) são efeitos excludentes entre si.

Lima (2015) acreditava que o pronome nulo seria mais difícil de processar quando estivesse sintaticamente distante do seu referente e, neste caso, o uso do pronome pleno ou do nome repetido seria opcional, o que apontaria para a não ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) em PB.

No primeiro experimento desenvolvido por Lima (2015, p.55), exemplificado em (90), foi observada a retomada anafórica de uma entidade ocupando a posição de sujeito simples, sendo retomado na mesma posição sintática, com e sem distância sintática<sup>164</sup>, a fim de observar o comportamento dos nomes repetidos, dos pronomes plenos e dos pronomes nulos nesses contextos.

(90)

*Sujeito simples com retomada executada por nome repetido,  
pronome pleno e pronome nulo sem sentença interveniente*

Jaime<sub>i</sub> conhece Diego<sub>j</sub> de vários lugares. Eles<sub>k</sub> estão sempre nas mesmas festas, principalmente as de música eletrônica. Jaime/Ele/Ø<sub>i</sub> é um dos produtores desse tipo de festa mais famosos de sua cidade e fez contato com seu conhecido para atrair mais público para suas festas

*Sujeito simples com retomada executada por nome repetido,  
pronome pleno e pronome nulo com sentença interveniente*

Jaime<sub>i</sub> conhece Diego<sub>j</sub> de vários lugares. Eles<sub>k</sub> estão sempre nas mesmas festas, principalmente as de música eletrônica. As festas de música eletrônica, também conhecidas como raves, têm sido ponto de encontro de muitos jovens, especialmente os jovens de classe média e alta. Porém, ainda existe muito preconceito da sociedade, de maneira geral, com relação

<sup>163</sup> Lima (2015) trabalhou com a Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995), com a teoria da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e com a teoria da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008).

<sup>164</sup> Lima (2015) chama a sentença criada para gerar a distância sintática de sentença interveniente.

a essas festas, pois se acredita que seus frequentadores estão sempre sob o efeito de drogas ou alucinógenos. Jaime/El<sub>e</sub>/Ø<sub>i</sub> é um dos produtores desse tipo de festa mais famosos de sua cidade e fez contato com seu conhecido para atrair mais público para suas festas

Segundo Lima (2015), os resultados do primeiro experimento apontam que, durante a leitura da região crítica, os pronomes nulos foram as formas anafóricas mais facilmente processadas, mostrando a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) nesse contexto. Contudo, é preciso destacar que, diferentemente dos demais trabalhos apresentados até aqui, Lima (2015) introduziu dois referentes a ser, somente um, retomada na região crítica, acontecendo, em seu experimento, ambiguidade referencial, o que pode ter interferido nos resultados encontrados, bem como torna relativa a comparação aos demais resultados anteriores.

Por não encontrar penalização para quando comparado o tempo de leitura do pronome pleno, Lima (2015) afirma que o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e o efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) são excludentes entre si, pelo menos no contexto estudado, e chama a atenção para o fato de que o aparecimento de efeito de uma penalidade ou outra em estudos anteriores pode ser resultado do *design* experimental criado e manipulado quando levadas em consideração as variáveis para o estudo desenvolvido por pesquisador.

No segundo experimento desenvolvido por Lima (2015, p.65), exemplificado em (91), foi observada a retomada anafórica de uma das duas entidades que ocupam a posição de sujeito composto, sendo retomada apenas uma delas na mesma posição sintática, com e sem distância sintática, a fim de observar o comportamento dos nomes repetidos, dos pronomes plenos e dos pronomes nulos nesses contextos.

(91)

*Sujeito composto com retomada executada por nome repetido,  
pronome pleno e pronome nulo sem sentença interveniente*

Erick<sub>i</sub> e Jaldo<sub>j</sub> trabalham assiduamente como voluntário. Eles<sub>k</sub> participam de muitos eventos, principalmente eventos de caridade. Erick/El<sub>e</sub>/Ø<sub>i</sub> trabalha com total disposição e afinco e seu amigo, apesar de gostar do trabalho, precisa de trabalho remunerado por ser de família carente.

*Sujeito composto com retomada executada por nome repetido,  
pronome pleno e pronome nulo com sentença interveniente*

Erick<sub>i</sub> e Jaldo<sub>j</sub> trabalham assiduamente como voluntário. Eles<sub>k</sub> participam de muitos eventos, principalmente eventos de caridade. Os eventos de caridade movimentam e mobilizam

as pessoas a participarem ativamente por causas que beneficiem os mais necessitados, fazendo com que as pessoas que se predispõem a desenvolver esse tipo de atividade não tenham bloqueios ou vergonha, já que, em muitos casos, é necessário pedir ajudas e doações batendo de porta em porta. Erick/Eløi trabalha com total disposição e afinco e seu amigo, apesar de gostar do trabalho, precisa de trabalho remunerado por ser de família carente.

Para Lima (2015), os resultados do segundo experimento mostram que, em contextos em que há dois antecedentes do mesmo gênero, com sentença interveniente, o nome repetido é processado mais facilmente do que as formas pronominais, apontando para a não ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008).

Lima (2015) afirma mais uma vez que as duas penalidades são excludentes entre si, em PB, porque quando o tempo de leitura das três possíveis formas anafóricas é analisado, não há efeito significativo para o tipo de retomada neste contexto. Lima (2015) reafirma que, a depender do *design* experimental, uma ou outra penalidade pode ser encontrada, mas não as duas ao mesmo tempo.

E, neste experimento, foi mostrado que a distância sintática não desempenha papel significativo durante o processamento correferencial.

A fim de analisar melhor os dados e os resultados encontrados nos dois experimentos, Lima (2015) fez um cruzamento dos dados, a fim de dar maiores explicações sobre seus achados.

Os resultados do cruzamento de dados levaram Lima (2015) a afirmar que a distância sintática, apesar de não desempenhar papel significativo durante o processamento correferencial, exerce algum tipo de efeito já que pode ter alguma relação com o esvaziamento ou com a manutenção da informação na memória de trabalho durante o processamento correferencial, mas não sendo a distância em si o fator gerador de penalidades entre as formas anafóricas.

Lima (2015) afirma que ficou comprovado o papel significativo que o tipo de retomada exerce durante o processamento correferencial, o que pode ser um forte indicativo de que a manipulação dos estímulos experimentais durante a realização de pesquisas acerca do processamento correferencial, em PB, pode aumentar as chances de ser encontrada a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) ou a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008), mas não ambas no mesmo contexto.

Além disso, Lima (2015) afirma que os resultados do cruzamento dos dados apontam para um efeito gerado pelo antecedente, ou seja, a retomada é sensível à presença de mais de um candidato.

Almor e colaboradores (2017)<sup>165</sup> examinaram, por meio de estudo usando a técnica da leitura automonitorada, teste de julgamento e dados de *corpus*, a extensão dos efeitos da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) em Português Brasileiro, a fim de comprovar que a frequência de uso do pronome nulo em contextos de uso real é similar à frequência deste nos estímulos experimentais.

Almor e colaboradores (2017) acreditavam que o PB é uma língua que está perdendo o sujeito nulo e passando a torná-lo marcado e que, por isso, é encontrado o efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008), o que daria evidências de que a preferência pelo uso de pronomes plenos em vez de pronomes nulos pode indicar outros princípios de processamento baseados apenas na expectativa de uso.

O primeiro experimento, exemplificado em (92), desenvolvido por meio da técnica da leitura automonitorada, levou Almor e colaboradores (2017, p.102) a observar o comportamento de nomes repetidos, de pronomes plenos e de pronomes nulos como formas anafóricas na posição de sujeito de uma entidade do mesmo gênero também na mesma posição sintática. Já o segundo experimento, desenvolvido por meio de testes de julgamento, levou Almor e colaboradores (2017) a observarem, usando o mesmo *design* do primeiro experimento, qual seria a forma anafórica preferida pelos participantes da pesquisa<sup>166</sup>.

- (92) a. Carla<sub>i</sub> desafiou Luís<sub>j</sub> para uma partida de xadrez.  
b. Carla/Ela/Ø<sub>i</sub> venceu(-o)<sub>j</sub> rapidamente e sem esforço.

De acordo com Almor e colaboradores (2017), os resultados do primeiro experimento apontam que, quando o antecedente está na posição de sujeito, as sentenças com pronomes plenos são lidas mais lentamente do que as sentenças

---

<sup>165</sup> Almor e colaboradores (2017) trabalharam apoiados nas premissas da Teoria da Centralização (GORDON, GROSZ e GUILLIOM, 1993), da teoria da Hipótese da Carga Informacional (ALMOR, 1999) e da teoria da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008)

<sup>166</sup> Para mais detalhes, ver Almor *et al.* (2017).

com pronomes nulos, mas sem efeitos significativos entre si, quando comparadas ao tempo de leitura das sentenças com nomes repetidos, o que indica a não ocorrência nem da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e nem da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) em Português Brasileiro.

Já os resultados do teste de julgamento dos itens utilizados no primeiro experimento revelam, segundo Almor e colaboradores (2017), que os falantes nativos do PB preferem pronomes plenos a pronomes nulos na posição de sujeito do centro anafórico. Contudo, para Almor e colaboradores (2017), este resultado é atenuado quando o centro anafórico está na posição de objeto, levando os autores a afirmarem que o pronome nulo é menos aceitável como correferente de entidades tanto na posição de sujeito quanto na posição de objeto.

É interessante frisar que Almor e colaboradores (2017) não encontraram, durante o experimento desenvolvido por meio da técnica de leitura automonitorada, tempos de leitura significativos que mostrassem a ocorrência da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) e isto veio a ser confirmado com os resultados dos testes de julgamento. Porém, mesmo com estes resultados encontrados, Almor e colaboradores (2017) são enfáticos ao afirmarem que continuam aceitando que a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) é um efeito existente em PB.

Já os resultados da análise dos dados de *corpora* revelaram uma grande variedade de uso entre o pronome pleno e o pronome nulo a depender do contexto de uso<sup>167</sup>.

Por fim, Almor e colaboradores (2017) alegam que a mudança de sujeito menos marcado para sujeito mais marcado em PB pode não ser tão drástica, uma vez que os pronomes nulos ainda são bastante frequentes, mesmo sendo mais raros do que os pronomes plenos quando traçadas observações comparativas.

A partir disso, Almor e colaboradores (2017) afirmam que em gêneros mais recentes, não somente os pronomes nulos têm se tornado raros, mas os

---

<sup>167</sup> Na verdade, Almor e colaboradores (2017) levam os leitores deste trabalho a entender que o uso do pronome pleno ou do pronome nulo varia a depender da esfera comunicativa à qual está vinculada, pois eles afirmam que os gêneros falados ou os gêneros escritos que contém características mais próximas da fala, como os jornais, mostram uma grande preferência pelo uso dos pronomes plenos ao invés dos pronomes nulos.

pronomes plenos também, já que os falantes têm usados outras formas marcadas para além do pronome como forma correferente ao nome.

E, tomando como base todas as análises feitas durante o desenvolvimento desta pesquisa, Almor e colaboradores (2017) são contundentes ao afirmarem que a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) existe em Português Brasileiro e que os pronomes nulos não estão sendo substituídos pelos pronomes plenos nesta língua.

Gondin (2017)<sup>168</sup> investigou, usando a técnica da leitura automonitorada, o comportamento de pronomes plenos e de nomes repetidos, focalizando o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) para explicar os resultados divergentes das pesquisas anteriores acerca dos efeitos de penalização das formas pronominais.

Gondin (2017) acreditava que a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) seria um efeito multifatorial e que deveriam ser levados em conta fatores linguísticos e metodológicos que poderiam estar na base dos resultados conflitantes acerca dos efeitos de penalidade em PB.

No primeiro experimento, exemplificado em (93), Gondim (2017, p.44) analisou o processamento dos nomes repetidos e dos pronomes plenos em condições com um ou com dois antecedentes no nível interssentencial em um conjunto de duas sentenças justapostas, a fim de obter dados sobre como a quantidade de antecedentes pode influenciar na ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) em Português Brasileiro.

- (93) a. Ana substituiu Nei na gincana da escola. Ela/Ana hoje joga na seleção.  
b. Ana participou bem da gincana da escola. Ela/Ana hoje joga na seleção.

De acordo com Gondin (2017), os resultados do primeiro experimento apontam que as sentenças com apenas um antecedente foram lidas mais rapidamente do que as sentenças que continham dois antecedentes, o que pode indicar que a quantidade de antecedentes pode ser um dos fatores que atuam durante o processamento correferencial. Entretanto, é importante levar em conta

---

<sup>168</sup> Gondin (2017) trabalhou adotando as premissas básicas da Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995), da teoria da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1995), da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA) e da teoria da Hipótese da Carga Informacional (ALMOR, 1999).

que, mesmo com dois antecedentes, não há ambiguidade atuando durante o processamento neste estudo.

Além disso, Gondim (2017) afirma que os pronomes plenos foram mais facilmente processados do que os nomes repetidos, mostrando a ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), independentemente da quantidade de antecedentes introduzidos na primeira sentença justaposta.

No segundo experimento, Gondim (2017) reutilizou, em nova rodada do procedimento experimental, os itens do primeiro experimento passando a aferir o tempo da segunda sentença em vez de diversas segmentações<sup>169</sup>, como no experimento anterior.

Os resultados do segundo experimento, segundo Gondin (2017) confirmaram os resultados encontrados no primeiro experimento e revelaram que a forma como as sentenças são segmentadas para a aferição/análise do tempo de leitura parece ser bastante influente na captura da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993). Entretanto, Gondim (2017) afirma que esta penalidade não pode ser encontrada em condições com sentenças justapostas, contendo dois antecedentes quando é aferido o tempo total de leitura da sentença crítica.

No terceiro experimento, Gondin (2017) replicou os itens experimentais do segundo experimento, a fim de analisar se a tarefa final em interação com outros fatores poderia interferir durante o processamento correferencial. Ao final da leitura dos itens foi usada uma pergunta de compreensão<sup>170</sup>, que não focava no antecedente, além disso, as passagens experimentais foram divididas em 09 (nove) segmentos<sup>171</sup>.

Segundo Gondin (2017), assim como no primeiro experimento, os resultados do terceiro experimento mostram que as sentenças com nomes repetidos foram lidas mais lentamente do que as sentenças com pronomes plenos, apontando para a ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993).

No que diz respeito ao processamento da tarefa final, Gondin (2017) não encontrou efeito significativo deste fator para a ocorrência da Penalidade do

---

<sup>169</sup> Para mais informações, ver Gondim (2017).

<sup>170</sup> No primeiro experimento, ao final da leitura, os participantes tinham de acessar a uma palavra sonda.

<sup>171</sup> A mesma segmentação usada no primeiro experimento.

Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), entretanto, a forma de segmentação/aferição do tempo de leitura, sim, o que aponta para uma apreensão desta penalidade de forma mais *on-line* do que em momentos tardios do processamento.

No quarto experimento, exemplificado em (94), Gondim (2017, p.62) analisou a interferência que pode ser causada pelo tipo de sentença em interação com o fator segmentação/aferição do tempo de leitura das sentenças, a fim de explicar se em orações coordenadas ocorreria a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), mas não em orações justapostas.

- (94) a. Bia encontrou Téo na frente da escola. Ela/Bia logo falou sobre o concerto.  
b. Bia encontrou Téo na frente da escola e ela/Bia logo falou sobre o concerto.

Segundo Gondin (2017), os resultados do quarto experimento mostram que a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) ocorre nos dois contextos estudados, não ocorrendo em contextos nos quais é aferido o tempo total da leitura do segmento crítico de sentenças justapostas com dois antecedentes, confirmando os resultados encontrados no segundo experimento.

O quinto experimento desenvolvido por Gondim (2017, p.66), exemplificado em (95), levou à análise do processamento correferencial em contextos nos quais é introduzida uma entidade na posição de sujeito a ser retomada na segunda sentença que ora é uma sentença coordenada, ora é uma sentença justaposta, sendo aferido o tempo total da leitura da sentença crítica.

- (95) a. Bia chegou cedo na sala de aula. Ela/Bia logo começou a tarefa.  
b. Bia chegou cedo na sala de aula e ela/Bia logo começou a tarefa.

De acordo com Gondin (2017), os resultados do quinto experimento mostram que não houve efeito significativo nem para o tipo de retomada, nem para o tipo de sentença, além de não ter havido interação entre estes fatores, o que levou ao processamento quase idêntico de nomes repetidos e pronomes plenos neste contexto, não contribuindo, portanto, para a ocorrência da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993).

Por fim, levando em consideração os resultados obtidos nos cinco experimentos, Gondim (2017) afirma que a forma como os itens experimentais

são segmentados para a análise do tempo de leitura é um dos fatores que, junto a outros fatores de ordem linguística, pode influenciar na captura da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993).

Outros trabalhos que levam em conta o processamento anafórico em Português Brasileiro não foram descritos nesta seção por não serem de interesse imediato para este trabalho, mas que merecem ser mencionados por tratarem sobre (i) o processamento anafórico em L2 (GADELHA, 2012), (ii) o processamento anafórico em pessoas com algum tipo de doença, distúrbio ou afasia (ALBUQUERQUE, 2008; VASCONCELOS; LEITÃO, 2012; ALVES, 2012, CORREIA; FERRARI-NETO; LEITÃO, 2013), (iii) o processamento das anáforas conceituais (GODOY, 2010; 2014) e (iv) questões de cunho teórico que podem atuar durante o processamento de expressões anafóricas (LEITÃO *et al.*, 2014; FERRARI-NETO; CORREIA, 2014; LEITÃO; BEZERRA; BRITO, 2014; LIMA; SOUZA, 2016).

Assim, podemos perceber que os estudos em Psicolinguística Experimental que tratam sobre o processamento correferencial em Português, sobretudo em Português Brasileiro, têm levado à discussão a questão de haver ou não o processamento mais fácil de uma forma anafórica, fazendo com que entendamos que não se deve usar aquela que parece ser sempre mais difícil de processar.

O fato é que, tomados em conjunto, os estudos, em geral, ainda não são capazes de dar uma explicação contundente e que possa ser aceita como um norte nos estudos desta área no Brasil, pois ainda há estudos que buscam mostrar que há Penalidade do Nome Repetido (LEITÃO, 2005a, 2005b; LEITÃO; MAIA, 2005; LEITÃO; PEIXOTO; SANTOS, 2008; QUEIROZ; LEITÃO, 2008; QUEIROZ, 2009; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; SIMÕES, 2014; LIMA, 2014; GONDIN, 2017), mas sem analisar a interação do pronome nulo como uma forma pronominal anafórica que pode ainda ser usada em PB como variável a ser verificada; e os estudos que buscam mostrar que há Penalidade do Pronome Pleno (MAIA; CUNHA-LIMA, 2011, 2012, 2014; MAIA, 2013; LIMA, 2015; ALMOR *et al.*, 2017) como fator de exclusão do outro tipo de penalidade.

Nos trabalhos citados ao longo desta seção, os autores falam sobre os participantes serem falantes nativos de Português Brasileiro ou de Português Europeu que são levados a tarefas de sentenças experimentais para serem

lidas, mas não especificam se os trabalhos partem de construções mais próximas da fala ou da escrita nesta língua, nem dos achados em pesquisas linguísticas ou da Linguística Cognitiva.

Além disso, levando em consideração os trabalhos descritos neste capítulo e no anterior, é possível perceber que a vasta literatura psicolinguística sobre o processamento anafórico possui uma lacuna, uma vez que são analisadas como formas anafóricas somente os nomes repetidos, os pronomes plenos e/ou os pronomes nulos, ficando de fora a possibilidade do aparecimento de nomes recategorizados, já que, de acordo com Koch (2014), Basseto (2015) e Magalhães e Brito (2016), uma das funções do nomes é recategorizar o referente, surgindo no texto por meio de diversas expressões nominais que não sejam somente a repetição do mesmo item lexical<sup>172</sup>.

Por conta disso, propomos fazer uma análise de dados de *corpora linguísticos* para averiguar a frequência de uso dos nomes repetidos, dos pronomes plenos e dos pronomes nulos e os possíveis fatores linguísticos que teriam influência no uso destas formas para, a partir de então, buscar possíveis explicações de como estas formas correferentes estão distribuídas em Português Brasileiro.

Veremos, portanto, no próximo capítulo, a metodologia e a descrição da pesquisa de coleta em dados de *corpora* com vistas a explicar a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito em Português Brasileiro.

---

<sup>172</sup> Há trabalhos como os de Almor (1999) e de Teixeira (2013) que explicam o processamento anafórico de hiperônimos de hipônimos, mas não como possibilidades de nomes recategorizados que podem ser opção para o usuário da língua para a não repetição do mesmo item lexical, como pretendemos, neste trabalho, analisar as novas formas usadas para retomar um referente.

## 4 FATORES LINGUÍSTICOS LIGADOS AO PROCESSAMENTO CORREFERENCIAL

Seguindo a linha sobre a resolução da correferência anafórica de base Psicolinguística, apresentamos, neste capítulo, a descrição linguística de dados de *corpora* escrito e, no capítulo seguinte, de fala, sendo necessário, de antemão, especificar os objetivos e as hipóteses que nortearam a pesquisa documental<sup>173</sup> da qual trata esta tese.

### 4.1 Objetivos e hipóteses

O objetivo em comum das análises que desenvolvemos e que serão reportadas nas próximas seções foi explicar como as expressões correferenciais, mais especificamente os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos são distribuídos em Português Brasileiro, levando em consideração fatores de ordem sintática, tomando como base os resultados encontrados em diversas pesquisas psicolinguísticas, conforme visto no capítulo anterior.

Mais especificamente, as análises dos *corpora* linguísticos objetivaram:

a) averiguar em dados de escrita e em dados de fala se a quantidade de palavras entre o termo antecedente e a expressão correferente geram usos preferenciais de uma forma em detrimento de outra, dando-nos, assim, evidências de quando o falante faz uso de nomes ou de pronomes na posição de sujeito, uma vez que pesquisas apontam que a distância em quantidade de sílabas (LEITÃO; SIMÕES, 2011) interfere na resolução anafórica, mas a a distância sintática criada por uma sentença interveniente entre o referente e a retomada (LIMA, 2015) não;

b) utilizar conceitos encontrados em pesquisas psicolinguísticas como Paralelismo Sintático (CHAMBERS; SMITH, 1998), Proeminência Sintática (GORDON; HENDRICK, 1997), Penalidade do Nome Repetido (GORDON;

---

<sup>173</sup> Segundo Gil (2008), as técnicas de coletas de dados podem ser seis tipos: (i) pesquisa bibliográfica, (ii) pesquisa documental, (iii) pesquisa experimental, (iv) estudo de campo, (v) estudo de caso e (vi) pesquisa-ação. Dentre esta classificação, o autor afirma que a pesquisa documental é bastante similar à pesquisa bibliográfica, mas difere desta ao tratar de “materiais que ainda não receberam dados estatísticos, ou que ainda podem ser reelaborados de acordo com o objeto de pesquisa” (GIL, 2002, p.46), o que nos leva a afirmar que, nesta tese, é feita uma pesquisa documental já que nos propomos a reanalisar um *corpus* já existente.

GROSZ; GUILLIOM, 1993) e Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) como norteadores para a criação das categorias de análise para, a partir da busca de evidências pautadas na frequência de uso em dados de *corpora*, termos possíveis indícios da materialização de tais conceitos no produto linguístico, o que levaria a compreender que a ação destes efeitos pode interferir não somente no processamento da retomada anafórica, mas também na produção textual;

c) averiguar se há algum contexto que favorece o aparecimento dos nomes, mostrando um possível uso preferencial desta forma, uma vez que todas as pesquisas psicolinguísticas feitas sobre o processamento anafórico em PB (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; QUEIROZ, 2009; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; MAIA; CUNHA-LIMA, 2011; LEITÃO; RIBEIRO; MAIA, 2012; MAIA; CUNHA-LIMA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; MAIA, 2014; LIMA, 2015; GONDIM, 2017) indicam que há penalização dos nomes, isto é, o nome é a forma mais custosa cognitivamente.

É importante frisar que, este estudo tem como foco a descrição linguística por meio da frequência de ocorrência das formas correferentes por nós estudadas em dados de *corpora*, ou seja, analisamos o produto linguístico, enquanto as pesquisas psicolinguísticas tratam do processamento do material linguístico e, por conta disto, não podemos relacionar diretamente a frequência de uso com efeitos de penalidade.

Por fim, este estudo tem como meta interpretar os resultados dos dados encontrados em *corpus* escrito e em *corpus* de fala à luz de modelos teóricos de inspiração pragmático-discursiva (ARIEL, 1990; GORDON; GROSZ; GUILLIOM, 1993; GORDON; HENDRICK, 1997; CHAMBERS; SMITH, 1998; GERLOMINI-LEZAMA, 2008), apontando o que efetivamente é encontrado no produto linguístico com relação aos efeitos e às penalidades descobertos em pesquisas psicolinguísticas e que servem como base para nossas categorias de análise. A partir dos resultados encontrados, esperamos elencar parâmetros que possam vir a ser levados em consideração em futuros *designs* experimentais de pesquisas nesta área.

No tocante às hipóteses que nortearam as nossas investigações, prevíamos que, em PB, houvesse alguma interferência da distância entre o termo antecedente e a sua expressão correferente, fazendo com que as

expressões correferenciais, mais especificamente, o nome repetido, o pronome pleno e o pronome nulo estivessem distribuídos ao longo do texto de forma determinada pelo uso, não impossibilitando o surgimento de formas menos frequentes quando em posição de sujeito. Com isso, esperávamos que o uso de um pronome em dada distância de seu antecedente não inviabilizasse a aparição de um nome, mas que, por meio da quantidade de ocorrências, ficasse clara a preferência de uso de uma forma ao invés de outra.

Esperávamos também que por meio da análise da quantidade de ocorrências dos centros anafóricos na posição de sujeito retomados por meio de nomes repetidos, de pronomes plenos ou de pronomes nulos, fôssemos capazes de encontrar indícios da ação dos efeitos de Proeminência Sintática (GORDON; HENDRICK, 1997) e Paralelismo Sintático (CHAMBERS; SMITH, 1998) no produto linguístico, corroborando, assim, pesquisas que tratam destes efeitos. Além disso, esperávamos que uma possível ação do efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) não fosse corroborado por acreditarmos que, em línguas *pro-drop*, o pronome pleno não concorre com o pronome nulo para ocupar a posição de sujeito, mas funciona como uma opção de uso a qual o falante pode ou não lançar mão e, por conta disto, encontrarmos evidências que mostrassem a ação do efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) corroborando, portanto, algumas pesquisas e outras não.

Esperávamos ainda encontrar indícios de que, em determinados contextos, o uso do nome repetido fosse mais viável para efetuar a retomada do antecedente do que o pronome pleno ou o pronome nulo, podendo estar diretamente ligado à distância, em quantidade de palavras, de seu termo antecedente. Assim, acreditávamos que houvesse determinado contexto no qual o uso do pronome pleno ou do pronome nulo seria menos frequente, o que seria refletido na baixa frequência de ocorrência destes, enquanto o nome repetido, por ser mais frequente, facilitaria a compreensão do texto, tornando-se, portanto, a forma mais usual, o que seria refletido na alta quantidade de ocorrências do nome.

Sobre o uso dos nomes como expressões correferenciais, é preciso comentar, desde já, que esperávamos que, ao longo do texto, dois tipos de nomes pudessem retomar o antecedente e, assim, configurar como parte da

categoria de nome. Por conta disto, criamos uma subcategoria para o nome, mas isto será melhor apresentado nas seções a seguir.

Chamamos a atenção para o fato de que, neste trabalho, associamos o termo penalizar aos resultados encontrados por meio de pesquisas experimentais que devem buscar evidências a partir da quantidade de ocorrências de determinado termo nos dados do *corpus*, ou seja, esperamos que o termo dito penalizado em pesquisas psicolinguísticas apareça com menos frequência, enquanto o termo mais acessível cognitivamente, de acordo com as pesquisas do capítulo anterior, surja com mais frequência. Destacamos, novamente, que, apesar de estarmos correlacionando custo cognitivo com frequência de uso, não podemos afirmar que um leve ao outro, uma vez que penalidade é um fenômeno exclusivamente cognitivo, o que nos leva a não tratar de penalidade, mas de distribuição de formas em dados de *corpora* linguísticos.

Por fim, esperávamos encontrar resultados que nos levassem a criar padrões de distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos parâmetros que possam vir a ser considerados durante a construção de futuros *designs* experimentais que tratam da correferência anafórica, sobretudo em posição de sujeito e, para tal, conduzimos uma análise descritiva em *corpora* escrito e de fala, a fim de comprovar nossas previsões. As análises feitas serão apresentadas nas próximas seções, nas quais serão expostas as categorias de análise dos *corpora* e comentários específicos tanto dos dados do *corpus* escrito como dos dados do *corpus* de fala, bem como serão expostos e discutidos os resultados encontrados.

#### **4.2 Procedimento para a obtenção de dados empíricos no *corpus* escrito**

Foi utilizado, para o levantamento das ocorrências de nomes repetidos, pronomes plenos e pronomes nulos em posição de sujeito, o material disponível no projeto Lácio-Web<sup>174</sup>.

---

<sup>174</sup> Durante a escolha dos *corpora* a serem utilizados nesta pesquisa, entramos em contato com três projetos: o *Corpus* Brasileiro, coordenado pelo pesquisador Tony Berber Sadinha; o projeto DISSERTES, coordenado pelo pesquisador João Bosco Figueiredo Gomes e o projeto Lácio-Web, coordenado pela pesquisadora Sandra Maria Aluísio. O primeiro é composto por *corpus* escrito e falado e subdividido em 13 *subcorpora* escritos e 25 *subcorpora* falados, mas não atendeu aos nossos requisitos por estar estratificado em sentenças e não em nível textual para que pudessemos analisar como a cadeia anafórica é construída ao longo do texto. O segundo não nos foi disponibilizado pelo coordenador que, ao nos enviar uma resposta, afirmou que o *corpus*

O projeto Lácio-Web foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) com parceria entre o Núcleo Interinstitucional de Linguística Computacional (NILC) e a Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH) da Universidade de São Paulo (USP) sob a orientação da pesquisadora Sandra Maria Alúcio e teve como principal objetivo divulgar e disponibilizar livremente na internet um *corpus* linguístico escrito que obedecesse aos parâmetros internacionais de constituição de *corpora*<sup>175</sup>.

O Lácio-Web é composto por seis *corpus* – Lácio-Ref, Lácio-Dev, Par-C, Comp-C, Mac-Morpho e Lácio-Sint. O Lácio-Ref<sup>176</sup> é um *corpus* aberto composto de textos de português brasileiro, tendo como característica serem escritos respeitando a norma culta; o Lácio-Dev é um *corpus* aberto composto de textos produzidos por candidatos aos vestibulares e graduandos da FFLCH-USP e por não serem revisados em relação à norma culta; o Par-C é um *corpus* aberto e paralelo Inglês-Português; o Comp-C é um *corpus* aberto e formado por textos originais para a comparação inglês/português e a ser estudado pelos critérios de gênero, tipo de texto e/ou data; o Mac-Morpho é um *corpus* fechado e anotado sendo composto por textos jornalísticos, contendo 1.167.183 palavras a serem analisadas morfossintaticamente e; o Lácio-Sint é um *corpus* fechado e etiquetado composto por textos de diversos gêneros.

A fim de delimitar os *corpora* para atender às nossas especificidades, optamos pela análise do *corpus* Lácio-Ref, já que há textos completos, sem notação, mas divididos por gênero discursivo e, sobretudo, por serem textos escritos em norma culta, o que poderia simplificar a análise de como os nomes e os pronomes estão distribuídos ao longo de um texto.

Abaixo, mostramos como está organizado o *corpus* Lácio-Web, conforme pode ser encontrado no *site* do projeto<sup>177</sup>.

---

de sua equipe não serviria ao nosso objetivo e, o terceiro, que atendeu às nossas necessidades, está descrito nesta seção.

<sup>175</sup> Para maiores informações, ver Alúcio, Oliveira e Pinheiro (2004).

<sup>176</sup> O Lácio-Ref é um *corpus* cru, ou seja, nele não há anotações das informações morfossintáticas, sintáticas ou de nível elevado, mas contém anotações de elementos gráficos. Neste *corpus*, a grande maioria dos textos está disponível de maneira integral e foram catalogados em três níveis a fim de atingirem o padrão de disponibilização: compilação-formatação, nomeação e catalogação. Para mais informações, ver Alúcio *et al.* (2003) e Alúcio *et al.* (2004).

<sup>177</sup> Para maiores informações, acessar <<https://sites.google.com/site/linguacorporus/lacioweb>> (último acesso em: 12/05/19).

**Quadro 06** – Composição do *corpus* Lácio-Ref (modificado)

<b>TEXTOS</b>	<b>GÊNERO</b>	<b>CARACTERÍSTICA</b>	<b>DISPONIBILIZAÇÃO</b>
1 ano da Revista Nova Escola	Informativo	Revista voltada para o trabalho docente e variedades relacionadas ao ambiente escolar	primeiro lançamento (20/01/2004)
1 ano da Revista Fapesp	Informativo	Revista de divulgação dos projetos científicos e tecnológicos financiados pela FAPESP	primeiro lançamento (20/01/2004)
1 ano da Folha de São Paulo	Informativo	Jornal brasileiro de grande circulação	segundo lançamento* (30/06/2004)
2 anos do Jornal Jorusp	Informativo	Jornal publicado pela Universidade de São Paulo (USP)	primeiro lançamento (20/01/2004)
1 ano da Revista Super Interessante	Informativo	Revista com temas científicos para divulgação ao grande público	Disponibilização Futura
2 edições da Revista das Religiões (1 e 2)	Informativo	Revista sobre as diversas religiões no mundo	Disponibilização Futura
8 anos do Informativo José Reis	Informativo	Boletim informativo da Escola de Comunicação e Artes (ECA), da Universidade de São Paulo	primeiro lançamento* (20/01/2004)
Teses e dissertações do site SABER (www.saber.usp.br)	Científico	Material científico produzido na Universidade de São Paulo	segundo lançamento* (30/06/2004)
<i>Caminhos da análise de acidentes de trabalho</i>	Científico	Livro, na área de saúde pública, produzido pelo professor Ildeberto Muniz de Almeida, da Universidade de São Paulo (USP)	Disponibilização Futura
8 anos da Revista Cerâmica Industrial	Científico	Revista da Associação Brasileira de Cerâmica para a divulgação de trabalhos científicos e tecnológicos	primeiro lançamento* (20/01/2004).
<i>A Economia da natureza</i>	Instrucional	Trechos do livro didático de Ensino Superior publicado pela Editora Guanabara/Koogan em 1993	segundo lançamento (30/06/2004).
<i>Ecologia</i>	Instrucional	Trechos do livro didático de Ensino Superior publicado pela Editora Guanabara em 1983	segundo lançamento (30/06/2004)
Publicações do NILC	Científico	Material científico de diversos tipos de texto produzidos pelos membros do NILC	segundo lançamento (30/06/2004)
Publicações do TIL 2003	Científico	Artigos do Primeiro workshop em <i>Tecnologia da Informação e da Linguagem Humana</i> (TIL), realizado em 2003 no NILC	segundo lançamento (30/06/2004)
Revista Brasil de Literatura	Científico	Revista de divulgação de textos científicos (ensaios e resenhas) sobre literatura, além de textos literários, propriamente, separados	primeiro lançamento (20/01/2004)

		nos gêneros poesia, prosa e drama	
Um montante de textos Jurídicos	Jurídico	Literatura jurídica com diversos tipos de texto, disponíveis na Web.	segundo lançamento* (30/06/2004)
Literatura de Ficção	Literário	Literatura de ficção separada nos gêneros poesia, prosa e drama	primeiro lançamento (20/01/2004)
Revista Brasil de Literatura	Literário	Textos literários (poesias e crônicas)	primeiro lançamento (20/01/2004)

(Fonte: Lácio-Web)

Para nossa pesquisa, não levamos em consideração a região na qual o texto foi produzido por não estarmos buscando evidências de variação ou de mudança linguística, mas buscamos respeitar, tanto para o *corpus* escrito quanto para o *corpus* falado, a característica do texto, pois as pesquisas psicolinguísticas usam *designs* experimentais com previsões de produção respeitando a norma culta, a escolaridade, pois é habitual que os participantes de experimentos psicolinguísticos estejam cursando a graduação ou já sejam graduados; e, também, o gênero, já que, nas análises descritas a seguir, havia gêneros que enquadrariam as duas características anteriores e, assim, optamos por analisar textos informativos.

Dentre os textos do *corpus* Lácio-Ref que se enquadram no gênero informativo, há oito *subcorpus*, sendo seis deles ligados à divulgação específicas e/ou técnico-científicas (Revista Nova Escola, Revista Fapesp, Revista Super Interessante, Revista das Religiões, Informativo José Reis e Revista Fapesp em inglês e português) e dois ligados à temas gerais (Folha de São Paulo e Jorusp). Optamos, então, por selecionar textos ligados a temas mais gerais e, dentre os textos do *subcorpus* Folha de São Paulo e do *subcorpus* Jorusp, selecionamos este último, por crermos que, na redação deste jornal trabalham graduandos, graduados e demais colaboradores, o que desemboca em um nível de escolaridade interessante para nossa pesquisa, enquanto os jornalistas, editores e colaboradores do jornal Folha de São Paulo são, no mínimo, graduados.

O *subcorpus* Jorusp, componente do *corpus* Lácio-Ref, é composto por 680 textos divididos em seções de acordo com os cadernos do Jornal da USP, a saber: (i) opinião – 55 textos, (ii) especial – 118 textos, (iii) cultura – 103 textos, (iv) pesquisa – 82 textos, (v) nacional – 69 textos, (vi) universidade – 227 textos e (vii) arquivo – 26 textos.

Para a delimitação final do *corpus* que comporia nossa pesquisa, buscamos, dentro do *subcorpus* Jorusp, os textos que tratassem de temas mais

gerais possível e, assim, optamos pelos textos da seção de cultura. Estes textos foram coletados, de acordo com a descrição no *corpus* entre fevereiro de 2001 e fevereiro de 2003, dos quais optamos por selecionar, em um primeiro momento, por meses que contemplassem o período intermediário da publicação no Jorusp e, em segundo momento, que contivessem uma média de linhas igual.

Após a criação deste parâmetro, chegamos a 10 (dez) textos a serem analisados. Estes textos tinham em média<sup>178</sup>, 1080 palavras, 5622 caracteres sem espaço, 6697 caracteres com espaço, 119 linhas e 17 parágrafos.

A partir das análises preliminares dos textos, de observar se estes obedeciam aos critérios que elencamos e, posteriormente, observar os critérios básicos para a participação dos sujeitos em pesquisas psicolinguísticas e as características dos textos em quantidade de palavras, de caracteres sem e com espaço, de linhas e de parágrafos, conforme descrito acima, partimos para a leitura e coleta dos dados aplicando as categorias de análises que serão apresentadas a seguir.

### 4.3 Categorias de análise

As categorias de análise que constituíram nosso formulário foram construídas a partir das pesquisas psicolinguísticas desenvolvidas em Português Brasileiro sobre a resolução da correferência anafórica (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; QUEIROZ, 2009; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; MAIA; CUNHA-LIMA, 2011; LEITÃO; RIBEIRO; MAIA, 2012; MAIA; CUNHA-LIMA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; MAIA, 2014; LIMA, 2015; GONDIM, 2017)<sup>179</sup> e,

---

<sup>178</sup> Para o cálculo da média de quantidade de palavras, quantidade de caracteres sem ou com espaço, quantidade de linhas e quantidade de parágrafos, foram somadas as quantidades específicas de cada texto até chegar a um resultado final a ser dividido por 10 (dez) que foi a quantidade total de textos analisados. Desta forma, podemos dizer que aplicamos um cálculo de média simples e que, quando necessário, fizemos aproximações para um número exato, como demonstramos agora no cálculo da média de parágrafos para ilustrar como procedemos. Temos, a seguir, a quantidade de parágrafos dos textos de 01 a 10, respectivamente, sendo, posteriormente dividido pela quantidade total de textos e, por fim, aproximado para o número exato mais próximo:  $18 + 17 + 32 + 16 + 11 + 24 + 09 + 09 + 06 + 25 / 10 = 16,7$ , portanto, constando como, a média de 17 parágrafos para os textos analisados.

<sup>179</sup> É necessário atentar para o fato de que não há um consenso nas pesquisas citadas, pois pode-se encontrar pesquisas que tratam sobre a resolução anafórica, resolução correferencial ou, ainda, resolução da correferência anafórica, termo pelo qual optamos por considerarmos o mais atual nas pesquisas desenvolvidas no Brasil.

assim, chegamos aos campos que serão explicados abaixo com vistas ao cumprimento de nossos objetivos nesta tese.

Contudo, antes de continuarmos a descrição das categorias de análise, é necessário dizer que, durante nossas leituras prévias dos textos do *corpus* para observar se estavam adequados a nossos critérios, percebemos que, ao longo da cadeia de referência no texto, um tipo diferente de nome surgiu, desempenhando a mesma função do nome repetido. Estes nomes eram por vezes sinônimos, antônimos, hiperônimos ou hipônimos<sup>180</sup>.

Levando em consideração que os nomes, ao introduzirem um referente no texto, podem ser retomados pelo mesmo item lexical ou por outras expressões nominais, entendemos a função recategorizados dos nomes durante a construção de retomada pelo emissor (KOCH, 2014; BASSETO, 2015; MAGALHÃES; BRITO, 2016). Porém, como estamos estudando o texto produzido, não a produção em si, acreditamos que o processo de recategorização já está finalizado, dando-nos o nome estável e já recategorizado no texto, o que nos levou a criar a subcategoria nomes recategorizados dentro da categoria dos nomes.

Além disso, notamos que, o que é conhecido nas pesquisas psicolinguísticas como nome repetido não aparece sempre da mesma maneira ao longo do texto, pois os nomes podem ser retomados de maneira idêntica ou quase idêntica, como ocorre com nomes próprios que podem aparecer completos – *Maria Gabriela*, reduzidos – *a Maria*, ou como apelidos – *a Gabi*, por exemplo e, assim, decidimos chamar de nomes repetidos os nomes idênticos ou quase idênticos que remetem ao mesmo antecedente, o que, notoriamente, difere dos nomes recategorizados que poderiam retomar *Maria Gabriela* por *a prima do Bruno* ou *a minha professora*<sup>181</sup>, por exemplo.

---

<sup>180</sup> Não é de nosso interesse atentar para a complexidade semântica existente entre as opções que o falante dispõe na escolha de um nome ou de outro para se referir ao antecedente, mas a como o falante constrói, por meio de descrições nominais, o sentido do texto, guiando o seu interlocutor a compreender o objeto do discurso.

<sup>181</sup> É importante frisar que, mesmo ocorrendo descrições nominais do tipo *a prima do Bruno* ou *a minha professora*, consideramos o núcleo do sintagma para então quantificá-lo. Ao longo de toda a nossa pesquisa, palavras que desempenhavam a função de adjunto adnominal (ou *spec*) foram desconsideradas e, por conta disso, alguns antecedentes não foram contabilizados por estarem em função de adjunto adnominal e não de núcleo do sujeito. Esta decisão foi tomada com vistas a respeitar o máximo possível o efeito de Proeminência Sintática (GORDON; HENDRICK, 1997).

Uma vez feitos estes comentários pertinentes para a compreensão da análise feita por nós e que terão os resultados mostrados e discutidos nas próximas seções, prosseguimos explicando as categorias de análise.

Para melhor compreensão de nossas análises, ao longo desta seção, a título ilustrativo, serão mostrados *prints* de nossos documentos de *Word* com imagens das categorias de análise sem a quantidade de ocorrências que foram sendo preenchidas em nossos formulários durante a coleta de dados.

Como é de consenso na área da Psicolinguística, as pesquisas que tratam da resolução da correferência anafórica, durante a explicação da metodologia, reportam a quantidade de palavras e, às vezes, a quantidade de caracteres existentes entre o termo antecedente e a sua expressão correferente, sendo, por isso, esta a nossa primeira categoria de análise: a quantidade de palavras.

Para a contagem de ocorrência do uso das expressões correferentes com relação à quantidade de palavras existentes entre o antecedente e a retomada, criamos intervalos a serem marcados para indicar a quantidade de ocorrências.

**Quadro 07** – Quantidade de palavras e de caracteres sem e com espaço entre o termo antecedente e a expressão correferente

<b>Retomada por nome repetido – idêntico ou quase idêntico</b>	<b>Retomada por pronome pleno</b>
( ) De 0 até 10 palavras	( ) De 0 até 10 palavras
( ) De 11 até 50 palavras	( ) De 11 até 50 palavras
( ) De 51 até 100 palavras	( ) De 51 até 100 palavras
( ) De 100 palavras em diante	( ) De 100 palavras em diante
( ) De 0 até 10 caracteres (sem espaço)	( ) De 0 até 10 caracteres (sem espaço)
( ) De 11 até 50 caracteres (sem espaço)	( ) De 11 até 50 caracteres (sem espaço)
( ) De 51 até 100 caracteres (sem espaço)	( ) De 51 até 100 caracteres (sem espaço)
( ) Mais de 100 caracteres (sem espaço)	( ) Mais de 100 caracteres (sem espaço)
( ) De 0 até 10 caracteres (com espaço)	( ) De 0 até 10 caracteres (com espaço)
( ) De 11 até 50 caracteres (com espaço)	( ) De 11 até 50 caracteres (com espaço)
( ) De 51 até 100 caracteres (com espaço)	( ) De 51 até 100 caracteres (com espaço)
( ) Mais de 100 caracteres (com espaço)	( ) Mais de 100 caracteres (com espaço)
<b>Retomada por nome repetido – nome recategorizado</b>	<b>Retomada por pronome nulo</b>
( ) De 0 até 10 palavras	( ) De 0 até 10 palavras
( ) De 11 até 50 palavras	( ) De 11 até 50 palavras
( ) De 51 até 100 palavras	( ) De 51 até 100 palavras
( ) De 100 palavras em diante	( ) De 100 palavras em diante
( ) De 0 até 10 caracteres (sem espaço)	( ) De 0 até 10 caracteres (sem espaço)
( ) De 11 até 50 caracteres (sem espaço)	( ) De 11 até 50 caracteres (sem espaço)
( ) De 51 até 100 caracteres (sem espaço)	( ) De 51 até 100 caracteres (sem espaço)
( ) Mais de 100 caracteres (sem espaço)	( ) Mais de 100 caracteres (sem espaço)
( ) De 0 até 10 caracteres (com espaço)	( ) De 0 até 10 caracteres (com espaço)
( ) De 11 até 50 caracteres (com espaço)	( ) De 11 até 50 caracteres (com espaço)
( ) De 51 até 100 caracteres (com espaço)	( ) De 51 até 100 caracteres (com espaço)
( ) Mais de 100 caracteres (com espaço)	( ) Mais de 100 caracteres (com espaço)

(Fonte: Os autores)

Como é possível ver no quadro 07, analisamos intervalos com variações de 01 a 10 palavras, 11 a 50 palavras, 51 a 100 palavras e mais de 100 palavras

entre o antecedente e o nome ou o pronome que desempenha a função de expressão correferente.

Uma vez que a quantidade de palavras e caracteres foi contabilizada em quantidade de ocorrências, partimos para as análises de cunho sintático-semântico a fim de encontrar indícios que nos levassem a compreender melhor como o falante retoma os antecedentes ao longo do texto e, assim, fazer uma discussão com os achados das pesquisas psicolinguísticas anteriormente citadas.

A primeira categoria por nós analisada foi o efeito de Paralelismo Sintático (CHAMBERS; SMITH, 1998) a fim de averiguar se este efeito pode ser notado em quantidade de ocorrências. Além disso, buscamos analisar se a forma mais proeminente tende a ser retomada, o que implicaria no aparecimento de mais sujeitos sendo retomados do que outras posições menos proeminentes, conforme previsto por Gordon e Hendrick (1997) ao postular o efeito da Proeminência Sintática. Por conta disto, subdividimos a categoria em possíveis formas de entrada do antecedente, tendo o antecedente aparecido na posição sintática de sujeito, de objeto, de adjunto adnominal ou de adjunto adverbial<sup>182</sup>, como pode ser visto no quadro abaixo.

**Quadro 08** – Averiguação do efeito de paralelismo sintático e do efeito de proeminência sintática

Em termos de estrutura sintática:			
<u>Paralelismo sintático</u>			
(	) sim	(	) não
<u>Entrada do referente como</u>			
(	) sujeito	(	) adj. <u>adn</u>
(	) adj. <u>adv</u>	(	) objeto

(Fonte: Os autores)

Na categoria seguinte, optamos por contabilizar a quantidade total de retomadas efetuadas por nomes repetidos, por nomes recategorizados, por pronomes plenos e por pronomes nulos, como pode ser visto no quadro 09.

<sup>182</sup> Outras possíveis posições sintáticas surgiram, como o antecedente estar dentro de aposto ou deslocado em expressões do tipo *segundo o autor, conforme João, etc.* e, por serem poucas ocorrências e não estarem diretamente ligados ao nosso objetivo, resolvemos excluir de nossas análises.

**Quadro 09** – Expressão correferente na posição de sujeito

Tipo de retomada	
(        ) nome repetido	{        } nome recategorizado
(        ) pronome nulo	{        } pronome pleno

(Fonte: Os autores)

Logo após a contabilização do tipo de retomada que apareceu na posição de sujeito, analisamos a quantidade e o tipo de sentença entre o antecedente e a retomada tendo criado, inicialmente, três subcategorias: (i) oração subordinada<sup>183</sup>, (ii) oração coordenada justaposta – orações coordenadas separadas por vírgulas ou unidas por conjunção<sup>184</sup>, mas não separadas em períodos distintos, e (iii) oração coordenada em períodos distintos – orações coordenadas separadas entre si por ponto final, mas que pertencem ao mesmo parágrafo.

Por trabalhar com a cadeia anafórica ao longo do texto, percebemos que era necessário analisar a retomada também entre os parágrafos e, por isso, criamos mais uma subcategoria: (iv) parágrafos distintos – esta subcategoria nos levou a observar como ocorre a primeira retomada do antecedente ao mudar de parágrafo.

Para cada uma das quatro subcategorias acima, foi possível contabilizar no formulário que tipo de expressão correferente apareceu em dado tipo de sentença/parágrafo, havendo a opção de contabilizar a quantidade de nomes repetidos, nomes recategorizados, pronomes plenos e pronome nulos, como podemos ver no quadro 10.

<sup>183</sup> Optamos por não analisar o tipo de oração subordinada em que se encontrava o nome repetido, o nome recategorizado, o pronome pleno ou o pronome nulo com função de correferente, por acreditarmos que não é de nosso objetivo, nesta pesquisa, mas contabilizamos todas as orações subordinadas em que uma das expressões correferentes por nós estudada apareceu.

<sup>184</sup> Em nossa pesquisa, optamos, também, por não atentar para a quantidade de conjunções que interligam orações coordenadas justapostas, uma vez que trabalhos anteriores (QUEIROZ; LEITÃO, 2008; QUEIROZ, 2009; LEITÃO; RIBEIRO; MAIA, 2012; GONDIM, 2017), dentre outros objetivos, averiguaram a influência da presença ou ausência da conjunção durante a resolução da correferência anafórica.

**Quadro 10** – Tipo de sentença entre o antecedente e a retomada

<u>Quantidade/tipo de orações entre a introdução do referente e a retomada</u>	
(        )	oração subordinada
	[pronome pleno ( ) / pronome nulo ( ) / nome repetido ( ) / nome recategorizado ( )]
(        )	oração coordenada [or. justaposta]
	[pronome pleno ( ) / pronome nulo ( ) / nome repetido ( ) / nome recategorizado ( )]
(        )	oração coordenada em períodos distintos
	[pronome pleno ( ) / pronome nulo ( ) / nome repetido ( ) / nome recategorizado ( )]
(        )	em parágrafos distintos
	[pronome pleno ( ) / pronome nulo ( ) / nome repetido ( ) / nome recategorizado ( )]

(Fonte: Os autores)

É preciso observar que as categorias de análise que criamos depois desta anterior não aparecem discutidas nas pesquisas que embasaram a criação de nosso formulário, mas surgem comentadas em quase todas pelo fato de que, em geral, são controlados<sup>185</sup> os tipos de verbo<sup>186</sup> presentes na sentença anterior à retomada e, por isso, analisamos este critério também, como pode ser visto no quadro a seguir.

**Quadro 11** – Tipo de verbo presente na sentença anterior à retomada

<u>Verbo da sentença anterior</u>	
(        )	verbo intransitivo
(        )	verbo transitivo
(        )	verbo de ligação

(Fonte: Os autores)

Para finalizar a explicação acerca das categorias de análise que compuseram os formulários que nos conduziram aos resultados que serão discutidos na seção seguinte, decidimos analisar o tipo de artigo precedente ao termo antecedente.

<sup>185</sup> Dentre as pesquisas revisitadas no capítulo anterior, a maioria delas faz uso de verbos transitivos diretos em seus designs experimentais, mas não há discussão ou justificativa do motivo de uso deste tipo de verbo ou de outros nas pesquisas e, por conta disto, resolvemos contabilizar o tipo de verbo que aparece na sentença anterior.

<sup>186</sup> Para o tipo de verbo analisado na sentença anterior, havíamos, inicialmente, dividido os verbos transitivos em (i) transitivos diretos, (ii) transitivos indiretos e (iii) bitransitivos. Ao final, por notarmos que alguns tipos eram tão pouco frequentes, optamos por enquadrá-los todos em uma mesma categoria, deixando para apontar a quantidade de ocorrências de cada um destes tipos especificamente na seção de resultados.

Sabemos que o português admite dois tipos de artigo precedendo substantivos, compondo sintagma nominal, sendo eles: (i) o artigo definido e (ii) o artigo indefinido. Contudo, percebemos que, ao longo do texto, um mesmo referente pode ora ser retomado por descrições nominais, ora por pronomes plenos ou por pronomes nulos e, assim, o antecedente passa a não ter nenhum artigo precedente ou pode ser precedido por outras palavras como pronomes indefinidos ou numerais, por exemplo<sup>187</sup>. Por conta disto, optamos por descrever esta categoria em três subcategorias: (i) artigo definido, (ii) artigo indefinido ou (iii) nenhum artigo, como pode ser visto no quadro seguinte.

**Quadro 12** – Tipo de artigo existente no SN

<u>Antecedente precedido por artigo</u>		
( ) definido	{ } indefinido	( ) nenhum

(Fonte: Os autores)

Concluimos, assim, a fase de descrição do percurso que fizemos desde seleção do *corpus* a ser utilizado em nossa pesquisa até a maneira que fomos criando as categorias de análise presentes em nosso formulário que foi preenchido durante a coleta de dados no *corpus* escrito, o que nos proporcionou chegar aos resultados que serão expostos logo a seguir.

#### 4.4 Resultados

Os resultados encontrados a partir da coleta de dados no *corpus* escrito Lácio-Ref que serão reportados logo a seguir servem para, na próxima seção, levar a uma discussão embasada em nossa coleta feita por meio de formulários a fim de quantificar as ocorrências dos nomes repetidos, dos pronomes plenos e dos pronomes nulos que funcionam como expressão correferente de um antecedente.

Nossos resultados serão apresentados por meio de gráficos que foram possíveis graças ao fato de termos feito o preenchimento dos formulários em

<sup>187</sup> Como nossa pesquisa não trata sobre o quão definida ou indefinida dada expressão ou palavra pode ser em português, optamos por, de maneira simples, observar se o antecedente é precedido por artigo definido ou indefinido e, quando precedido por qualquer outro tipo de palavra ou expressão, ou mesmo quando o último antecedente for um pronome pleno ou nulo, enquadrando-os na subcategoria de nenhum.

documentos do formato *Word* e transposto os valores para planilhas de *Excel*, conforme ilustrado por meio da figura abaixo e, assim, dando-nos possibilidade de construção dos gráficos.

**Quadro 13** – Exemplo da planilha em Excel do parâmetro de contabilização dos dados com ou sem o efeito de Paralelismo Sintático

Entrevista	Nº	T	Parâmetro	Paralelismo sintático
01.JO-IF-JORUSP-cul-02set02_01	1		Existe	21
			Não Existe	10
02.JO-IF-JORUSP-cul-02set02_02	2		Existe	39
			Não Existe	9
03.JO-IF-JORUSP-cul-02set02_03	3		Existe	14
			Não Existe	12
04.JO-IF-JORUSP-cul-03dez01_03	4		Existe	35
			Não Existe	4
05.JO-IF-JORUSP-cul-03jun02_01	5		Existe	40
			Não Existe	12
06.JO-IF-JORUSP-cul-02set02_03	6		Existe	30
			Não Existe	3
07.JO-IF-JORUSP-cul-02set02_04	7		Existe	31
			Não Existe	9
08.JO-IF-JORUSP-cul-03jun02_02	8		Existe	35
			Não Existe	15
09.JO-IF-JORUSP-cul-03jun02_03	9		Existe	7
			Não Existe	3
10.JO-IF-JORUSP-cul-03jun02_04	10		Existe	29
			Não Existe	3

(Fonte: Os autores)

Cada categoria de análise empreendida por nós nesta pesquisa e descritas na seção anterior continuam uma planilha de *Excel* e, para cada planilha, valores específicos foram colocados durante a coleta de dados. Assim, fomos capazes de comparar a quantidade de cada fator, conforme reportaremos nesta seção.

Como nosso trabalho trata da distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito em Português Brasileiro, optamos por descrever as frequências de uso encontradas em cada texto e somá-las para observar quais as formas mais frequentes e, posteriormente, comparar quais formas foram mais usadas. Entretanto, uma vez que os textos analisados por nós, em geral, têm tamanhos variados, foi necessário dar tratamento estatístico em busca de valores de significância entre as formas aparentemente mais frequentes com relação às menos frequentes.

Pelo fato de termos contabilizado, em nossas planilhas de Excel, a quantidade exata de ocorrência para cada fator analisado, utilizamos o programa estatístico PAST (Paleontological Statistics Software) desenvolvido por Hammer,

Harper e Ryan (2001) para, nos casos em que há cruzamento de variáveis, como no caso da possível relação entre a forma anafórica usada na posição de sujeito e a distância entre o referente e o antecedente ou, no caso da relação entre a forma anafórica e a disposição textual, ser aplicado o Qui-quadrado. Quando buscamos comparar os resultados das formas entre si, aplicamos análise de variância, mais especificamente, o test-T, para, assim, termos o p-valor que poderia apontar para a significância ou não dos resultados encontrados.

Por considerarmos o tipo de expressão correferencial que retoma o termo antecedente o ponto de partida de nosso trabalho, começaremos reportando o resultado em quantidade de ocorrências para as duas macrocategorias que funcionam como formas correferentes no texto escrito: os nomes e os pronomes.

Os três primeiros gráficos nos levam a analisar inicialmente como nomes e pronomes aparecem nos textos em termos de quantidade (gráfico 01). Depois subdividimos a categoria dos pronomes para descrever, conforme estudado em pesquisas psicolinguísticas sobre a correferência anafórica em PB, como são usados os nomes repetidos, os pronomes plenos e os pronomes nulos (gráfico 02). Por fim, subdividimos a categoria de nomes para reportar como são usados os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos (gráfico 03), uma vez que, pelos motivos apontados na seção anterior, tivemos de criar uma nova subcategoria para os nomes.

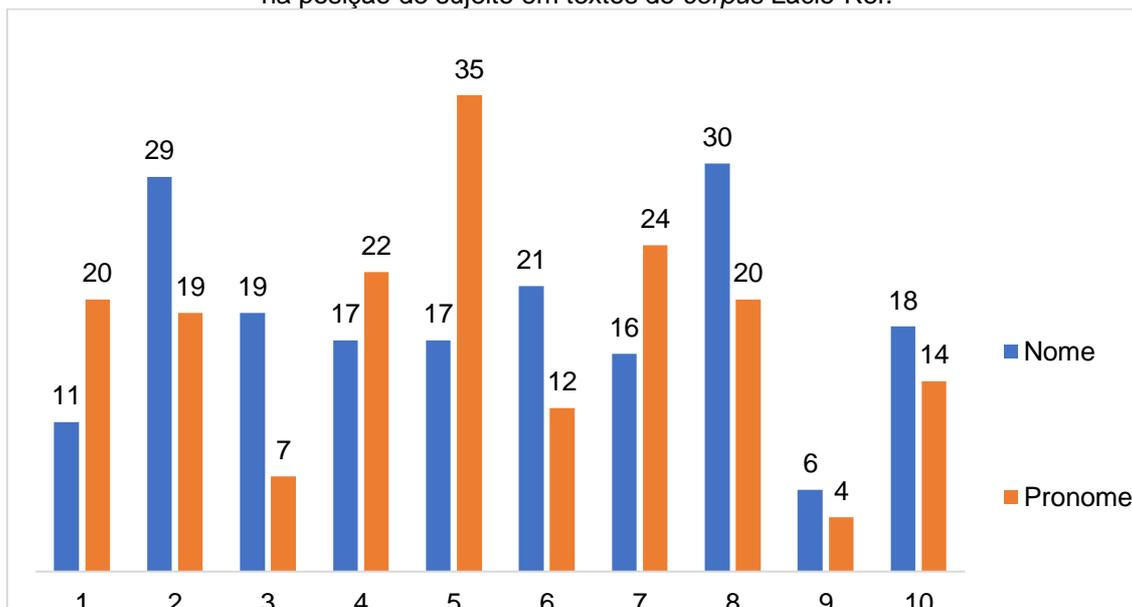
A fim de facilitar a compreensão de nossos resultados, optamos por deixar em cada gráfico a quantidade de ocorrências de cada termo estudado e, durante a descrição, decidimos calcular, em porcentagem<sup>188</sup>, o valor encontrado para, após cada gráfico, tornar mais simples a comparação dos resultados que nos proporcionaram a discussão subsequente. Além do valor em porcentagem, mostramos, para cada forma mais frequente, a comparação com as demais tanto em valor de frequência quanto o p-valor, a fim de observar se, quando levada em consideração a variabilidade dos textos analisados, há significância estatística para os resultados encontrados.

Uma vez feitos estes comentários que consideramos pertinentes, mostraremos, no gráfico 01, como nomes e pronomes são usados, em termos de quantidade de ocorrências, na posição de sujeito em textos escritos em PB.

---

<sup>188</sup> Para reportar a frequência simples através de porcentagem, os valores foram calculados por meio de regra de três simples.

**Gráfico 01** – Quantidade de ocorrência de nomes e pronomes como expressões correferentes na posição de sujeito em textos do *corpus* Lácio-Ref.



A fim de condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 361 retomadas na posição de sujeito, sendo 184 por nomes e 177 por pronomes, implicando em 51% das retomadas efetuadas por nomes e 49% por pronomes. A diferença de frequência de uso de 0,2 percentuais, com p-valor = 0,14, não aponta para diferenças significativa quando levadas em consideração as ocorrências de nomes com relação a pronomes nos textos escritos analisados.

Desta forma, é possível que o usuário da língua, de maneira geral, faça uso tanto do nome quanto do pronome no centro anafórico marcado<sup>189</sup> em (124).

(124) [...] Jó<sub>i</sub> não é um livro<sub>i</sub> sobre ciência. Um dos mais antigos textos do Velho Testamento<sub>i</sub>, cujo autor pode ter sido contemporâneo do patriarca Abraão, que viveu há quase 4 mil anos, ele<sub>i</sub> relata a história de um homem rico e piedoso que, fustigado pelo diabo, perdeu tudo o que possuía. [...]

(JO-IF-JORUSP-cul-02set02\_03)

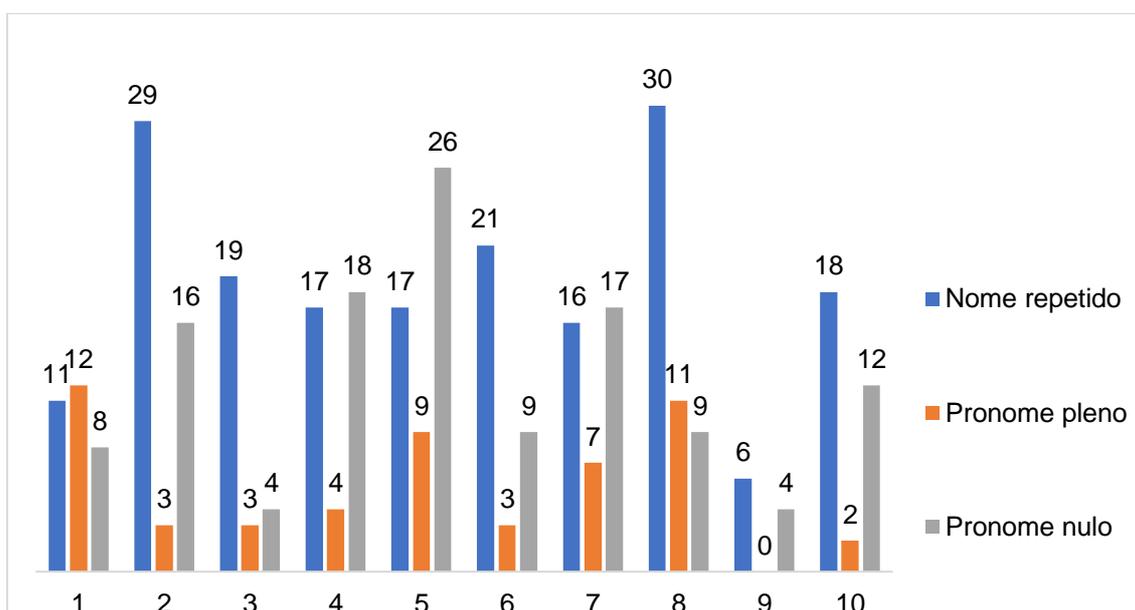
Além do resultado geral e levando em consideração a grande variabilidade dos textos analisados, é possível notarmos que há grande variação de uso entre os nomes e pronomes por parte de cada usuário a partir dos dados

<sup>189</sup> Durante a exposição de nossos resultados, mostraremos a marcação de indicação da correferência entre o referente e a forma anafórica usada para a retomada no exemplo dado.

encontrados nos textos separadamente, havendo sempre a preferência de um tipo de forma correferente por parte do usuário.

O gráfico 02, mostra a categoria de pronome subdividida em pronomes plenos e pronomes nulos, conforme estudado nas pesquisas psicolinguísticas que nos motivaram a fazer o presente estudo para que assim, possamos analisar, em termos de quantidade de ocorrências, como são usados os nomes (englobando aqui tanto nomes repetidos quanto nomes recategorizados), os pronomes plenos e os pronomes nulos em textos escritos em PB.

**Gráfico 02** – Quantidade de ocorrência de nomes, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do *corpus* Lácio-Ref



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 361 retomadas na posição de sujeito, sendo 184 por nomes e 177 por pronomes. Destas retomadas, foram feitas 54 por pronomes plenos e 123 por pronomes nulos, implicando em 51% das retomadas efetuadas por nomes, 15% por pronomes plenos e 34% por pronomes nulos.

Considerando os resultados mostrados no gráfico 02, optamos por considerar apenas os resultados da categoria de pronomes, neste momento, já que os resultados da categoria de nomes não foram subdivididos, estando inalterado como no gráfico 01. Já a categoria de pronomes, por ter sido subdividida entre as duas formas correferentes possíveis – pronomes plenos e pronome nulos – traz resultados interessantes.

Podemos, através destes resultados, verificar, a partir da quantidade de ocorrências, que os pronomes nulos tendem a retomar mais antecedentes em textos escritos, com 69,5% das retomadas por pronomes nulos, do que os pronomes plenos, com 30,5%, mas que não há diferença significativa (p-valor = 1,18) entre o uso do pronome nulo com relação ao pronome pleno, quando levada em consideração a variabilidade dos textos analisados.

Estes resultados indicam que, levando em conta a cadeia referencial ao longo do texto, é mais frequente o uso de pronomes nulos do que o uso de pronomes plenos nos centros anafóricos na posição de sujeito de textos escritos, de maneira geral, o que pode ser percebido em (125) durante o rastreamento das retomadas do referente *Jó*.

(125)<sup>190</sup> [...] *Jó*<sub>i</sub> não é um livro<sub>i</sub> sobre ciência. Um dos mais antigos textos do Velho Testamento<sub>i</sub>, cujo autor pode ter sido contemporâneo do patriarca Abraão, que viveu há quase 4 mil anos, ele<sub>i</sub> relata a história de um homem rico e piedoso que, fustigado pelo diabo, perdeu tudo o que possuía. Apesar das enormes dificuldades e da censura dos amigos que viam sua miséria como fruto de uma conduta má, *Jó* manteve uma inabalável confiança no Criador, que depois lhe restituiu os bens e a alegria. O livro<sub>i</sub>, portanto, procura fortalecer a fé do leitor, Ø<sub>i</sub> consolar os aflitos e Ø<sub>i</sub> celebrar o cuidado de Deus para com seus filhos. Nada a ver com dados científicos. [...]

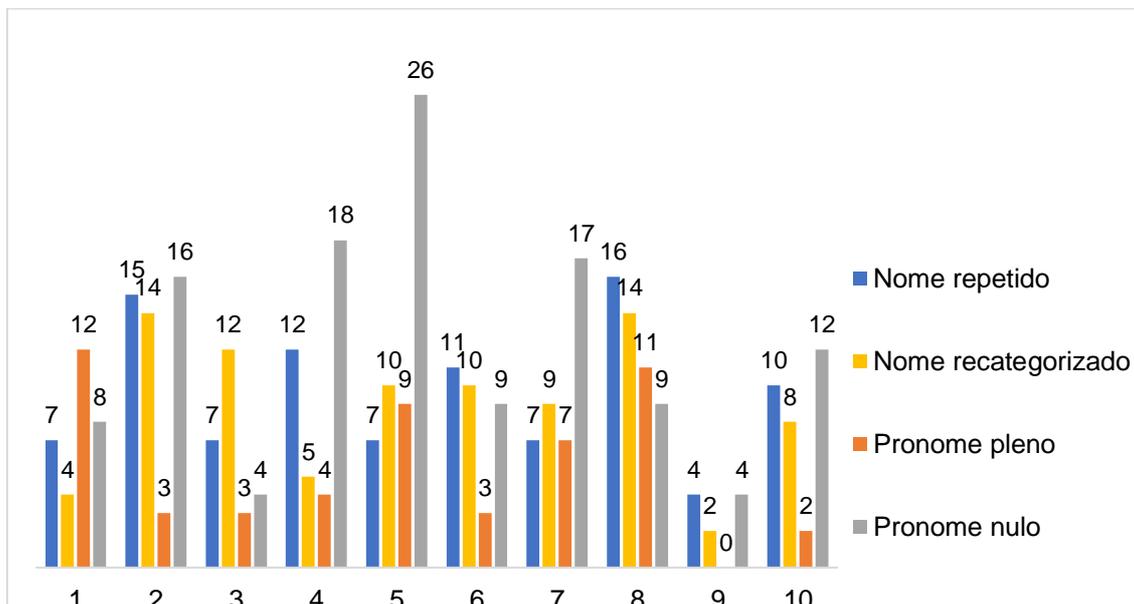
(JO-IF-JORUSP-cul-02set02\_03)

Para finalizar os gráficos que mostram a quantidade de retomadas para cada expressão correferencial estudada por nós neste trabalho, apresentamos a seguir o gráfico em que subdividimos a categoria de nomes em nomes repetidos e nomes recategorizados, conforme explicado anteriormente neste mesmo capítulo.

---

<sup>190</sup> Reiteramos que, neste momento, questões sintáticas não foram levadas em conta, pois era de interesse observar, inicialmente, em termos de frequência, quais as formas anafóricas mais comuns na cadeia correferencial ao longo do texto. Questões como tipo de oração e distância entre o referente e sua retomada anafórica serão mostradas ao longo desta seção.

**Gráfico 03** – Quantidade de ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do *corpus* Lácio-Ref



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 361 retomadas na posição de sujeito, sendo 184 por nomes e 177 por pronomes. Das retomadas feitas por nomes, foram feitas 96 por nomes repetidos e 88 por nomes recategorizados, implicando em 26,6% das retomadas efetuadas por nomes repetidos, 24,4% por nomes recategorizados, 15% por pronomes plenos e 34% por pronomes nulos. Por, neste momento, levarmos em consideração o uso dos nomes repetidos com relação ao uso dos nomes recategorizados, temos indícios de que não há diferença significativa ( $p$ -valor = 0,58) entre o uso destas formas, quando levada em consideração a variabilidade dos textos analisados.

Mostramos, em (126), um exemplo de retomada sendo feita por meio do uso do nome recategorizado *coordenadora* para retomar o referente *Maria Imaculada*.

(126) [...] Maria Imaculada<sub>i</sub> disse que a consulta a boas fontes bibliográficas de biblioteconomia ajudou a indicar os temas que deveriam ser avaliados. A cada 15 dias o formulário será modificado, a fim de que todos os temas sejam abordados. A coordenadora<sub>i</sub> do projeto explicou que isso faz parte do método adotado, pois, como há muitos itens, um questionário só seria grande e cansativo para ser respondido de uma vez. [...]

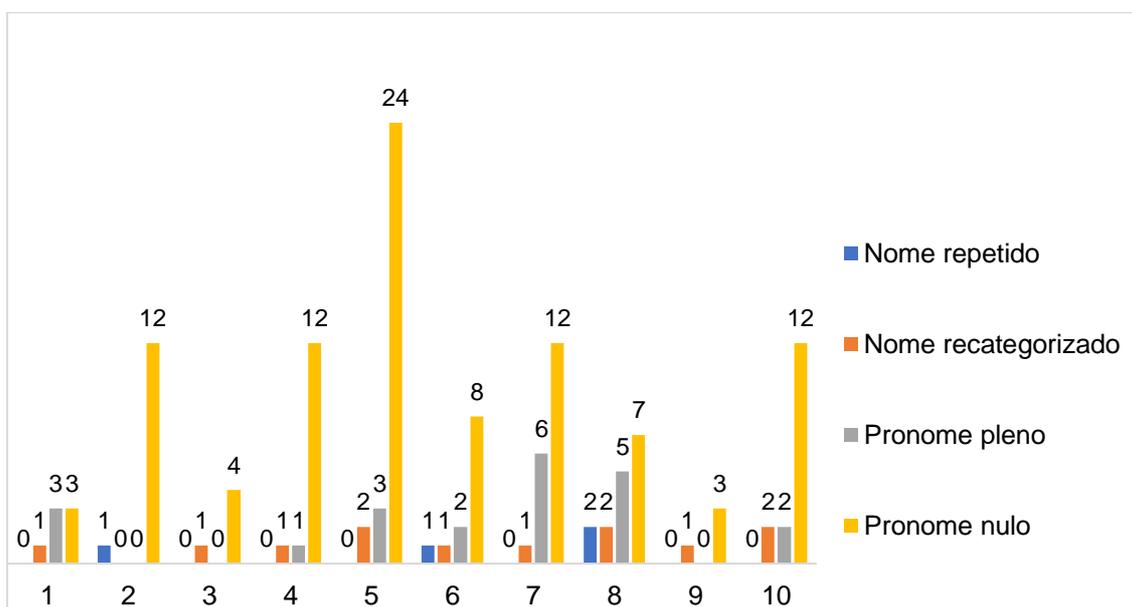
(JO-IF-JORUSP-cul-02set02\_02)

Após a conclusão da exposição dos gráficos sobre o tipo de expressão correferencial preferencialmente usada na posição de sujeito em dados do *corpus* Lácio-Ref, iremos mostrar, logo adiante, uma sequência de 04 (quatro) gráficos com os resultados em quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com relação à quantidade de palavras entre o antecedente e a expressão correferente.

Cada gráfico traz os resultados em quantidade de ocorrências das expressões correferenciais estudadas por nós neste trabalho, de acordo com os intervalos mostrados na seção anterior.

Por isto, o primeiro gráfico desta sequência mostra as quantidades de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos quando há, entre o antecedente e a forma correferente, entre 0 (zero) até 10 (dez) palavras.

**Gráfico 04** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há até 10 palavras entre o antecedente a retomada em textos do *corpus* Lácio-Ref



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas no intervalo de 0 até 10 palavras, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 135 retomadas nesta distância, sendo 04 por nomes repetidos, 12 por nomes recategorizados, 22 por pronomes plenos e 97 por pronomes nulos, implicando em 2,9% das

retomadas efetuadas por nomes repetidos, 8,9% por nomes recategorizados, 16,3% por pronomes plenos e 71,9% por pronomes nulos.

Com estes resultados, temos indícios de que, quando o antecedente e a expressão correferente estão próximas em uma quantidade de até 10 (dez) palavras, há grande tendência de surgimento do pronome nulo, uma vez que, o pronome nulo aparece em 96% e os nomes repetidos em apenas 4% das ocorrências. Apesar dessa diferença percentual, o valor de  $p$ , igual a 0,52, não aponta efeito significativo. Para a comparação das demais formas, encontramos indícios de tendência de uso, mas sem  $p$ -valor significativo, quando levada em consideração a variabilidade do textos, nem entre o pronome nulo e o nome recategorizado ( $p$ -valor = 0,69) – 89% das retomadas efetuadas por pronomes nulos e 11% por nomes recategorizados – nem entre o pronome nulo e o pronome pleno ( $p$ -valor = 1,34) – 81,5% das retomadas efetuadas por pronomes nulos e 18,5% por pronomes plenos.

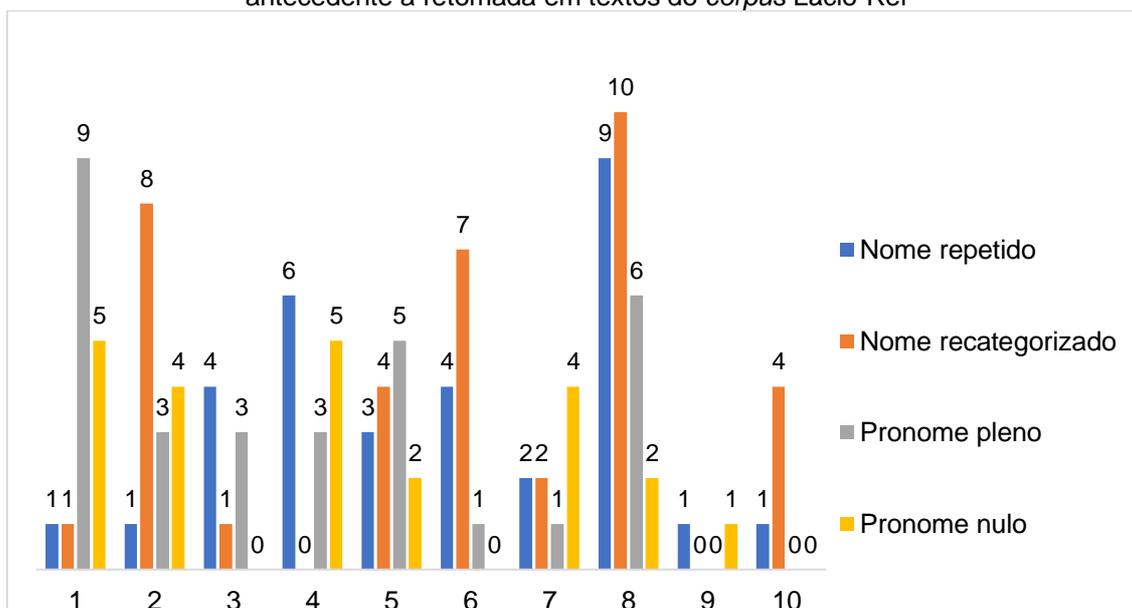
Podemos notar, em (127), a retomada sendo feita pelo uso do pronome nulo quando afastado em 07 (sete) palavras de seu referente *ficção televisiva*.

(127) [...] Se a ficção televisiva<sub>i</sub> pode tocar tantos ao mesmo tempo e  $\emptyset$ <sub>i</sub> discutir, dentro dos limites da ficção, questões delicadas, difíceis e cruciais do momento que vivemos com seriedade, cabe-lhe<sub>i</sub> tratar desses problemas não como documentário ou com o didatismo de uma lição. [...]

(JO-IF-JORUSP-cul-03jun02\_01)

No gráfico a seguir, são mostrados os resultados das quantidades de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos quando há, entre o antecedente e a forma correferente, o intervalo de 11 (onze) até 50 (cinquenta) palavras.

**Gráfico 05** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 11 a 50 palavras entre o antecedente a retomada em textos do *corpus* Lácio-Ref



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas no intervalo de 11 (onze) até 50 (cinquenta) palavras. Foram encontradas, na soma de ocorrência dos 10 (dez) textos, 124 retomadas nesta distância, sendo 32 por nomes repetidos, 38 por nomes recategorizados, 31 por pronomes plenos e 23 por pronomes nulos, implicando em 26% das retomadas efetuadas por nomes repetidos, 30,6% por nomes recategorizados, 25% por pronomes plenos e 18,4% por pronomes nulos.

Com estes resultados, inferimos que, nesta distância, os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos são usados quase de forma aleatória, pois não temos indícios a partir da frequência de ocorrência para o uso de uma forma preferível. A única coisa que nos chama atenção é a queda de ocorrências dos pronomes nulos, mas, mesmo assim, não encontramos indícios de significância estatística entre as formas.

Isto pode ser percebido ainda no rastreamento da referente *ficção televisiva*, mostrado no exemplo (128), que, ao ser distanciada da última entrada do referente em 14 (quatorze) palavras, levou o usuário da língua a fazer a retomada por meio do pronome pleno, o que, de acordo com os dados mostrados no gráfico 05, é comum em textos escritos.

(128) [...] Se a ficção televisiva<sub>i</sub> pode tocar tantos ao mesmo tempo e  $\emptyset$ <sub>i</sub> discutir, dentro dos limites da ficção, questões delicadas, difíceis e cruciais do momento que vivemos com

seriedade, cabe-lhe; tratar desses problemas não como documentário ou com o didatismo de uma lição. Se ela; atravessar a fronteira do ficcional para o factual, os aparelhos serão desligados ou sintonizados em outro canal. [...]

(JO-IF-JORUSP-cul-03jun02\_01)

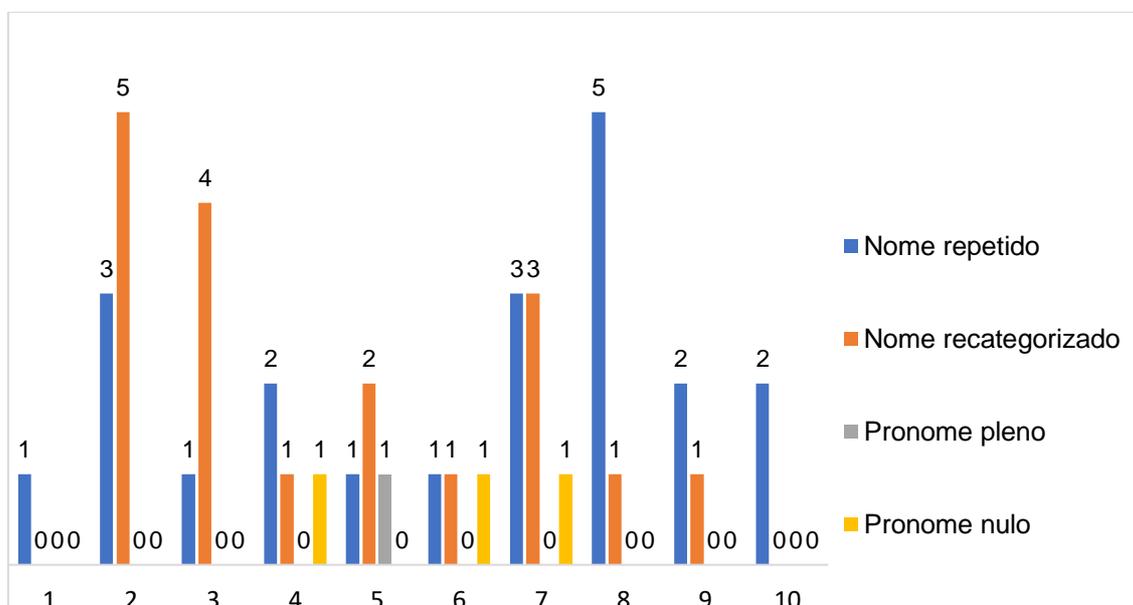
Podemos ver, no exemplo (129), o uso do nome recategorizado *pianista brasileiro* para retomar o referente *Martins* quando há entre a anáfora e seu antecedente 37 (trinta e sete) palavras.

(129) Martins; mostrou, de forma prática no disco gravado e em teoria, 15 anos antes, em um artigo escrito para Cahiers Debussy, como o francês foi influenciado por Moussorgsky. Nesse interesse pelo romantismo da segunda metade do século 19, o pianista brasileiro; que prefere tocar compositores que conhece em profundidade, desvenda o mundo infantil do compositor russo.

(JO-IF-JORUSP-cul-03jun02\_02)

No gráfico a seguir, mostramos as quantidades de ocorrências das formas correferentes no intervalo de 51 (cinquenta e uma) até 100 (cem) palavras.

**Gráfico 06** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 51 a 100 palavras entre o antecedente a retomada em textos do *corpus* Lácio-Ref



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas no intervalo de 51 (cinquenta e uma) até 100 (cem) palavras. Foram encontradas, na soma de ocorrência dos 10 (dez)

textos, 43 retomadas nesta distância, sendo 21 por nomes repetidos, 18 por nomes recategorizados, 01 por pronome pleno e 03 por pronomes nulos, implicando em 48,9% das retomadas efetuadas por nomes repetidos, 41,9% por nomes recategorizados, 2,2% por pronomes plenos e 7% por pronomes nulos.

Com estes resultados, podemos perceber, nitidamente, a queda no uso dos pronomes de maneira geral, sendo os nomes, tanto os repetidos quanto os recategorizados, as formas correferentes mais frequentes quando há distância maior em quantidade de palavras entre o antecedente e a retomada.

A fim de comprovar esta queda de frequência dos pronomes, nesta distância, optamos por analisar a porcentagem de ocorrência dos nomes repetidos com relação às demais formas e, assim, concluímos que, há 53,9% das retomadas sendo efetuados por nomes repetidos e 46,1% por nomes recategorizados (p-valor = 0,99) , 95,5% das retomadas sendo efetuadas por nomes repetidos e 4,5% por pronomes plenos (p-valor = 0,14) e, 87,5% das retomadas sendo efetuadas por nomes repetidos e 12,5% por pronomes nulos (p-valor = 0,64).

Em (130), o nome *Jó* introduz o referente que é recategorizado como *livro*, sendo retomado pelo nome recategorizado *um dos mais antigos textos do Velho Testamento* com reforço, após o aposto, pelo pronome *ele*, quando distante de seu referente em 26 (vinte e seis) palavras e, posteriormente, sendo retomado, na posição de sujeito, pelo nome recategorizado *livro*, quando há, entre a última retomada e novo centro anafórico, 54 (cinquenta e quatro) palavras.

(130) [...] *Jó*<sub>i</sub> não é um *livro*<sub>i</sub> sobre ciência. Um dos mais antigos textos do Velho Testamento<sub>i</sub>, cujo autor pode ter sido contemporâneo do patriarca Abraão, que viveu há quase 4 mil anos, *ele*<sub>i</sub> relata a história de um homem rico e piedoso que, fustigado pelo diabo, perdeu tudo o que possuía. Apesar das enormes dificuldades e da censura dos amigos que viam sua miséria como fruto de uma conduta má, *Jó* manteve uma inabalável confiança no Criador, que depois lhe restituiu os bens e a alegria. O *livro*<sub>i</sub>, portanto, procura fortalecer a fé do leitor, consolar os aflitos e celebrar o cuidado de Deus para com seus filhos. Nada a ver com dados científicos. [...]

(JO-IF-JORUSP-cul-03dez01\_02)

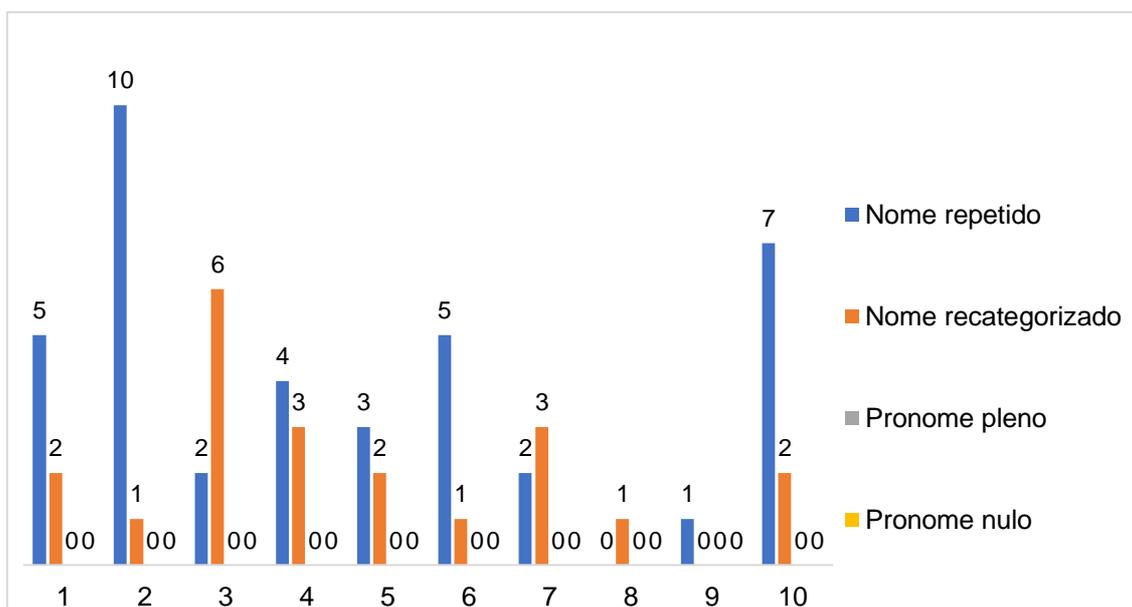
Já em (131), o referente introduzido pelo nome José Eduardo Martins é retomado pela repetição no nome Martins, quando há, entre a retomada e a introdução do referente, 81 (oitenta e uma) palavras.

(131) [...] Todos eles invadem uma capela belga do século 11 quando o pianista José Eduardo Martins; toca a música do russo Moussorgsky e do francês Claude Debussy. As obras dos dois compositores evocam motivos infantis e buscaram, ambas, inspiração em aquarelas. Para compor *La Boîte à Joux*, Debussy inspirou-se nas telas de André Hélé. Moussorgsky, depois de ver as aquarelas de seu amigo Victor Hartmann, que havia falecido dias antes, ficou tão impressionado que fez em quinze dias todas as peças musicais da famosa obra *Quadros de uma exposição*. É essa semelhança entre os dois músicos que Martins; procura mostrar no CD que foi lançado este ano pelo selo belga *De Rode Pomp*, um dos mais importantes da Europa. [...]

(JO-IF-JORUSP-cul-03jun02\_02)

No gráfico a seguir, são mostrados os resultados das quantidades de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos quando há, entre o antecedente e a forma correferente, mais de 100 (cem) palavras.

**Gráfico 07** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há mais de 100 palavras entre o antecedente a retomada em textos do *corpus* Lácio-Ref



No gráfico acima, a primeira coisa que recebe destaque é o fato de, neste contexto, não haver nenhuma retomada por pronome pleno ou nulo quando há mais de 100 (cem) palavras entre o antecedente e a retomada. Pelo fato de não haver dados a serem quantificados para os pronomes plenos e para os pronomes nulos neste contexto, descreveremos o que acontece entre os nomes repetidos e os nomes recategorizados.

Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas quando há 100 palavras entre o antecedente e a retomada, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 60 retomadas nesta distância, sendo 39 por nomes repetidos e 21 por nomes recategorizados, implicando em 65% das retomadas efetuadas por nomes repetidos e 35% por nomes recategorizados ( $p$ -valor = 0, 57).

Estes resultados estão em consonância com os dados do gráfico anterior, pois os nomes repetidos e os nomes recategorizados são mais frequentemente usados quando a distância, em quantidade de palavras, entre o antecedente e a retomada é longa, começando a ser percebido no intervalo de 51 a 100 palavras, sendo reconfirmado quando há mais de 100 palavras de distância entre a retomada e a última referência feita.

No exemplo (132), temos o nome *Adriana Ferrari*, recategorizado como *diretora do Sibi/USP* sendo retomado após o uso de 167 (cento e sessenta e sete) palavras, pela repetição do nome *diretora do Sibi*.

(132) [...] De acordo com Adriana Ferrari, diretora do Sibi/USP<sub>i</sub>, o orçamento anual aprovado para as bibliotecas da USP está em torno de R\$ 25 milhões, sendo que R\$ 22 milhões são gastos com a renovação de assinaturas de periódicos. Parece ser um número alto a princípio, mas é preciso levar em conta, por exemplo, que a assinatura de uma das principais revistas internacionais de psicologia custa aproximadamente US\$ 7.500. Todavia, grande parte dos usuários das bibliotecas da Universidade nem sabe da existência e do valor desse acervo de periódicos. A USP dispõe de aproximadamente 1,7 milhão de livros.

Para divulgar esse acervo e sua importância, a nova diretoria do Sibi<sub>i</sub> tem como uma de suas propostas de trabalho um projeto eficiente na área de marketing. [...]

(JO-IF-JORUSP-cul-02set02\_02)

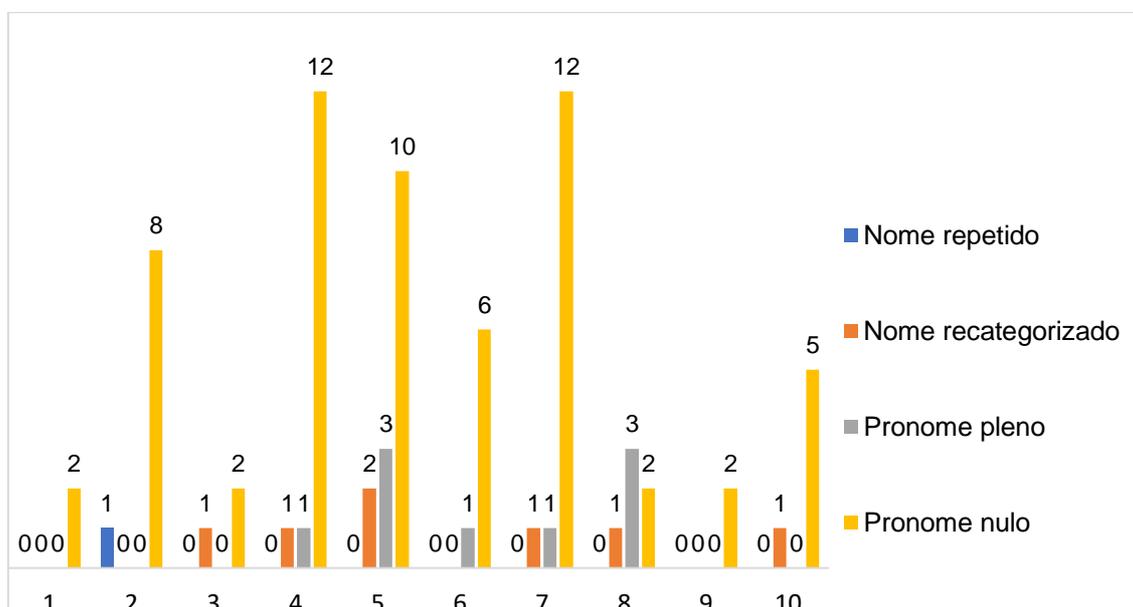
Uma vez finalizada a exposição dos resultados do tipo de retomada encontrado em intervalos de quantidade de palavras, conforme exposto nos quatro gráficos acima, damos prosseguimento aos resultados do uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em intervalos de caracteres sem espaço, respeitando os mesmos intervalos anteriores, ou seja, buscamos encontrar a frequência de ocorrência das expressões correferenciais em caracteres sem espaço.

Antes de começar a apresentação dos gráficos que mostram a distância entre o referente e seu correferente, é necessário advertirmos que, por acreditar que a distância entre os termos em quantidade de caracteres serve para

confirmar o que é encontrado quando levando em consideração a distância em quantidade de palavras, optamos por não apresentar exemplos para os gráficos que tratam do distanciamento do par referente/correferente em quantidade de caracteres, tampouco o p-valor entre as formas por não termos encontrado variância significativa entre nenhuma das formas anteriormente.

No gráfico que segue, estão expostos os valores em frequência de ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em quantidade de até 10 (dez) caracteres sem espaço.

**Gráfico 08** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes em até 10 caracteres sem espaço entre o antecedente a retomada em textos do *corpus* Lácio-Ref



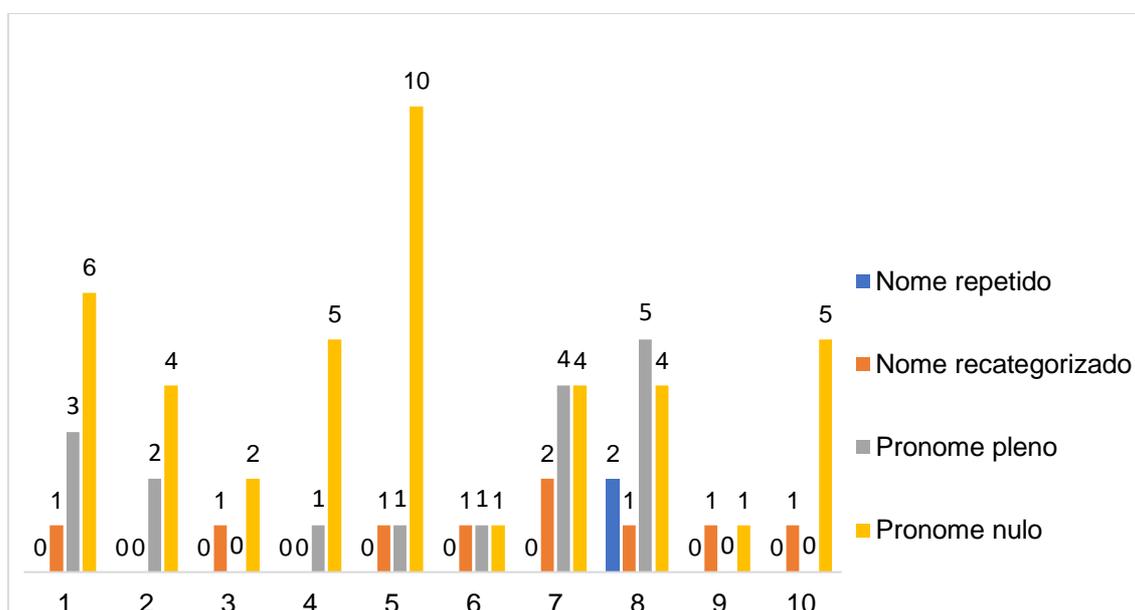
Ao condensar os resultados acima descritos e somar todas as ocorrências de retomadas encontradas no intervalo de 0 (zero) até 10 (dez) caracteres sem espaço nos 10 (dez) textos temos, 78 retomadas nesta distância, sendo 01 por nome repetido, 07 por nomes recategorizados, 09 por pronomes plenos e 61 por pronomes nulos, implicando em 1,2% das retomadas efetuadas por nomes repetidos, 9% por nomes recategorizados, 11,6% por pronomes plenos e 78,2% por pronomes nulos.

Percebemos que os pronomes nulos ocorrem com mais frequência nesta distância em quantidade de caracteres sem espaço e, quando comparadas as ocorrências do pronome nulo em relação às demais formas, temos a ocorrência de 87,1% de pronomes nulos e 12,9% de pronomes plenos, 89,7% de pronomes

nulos e 10,3% de nomes recategorizados e 98,4% de pronomes nulos e 1,6% de nomes repetidos. E, assim, os dados confirmam que, em contextos nos quais o antecedente e a retomada estão muito próximos, os pronomes nulos são as formas correferente mais frequentemente usadas.

O gráfico a seguir mostra a quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em quantidade de 11 a 50 caracteres sem espaço.

**Gráfico 09** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes de 11 a 50 caracteres sem espaço entre o antecedente e a retomada em textos do *corpus* Lácio-Ref



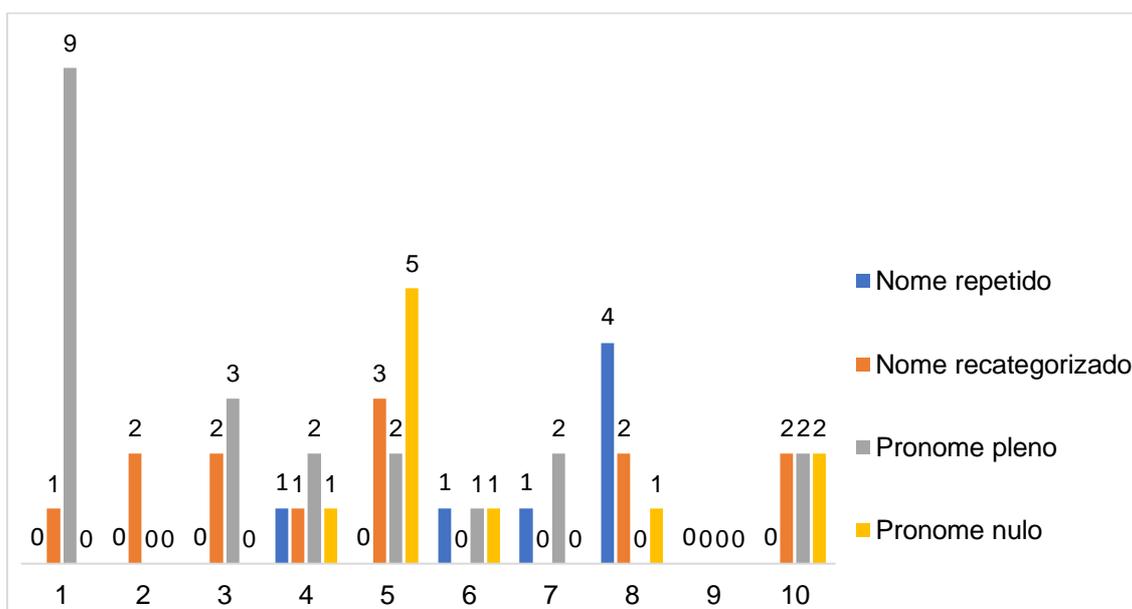
Ao condensar os resultados acima descritos e somar todas as ocorrências de retomadas encontradas no intervalo de 11 (onze) até 50 (cinquenta) caracteres sem espaço nos 10 (dez) textos temos, 70 retomadas nesta distância, sendo 02 por nomes repetidos, 09 por nomes recategorizados, 17 por pronomes plenos e 42 por pronomes nulos, implicando em 2,9% das retomadas efetuadas por nomes repetidos, 12,9% por nomes recategorizados, 24,2% por pronomes plenos e 60% por pronomes nulos.

Notamos, por meio destes resultados, que, sem espaçamento entre as palavras, o uso do pronome nulo ainda é a forma mais frequente, já que há 95,5% das retomadas efetuadas por pronomes nulos e 4,5% por nomes repetidos, 82,3% por pronomes nulos e 17,7% por nomes recategorizados e, por fim, 71,2% por pronomes nulos e 28,8% por pronomes plenos.

Estes resultados corroboram os achados anteriores quando considerada a quantidade de palavras, indicando que o pronome nulo é usado preferencialmente em distância curta, tanto em quantidade de palavras quanto em quantidade de caracteres entre a retomada e o antecedente. Mais adiante, veremos mais indícios destes dados mais adiante quando reportarmos os resultados da quantidade de ocorrência das expressões em caracteres com espaçamento entre as palavras.

No gráfico 10, mostramos a frequência de ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em quantidade de 51 a 100 caracteres sem espaço entre as palavras nos dados do *corpus* escrito Lácio-Ref.

**Gráfico 10** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes de 51 a 100 caracteres sem espaço entre o antecedente e a retomada em textos do *corpus* Lácio-Ref



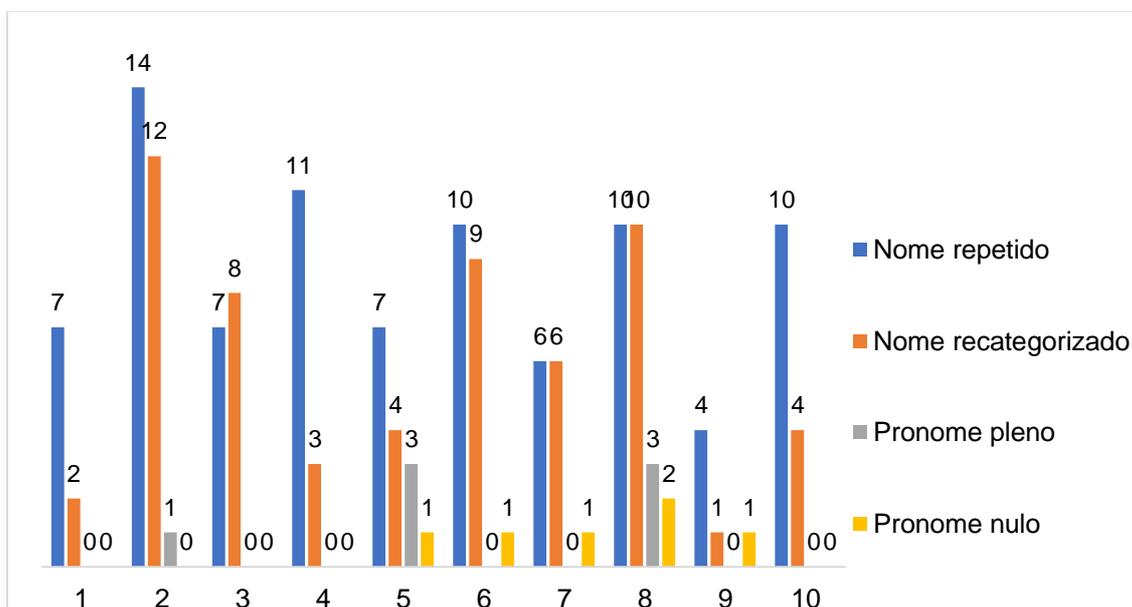
Ao condensar os resultados acima descritos e somar todas as ocorrências de retomadas encontradas no intervalo de 51 (cinquenta e um) até 100 (cem) caracteres sem espaço nos 10 (dez) textos temos, 51 retomadas nesta distância, sendo 07 por nomes repetidos, 13 por nomes recategorizados, 21 por pronomes plenos e 10 por pronomes nulos, implicando em 13,8% das retomadas efetuadas por nomes repetidos, 25,5% por nomes recategorizados, 41,2% por pronomes plenos e 19,5% por pronomes nulos.

Notamos que, nesta distância em caracteres sem espaçamento entre o referente e sua retomada, o pronome pleno é mais frequente, assim como o

nome recategorizado. Nesta distância, encontramos 75% das retomadas efetuadas por pronomes plenos e 25% por nomes repetidos, 61,8% por pronomes plenos e 38,2% por nomes recategorizados e, por fim, 67,8% por pronomes plenos e 32,2% por pronomes nulos.

O próximo gráfico traz os resultados do último intervalo proposto para esta subcategoria de análise.

**Gráfico 11** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes em mais de 100 caracteres sem espaço entre o antecedente e a retomada em textos do *corpus* Lácio-Ref



Ao condensar os resultados acima descritos e somar todas as ocorrências de retomadas encontradas em mais de 100 (cem) caracteres sem espaço nos 10 (dez) textos temos 158 retomadas, sendo 86 por nomes repetidos, 59 por nomes recategorizados, 07 por pronomes plenos e 06 por pronomes nulos, implicando em 54,4% das retomadas efetuadas por nomes repetidos, 37,4% por nomes recategorizados, 4,4% por pronomes plenos e 3,8% por pronomes nulos.

Percebemos que, nesta distância em caracteres sem espaçamento entre as palavras, o nome repetido é mais frequente, mas não muito mais do que o nome recategorizado, apontando para uma distância em que o uso do nome funciona melhor que o uso dos pronomes, uma vez que a frequência destes cai com relação aos intervalos reportados anteriormente. Nesta distância, encontramos 59,3% das retomadas efetuadas por nomes repetidos e 40,7% por

nomes recategorizados, 92,5% por nomes repetidos e 7,5% por pronomes plenos e, por fim, 93,5% por nomes repetidos e 6,5% por pronomes nulos.

A fim de confirmar nossos resultados anteriores, reportaremos a seguir os resultados da frequência de ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em textos escritos do *corpus* Lácio-Ref com relação à quantidade de caracteres com espaço entre o referente e a retomada. Porém, como os dados são bastante próximos dos encontrados nos mesmos intervalos de caracteres sem espaço entre o par referente/expressão correferente, conforme mostrado nos gráficos de 08 a 11, optamos por apenas colocar os resultados gerais dos intervalos de caracteres com espaço, sem reportar, novamente, o p-valor comparativo das formas, por não termos encontrado efeitos de significância estatística entre as formas quando comparadas entre si, levando em consideração a variabilidade dos textos.

Estes resultados tão próximos ocorrem pelo fato de, por termos contabilizado intervalos e não quantidade exata para cada retomada, o limite entre uma forma e outra chega a ser tão fluido que, uma mesma forma pode estar a 09 (nove) caracteres sem espaço e a 11 (onze) caracteres com espaço de seu antecedente, sendo, portanto, uma variação muito curta nos resultados encontrados.

Para deixar estes limites ainda mais claros, optamos por mostrar, no exemplo (133), extraído<sup>191</sup> do *corpus* analisado, os limites em quantidade de (1) caracteres com espaço, (2) caracteres sem espaço e (3) palavras quando quantificamos as ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito ao longo do texto.

(133) [...] A exposição permanecia intocada, tal qual na época em que fora idealizada, no início do século, quando o museu ainda era a seção de História Natural do Museu Paulista. “Parecia um enorme catálogo e para os padrões daquela época quanto mais animais fossem expostos, melhor. Havia tantos bichos<sub>i</sub> que ficava difícil saber onde começava um e terminava o outro”, conta a professora Miriam David Mar**(1)**ques, diretora da di**(2)**visão de difusão cultural. Com a casa nova, o conceito da exposição também será diferente. “Será um ponto de virada na história do museu. Não queremos mostrar uma porção de bichos mortos, empalhados e sujos, e sim qual é a nossa função e o nosso trabalho.” Os exemplares<sub>i</sub> foram limpos e agora Ø<sub>i</sub> serão

---

<sup>191</sup> Neste exemplo foram feitas as modificações necessárias para mostrar o exato momento em aparece a centésima palavra, o centésimo caractere sem espaço e o centésimo caractere com espaço, criando, portanto, modificações ao texto original.

apresentados não mais soltos em vitrines. Eles<sub>i</sub> estarão dentro de cenários que reproduzem o seu hábitat natural. O mote da mostra **(3)** — a pesquisa em zoologia — saiu das discussões feitas durante seis seminários que reuniram todo o corpo científico do museu. [...]

(JO-IF-JORUSP-cul-02set02\_04 - *corpus* Lácio-Ref. Modificado)

No fragmento acima, dentre outros referentes, estávamos analisando todas as retomadas do referente *bichos* que aparece no começo do texto e depois passa a ser retomado por nomes recategorizados, pronomes plenos e/ou pronomes nulos em posição de sujeito ou objeto, majoritariamente.

Por questões de delimitação de nosso objeto de estudo, atentamo-nos para as retomadas deste referente na posição de sujeito mapeando a distância entre a última referência, mesmo que esta não seja na posição de sujeito, e sua retomada.

Assim, no fragmento acima, na linha 04 temos o referente *bichos* reaparecendo no texto, em posição de objeto e sendo retomado posteriormente pelo nome recategorizado *exemplares* em posição de sujeito, havendo entre o referente e a retomada 65 palavras, 311 caracteres sem espaço e 376 caracteres com espaço. Há, por sua vez, uma segunda retomada sendo feita pelo pronome nulo, tendo como última entrada do referente o nome recategorizado *exemplares*, sendo distanciado do antecedente por 04 palavras, 17 caracteres sem espaço e 22 caracteres com espaço. E, na terceira retomada deste exemplo, temos o pronome pleno retomando a última entrada do referente por pronome nulo, estando distante deste por 07 palavras, 41 caracteres sem espaço e 48 caracteres com espaço.

Constatamos, portanto, que, no primeiro caso, há uma diferença de 60 caracteres ao contabilizar os caracteres sem e com espaço, enquanto no segundo caso, esta diferença cai para 05 caracteres, e, no terceiro, para 48 caracteres, o que nos mostra o motivo de a variação dos referentes com relação à quantidade de caracteres ser muito fluída, pois o mesmo referente pode estar a tão curta distância da retomada em quantidade de caracteres que, a mesma retomada ora está em um intervalo ora em outro, de acordo com os intervalos que estabelecemos.

Acreditamos que, devido a este limiar fluído na quantidade de caracteres, verificamos pouquíssimas mudanças nos resultados de ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos. Essa observação parece confirmar dados já encontrados e, por isso, optamos por, ao

invés de mostrar mais 04 (quatro) gráficos quase idênticos aos anteriores, reportar os resultados gerais com os respectivos comentários.

Assim, temos para a quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos funcionando como expressão correferente na posição de sujeito em dados de *corpus* escrito a uma distância de 0 (zero) até 10 (dez) caracteres com espaço, os mesmos resultados para este intervalo em caracteres sem espaço.

Desta forma, para o intervalo de 0 (zero) a 10 (dez) caracteres com espaço, houve 78 retomadas, sendo 01 por nome repetido, 07 por nomes recategorizados, 09 por pronomes plenos e 61 por pronomes nulos, implicando em 1,2% por nomes repetidos, 9% por nomes recategorizados, 11,6% por pronomes plenos e 78,2% por pronomes nulos.

Percebemos que os pronomes nulos ocorrem com mais frequência nesta distância em quantidade de caracteres sem espaço e, quando comparadas as ocorrências do pronome nulo em relação às demais formas, temos a ocorrência de 87,1% de pronomes nulos e 12,9% de pronomes plenos, 89,7% de pronomes nulos e 10,3% de nomes recategorizados e 98,4% de pronomes nulos e 1,6% de nomes repetidos.

Notamos que, no intervalo acima, temos os resultados iguais ao mesmo intervalo quando contabilizada a quantidade de caracteres sem espaço, dando-nos a mesma interpretação. Já nos intervalos seguintes, houve leve variação de valores, como reportamos a seguir.

No intervalo de 11 (onze) a 50 (cinquenta) caracteres com espaço, houve 66 retomadas, sendo 02 por nomes repetidos, 09 por nomes recategorizados, 16 por pronomes plenos e 39 por pronomes nulos, implicando em 3,4% das retomadas efetuadas por nomes repetidos, 13,6% por nomes recategorizados, 19,7% por pronomes plenos e 66,3% por pronomes nulos.

Percebemos que o uso dos pronomes nulos, nesta distância, ainda é mais frequente do que o uso das outras formas, já que encontramos a ocorrência de 95,1% de pronomes nulos e 4,9% de nomes repetidos, 81,3% de pronomes nulos e 18,7% de nomes recategorizados e, por fim, 71% de pronomes nulos e 29% de pronomes plenos.

Já no intervalo de 51 (cinquenta e um) a 100 (cem) caracteres com espaço, houve 50 retomadas, sendo 05 por nomes repetidos, 10 por nomes recategorizados, 22 por pronomes plenos e 13 por pronomes nulos, implicando

em 10% das retomadas efetuadas por nomes repetidos, 20% por nomes recategorizados, 44% por pronomes plenos e 26% por pronomes nulos.

Nesta distância, o uso do pronome pleno é mais frequente, tendo sido encontrada a ocorrência de 81,5% de pronomes plenos e 18,5% de nomes repetidos, 68,8% de pronomes plenos e 31,2% de nomes recategorizados e, por fim, 62,9% de pronomes plenos e 37,1% de pronomes nulos.

Quando há mais de 100 (cem) caracteres com espaço entre o antecedente e a retomada, encontramos 164 retomadas, sendo 88 por nomes repetidos, 63 por nomes recategorizados, 7 por pronomes plenos e 6 por pronomes nulos, implicando em 53,7% das retomadas efetuadas por nomes repetidos, 38,4% por nomes recategorizados, 4,3% por pronomes plenos e 3,6% por pronomes nulos.

Desta forma, temos indícios de que, quando há distância maior em quantidade de caracteres com espaço, assim como aconteceu quando a distância é grande em quantidade de caracteres sem espaço, a forma preferencialmente usada como expressão correferencial é o nome repetido, já que encontramos a ocorrência de 58,3% de nomes repetidos e 41,7% de nomes recategorizados, 92,7% de nomes repetidos e 7,2% de pronomes plenos e, por fim, 93,7% de nomes repetidos e 6,3% de pronomes nulos.

Para finalizar estas categorias de análise, gostaríamos de reiterar que encontramos indícios de que (i) quanto maior a distância em quantidade de caracteres com e sem espaço, o nome repetido é mais usado, (ii) não encontramos a quantidade exata de caracteres sem ou com espaço em que o uso do nome repetido ou do nome recategorizado passa a ser mais frequente do que os pronomes nulos, mas somente indícios de que, quanto maior a distância em quantidade de caracteres sem e com espaço, assim como acontece quando analisada a quantidade de palavras, menos provável é o uso do pronome nulo.

Feitas as reiterações e em busca de resultados de significância estatística, procuramos verificar estatisticamente uma possível influência da distância em relação à forma correferente utilizada. Utilizando o *software* PAST, uma vez que esse software leva em consideração a variabilidade de nossos textos, inserimos os valores de frequência de uso para cada intervalo de distância por nós mapeado. Calculamos o Qui-quadrado e encontramos que a distância, em quantidade de palavras, em quantidade de caracteres sem e com

espaço, pode ter influência direta na forma correferente usada no texto (considerando a distância *versus* a forma correferente, o p-valor foi de 0,0001).

Uma vez que nossas três primeiras categorias de análise foram minuciosamente reportadas, partiremos para os resultados de outras questões que podem nos ajudar a entender melhor como acontece a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos funcionando como expressões correferenciais em posição de sujeito em textos escritos em Português Brasileiro.

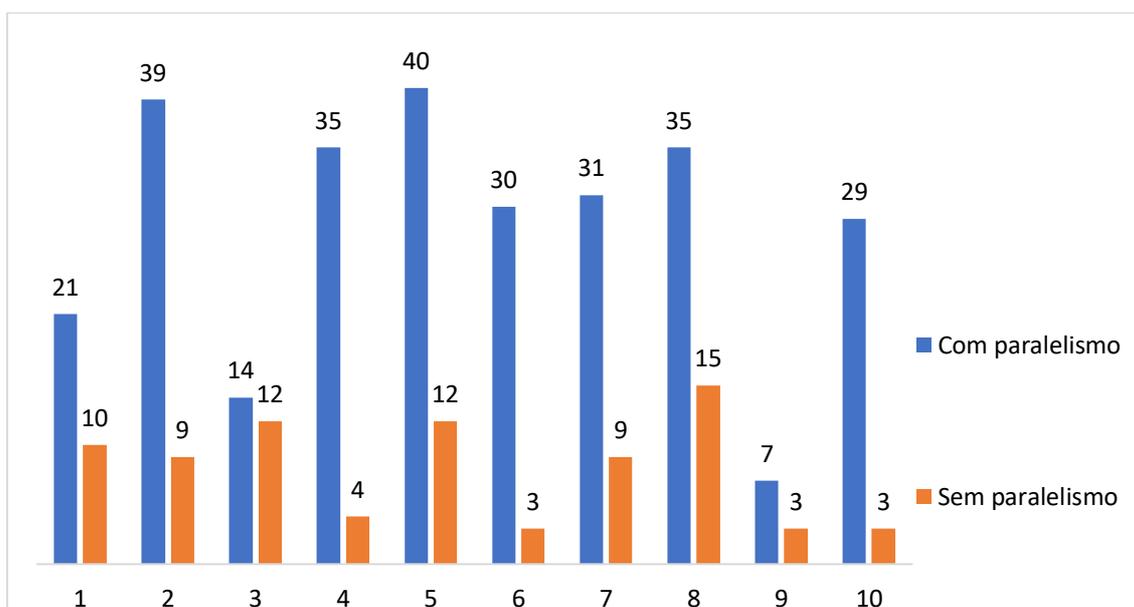
Seguindo a sequência de nosso formulário, conforme detalhadamente exposto na seção anterior, apresentamos abaixo a quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos que, em posição de sujeito, retomaram um antecedente que estava em posição de sujeito ou não para, assim, termos indícios sobre o efeito do Paralelismo Sintático (CHAMBERS; SMITH, 1998) durante a resolução da correferência anafórica em textos escritos em PB<sup>192</sup>.

Como o efeito de paralelismo sintático trata sobre o antecedente e a retomada ocorrerem na mesma posição sintática, observamos se o antecedente era retomado na posição de sujeito ou não, considerando haver efeito de paralelismo quando o antecedente é retomado por uma expressão correferencial estando ambos em posição de sujeito e, consideramos sem paralelismo quando o antecedente está em outra posição que não seja a de sujeito da sentença.

---

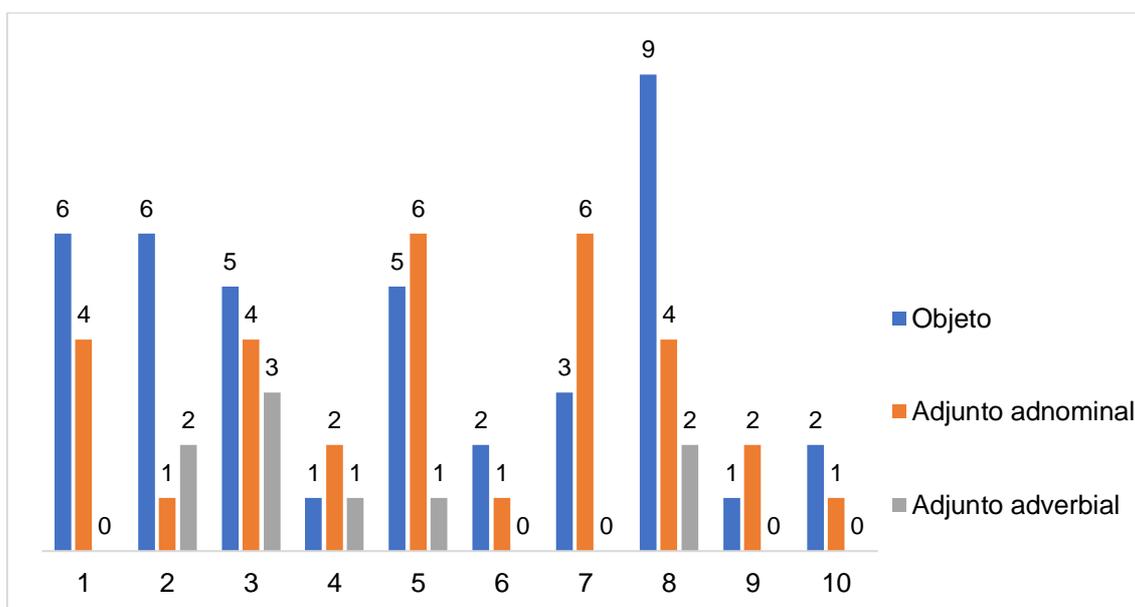
<sup>192</sup> Por não tratar diretamente da distribuição das formas correntes na posição de sujeito de textos escritos em PB, não apresentaremos exemplos para os resultados dos gráficos que tratam do paralelismo sintático, nem das demais posições sintáticas que podem introduzir referentes.

**Gráfico 12** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes em posição ou não de sujeito nos textos do *corpus* Lácio-Ref



Somando todas as ocorrências contabilizadas a fim de averiguar o efeito de paralelismo sintático, verificamos 361 retomadas em posição de sujeito, sendo 281 retomando um antecedente também em posição de sujeito e 80 retomando antecedente em outra posição sintática, ou seja, 77,8% de ocorrências sujeito/sujeito e 22,2% outras posições sintáticas/sujeito, tendo essa diferença  $p\text{-valor} = 0,38$ . Levando em consideração a frequência, temos indícios de que há grande tendência para o termo antecedente em posição de sujeito ser retomado por uma expressão correferencial também em posição de sujeito, corroborando possível efeito de Paralelismo Sintático em textos escritos em PB.

**Gráfico 13** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes em posições sintáticas que não de sujeito nos textos do *corpus* Lácio-Ref



O gráfico acima mostra as entradas de antecedentes que não estavam em posição de sujeito, tendo sido recorrente o fato de os antecedentes surgirem em posição de objeto, de adjunto adnominal ou adjunto adverbial. Além destas entradas, ocorreram expressões do tipo *conforme o autor*, *segundo ele* e outras do tipo, mas optamos por enquadrá-las na subcategoria dos adjuntos adverbiais, pelo fato de ambas poderem ser movidas dentro do texto.

Somando todas as ocorrências contabilizadas, a fim de averiguar a frequência das posições sintáticas em que os referentes tendem a ocorrer quando não há possível efeito de paralelismo sintático, observamos 80 retomadas em posição de sujeito, sendo 40 retomadas de um antecedente em posição de objeto, 31 em posição de adjunto adnominal e 09 em posição de adjunto adverbial, implicando em 50% de ocorrências objeto/sujeito ( $p$ -valor = 0,99), 38,7% adjunto adnominal/sujeito ( $p$ -valor = 0,31) e 11,3% adjunto adverbial/sujeito ( $p$ -valor = 0,10). Esses valores de  $p$  implicam que as diferenças de ocorrência nas posições objeto/sujeito, adjunto adnominal/sujeito e adjunto adverbial/sujeito não são significativas.

Por tratarmos de frequência de uso, acreditamos ter indícios de que o efeito de paralelismo sintático pode ser percebido quando a língua está em uso. No entanto podemos também notar que, mesmo quando o antecedente aparece em outras posições sintáticas, há uma tendência das posições que (re)introduzem referentes ao longo do texto, uma vez que a frequência de um antecedente em

posição de objeto a ser retomado por uma expressão correferente em posição de sujeito parece ser maior do que um adjunto adverbial contendo um referente a ser retomado em posição de sujeito.

Creemos, portanto, ter dado conta de mais uma de nossas categorias de análise do *corpus* escrito, avançando na exposição de nossos resultados para algo também explorado em pesquisas psicolinguísticas: o tipo de oração na qual a retomada é feita.

Em nossa revisão da literatura, encontramos pesquisas que trabalham com três tipos de orações onde ocorre a retomada: (i) orações coordenadas justapostas, (ii) orações coordenadas ligadas por conjunção e (iii) orações subordinadas.

Nesta pesquisa, em contrapartida, buscamos evidências de acordo com a disposição textual, podendo as retomadas serem feitas como sujeitos em: (i) orações subordinadas, (ii) orações coordenadas justapostas – aquelas que estão ligadas por vírgulas e/ou conjunções no mesmo período, (iii) orações coordenadas em períodos distintos – aquelas que aparecem dispostas uma ao lado da outra, mas separadas por ponto final, ou seja, em períodos diferentes e, (iv) a primeira retomada em parágrafo subsequente.

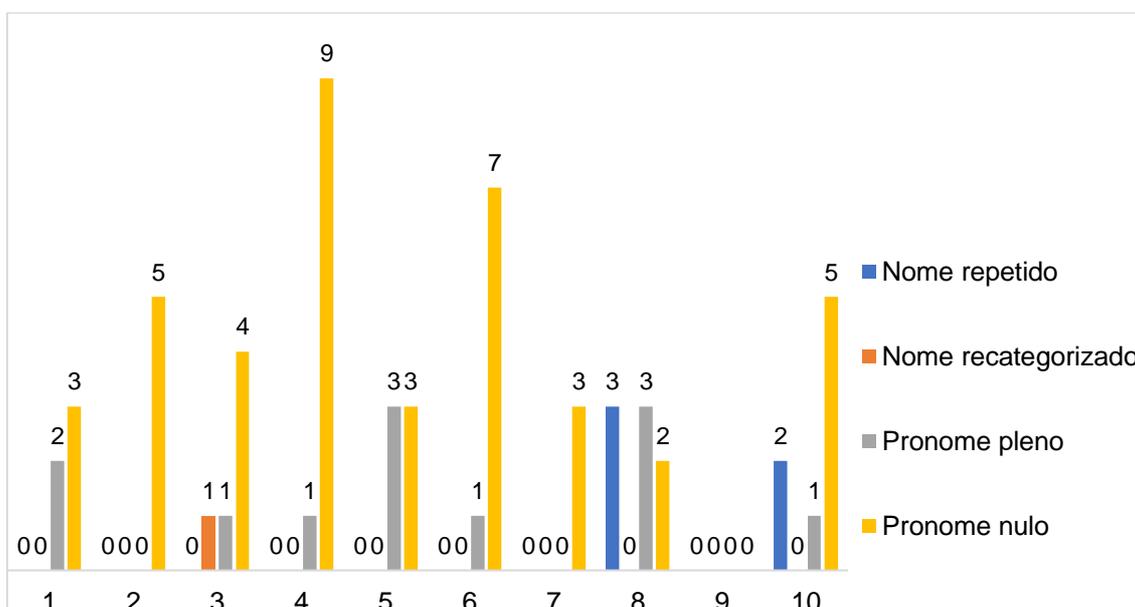
Como, nesta pesquisa, propusemo-nos a observar como nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos estão distribuídos ao longo do texto, ou seja, funcionando como uma cadeia de correferenciação, percebemos a necessidade de analisar como ocorre a retomada do referente entre parágrafos, uma vez que os três tipos citados acima e amplamente pesquisados em psicolinguística não dão conta de como acontece a retomada quando são mudados os parágrafos

A fim de expor os resultados destas categorias de análise, começaremos mostrando os resultados de como acontece, em textos escritos, o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações subordinadas<sup>193</sup>.

---

<sup>193</sup> Neste trabalho, não fizemos distinção entre o tipo de oração subordinada que foi contabilizada e, assim, não diferenciamos orações subordinadas substantiva de orações subordinadas adverbiais, nem orações subordinadas desenvolvidas de orações subordinadas reduzidas, pois não faz parte de nosso objeto de estudo. Tivemos cuidado de não reportar as orações subordinadas adjetiva, conhecidas como encaixadas, pelo fato de o pronome relativo ser visto como uma forma realizada do sujeito, não cabendo, portanto, a contabilização como pronome nulo.

**Gráfico 14** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações subordinadas nos textos do *corpus* Lácio-Ref



Somando as quantidades totais de retomadas em orações subordinadas dos 10 (dez) textos, houve 59 retomadas, sendo 05 por nomes repetidos, 01 por nome recategorizado, 12 por pronomes plenos e 41 por pronomes nulos, implicando em 8,5% das retomadas efetuadas por nomes repetidos, 1,6% por nomes recategorizados, 20,4% por pronomes plenos e 69,5% por pronomes nulos.

Ao compararmos as formas entre si, observamos que o pronome nulo é a forma mais frequente e, assim, comparamos a frequência deste com as demais formas, a fim de encontrar possíveis efeitos de penalidade. Foram contabilizadas 89,1% retomadas efetuadas por pronomes nulos e 10,9% por nomes repetidos, diferença não significativa ( $p$ -valor = 0,66), 97,6% por pronomes nulos e 2,4% por nomes recategorizados, também não significativa ( $p$ -valor = 0,60) e, por fim, 77,3% por pronomes nulos e 22,7% por pronomes plenos, da mesma forma não significativa ( $p$ -valor = 0,77).

Percebemos, portanto, que, em textos escritos, é mais comum surgir o pronome nulo na posição de sujeito da oração subordinada, retomando um referente, como em (134).

(134) [...] Uma forma diferente de ver o Direito, a linguagem e a cultura será exposta pela USP na próxima semana. Nos dias 9, 10 e 11 de setembro, a Faculdade de Educação reunirá

especialistas de diferentes universidades para  $\emptyset_i$  tratar desses temas durante o 4o. Seminário Internacional Filosofia, Educação e Arte. [...] <sup>194</sup>

(JO-IF-JORUSP-cul-02set02\_03)

Tais resultados vão na mesma direção de nossos achados quanto à quantidade de palavras e à quantidade de caracteres sem e com espaço, pois notamos que os pronomes nulos, em orações subordinadas, são mais frequentes do que os nomes.

Para reportar as orações coordenadas, foi necessário diferenciarmos o que classificamos como dois tipos diferentes, uma vez que, na literatura psicolinguística recebe o nome de oração coordenada justaposta aquela que está disposta ao lado da outra, estando ou não ligada por conjunção. Entretanto, ao longo da cadeia referencial ao longo do texto, percebemos duas possíveis orações coordenadas: (i) fazendo parte do mesmo período e podendo estar disposta uma ao lado da outra ligada por vírgulas e/ou conjunções e (ii) estando em períodos distintos, mas no mesmo parágrafo, ou seja, separadas por ponto final de um período ao lado do outro que se inicia no mesmo parágrafo. Por conta desta diferenciação, tivemos de criar duas subcategorias para as orações coordenadas, chamando de (i) orações coordenadas justapostas aquelas que estão no mesmo período, ligadas por vírgulas e/ou conjunções e (ii) orações coordenadas em períodos distintos, aquelas que integram o mesmo parágrafo, mas são separadas por ponto final.

Entretanto, reiteramos que, apesar de estarmos reportando a quantidade de ocorrências de nomes recategorizados, não estamos procurando explicações assertivas sobre esta forma, pois acreditamos que as motivações de uso de nomes repetidos ou de nomes recategorizados está mais para uma questão semântica do que sintática.

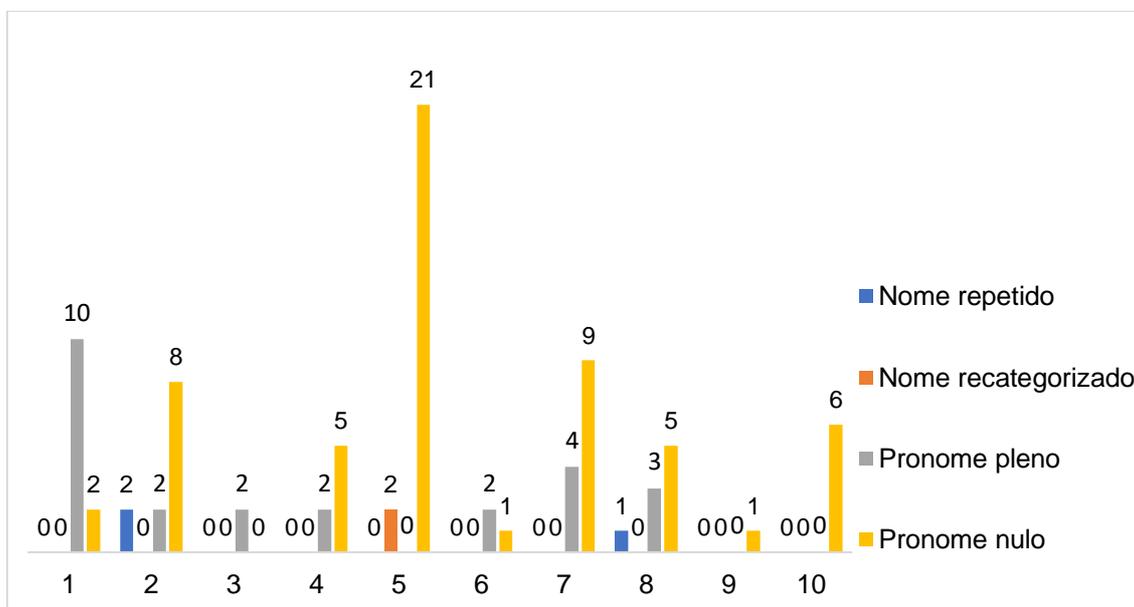
Como o objeto de nossa pesquisa é explicar a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB, continuaremos descrevendo as ocorrências destas formas, buscando explicações sintáticas para o aparecimento de cada uma delas.

---

<sup>194</sup> No exemplo (134), mostramos uma oração subordinada adverbial final reduzida de infinitivo que, na forma desenvolvida pode ser “a Faculdade de Educação reunirá especialistas de diferentes universidades para que os especialistas/eles tratem desses temas.

Reportamos, a seguir, o gráfico com os resultados das ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas justapostas.

**Gráfico 15** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas justapostas nos textos do *corpus* Lácio-Ref



Somando as quantidades totais de retomadas em orações coordenadas justapostas dos 10 (dez) textos, houve 88 retomadas, sendo 03 por nomes repetidos, 02 por nomes recategorizados, 25 por pronomes plenos e 58 por pronomes nulos, implicando em 3,4% das retomadas efetuadas por nomes repetidos, 2,3% por nomes recategorizados, 28,4% por pronomes plenos e 65,9% por pronomes nulos.

Ao compararmos as formas entre si, observamos que o pronome nulo é ainda a forma mais frequentemente usada embora as diferenças de uso do pronome nulo e outras formas não sejam estatisticamente significativas, pois encontramos 95% das retomadas efetuadas por pronomes nulos e 5% por nomes repetidos ( $p$ -valor = 0,99), 96,6% por pronomes nulos e 3,4% por nomes recategorizados ( $p$ -valor = 0,86) e, por fim, 69,9% por pronomes nulos e 30,1% por pronomes plenos ( $p$ -valor = 0,99).

Em (135), temos há a retomada por pronome nulo do referente trabalhos em orações coordenadas ligadas por conjunção *ou*, como nomeamos, em orações coordenadas justaposta, o que, levando em consideração os resultados

do gráfico 15, parece ser a retomada mais natural neste contexto em textos escritos.

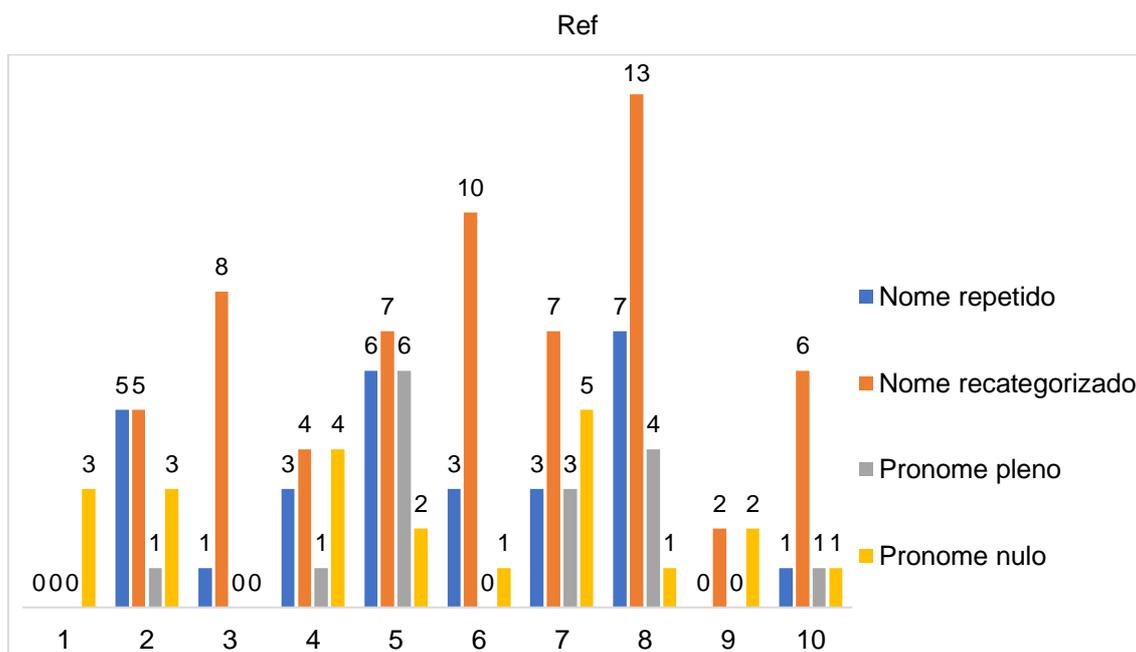
(135) [...] Os trabalhos têm propostas diferentes e  $\emptyset_i$  mostram bem as trilhas infinitas da arte contemporânea que dá vazão à criatividade, conceito e técnica. [...]

(JO-IF-JORUSP-cul-03dez01\_03)

Encontramos aqui, portanto, o uso dos nomes repetidos sendo mais frequente do o uso dos nomes recategorizados, o que pode ser interpretado como mais um indício da influência da distância curta em quantidade de palavras e de caracteres com ou sem espaço, já que as orações coordenadas justapostas frequentemente apresentam os referentes próximos de suas retomadas.

Vejamos os resultados do gráfico 16.

Gráfico 16 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas em períodos distintos nos textos do *corpus* Lácio-Ref



Somando as quantidades totais de retomadas em orações coordenadas em períodos distintos dos 10 (dez) textos, houve 129 retomadas, sendo 29 por nomes repetidos, 62 por nomes recategorizados, 16 por pronomes plenos e 22 por pronomes nulos, implicando em 22,4% das retomadas efetuadas por nomes repetidos, 48% por nomes recategorizados, 12,4% por pronomes plenos e 17,2% por pronomes nulos.

Ao compararmos as formas entre si, encontramos, neste contexto, 68,1% das retomadas efetuadas por nomes recategorizados e 31,9% por nomes repetidos ( $p$ -valor = 1), 79,5% por nomes recategorizados e 20,5% por pronomes plenos ( $p$ -valor = 0,72) e, por fim, 73,8% por nomes recategorizados e 26,2% por pronomes nulos ( $p$ -valor = 1,77). Observamos, assim, que, apesar das diferenças de uso do nome categorizado e outras formas não serem estatisticamente significativas, o uso do nome recategorizado é, neste contexto, mais frequente, como em (136), sendo seguido pelo uso do nome repetido.

(136) [...] Com uma apresentação didática interativa sobre orquestra de cordas, o diretor e regente titular<sub>i</sub> da Orquestra Sinfônica da USP, Carlos Moreno<sub>i</sub>, realiza seu sonho de levar música erudita para a praça, para as pessoas que não têm como ir ao teatro. “Com essa apresentação queremos formar uma nova plateia e apresentar um trabalho artístico de alto nível para crianças, introduzindo-as num mundo um pouco diferenciado daquele a que estão acostumadas”, explica o maestro<sub>i</sub>. [...]

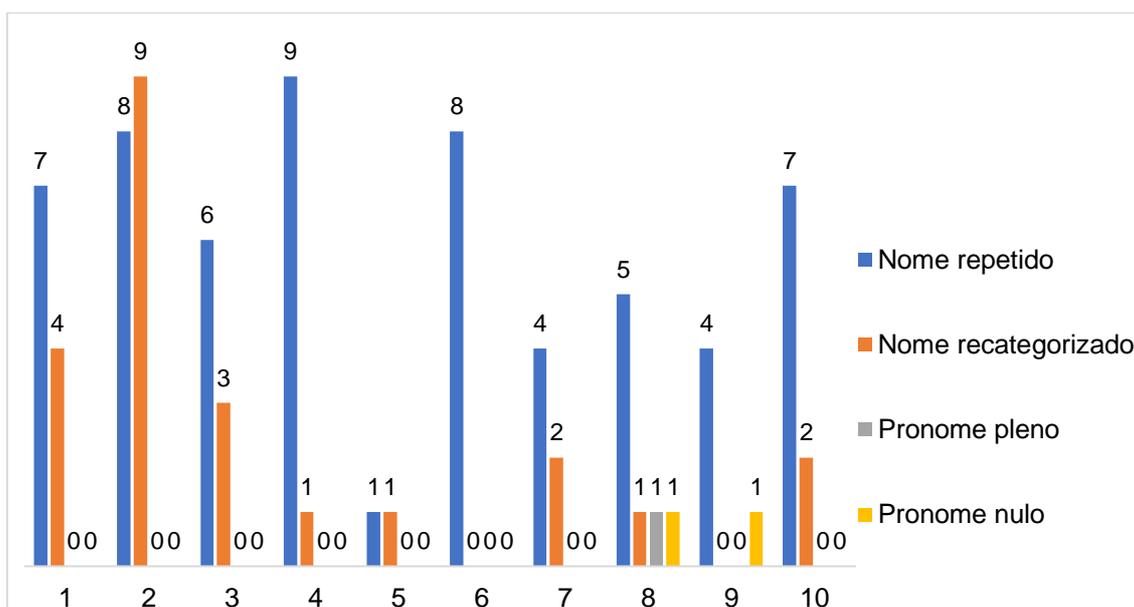
(JO-IF-JORUSP-cul-02set02\_01)

Reportamos, abaixo, os resultados de ocorrência dos nomes repetidos, dos nomes recategorizados, dos pronomes plenos e dos pronomes nulos quando correferentes na primeira retomada em parágrafo subsequente<sup>195</sup>.

---

<sup>195</sup> Reforçamos que, neste trabalho, estamos analisando a disposição textual das retomadas executadas por nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em dados de *corpus* escrito. Ao fazer este tipo de análise, encontramos uma cadeia referencial e diversas retomadas em posições sintáticas distintas e em orações e parágrafos distintos. Conforme explicamos anteriormente nesse capítulo, uma de nossas subcategorias de análise trata da primeira retomada quando em parágrafos subsequente, isto é, analisamos a última vez em que o referente rastreado entrou em posição de sujeito e a primeira vez em que ele foi retomado no parágrafo seguinte para, assim, analisar se os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos aparecem com a mesma frequência neste contexto. Além disso, aproveitamos para reiterar que este contexto nunca foi analisado no processamento correferencial em pesquisas psicolinguísticas.

Gráfico 17 – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em parágrafos distintos nos textos do *corpus* Lácio-Ref



Somando as quantidades totais de retomadas que funcionam como a primeira retomada em parágrafo subsequente dos 10 (dez) textos, houve 85 retomadas, sendo 59 por nomes repetidos, 23 por nomes recategorizados, 01 por pronome pleno e 02 por pronomes nulos, implicando em 69,5% das retomadas efetuadas por nomes repetidos, 27% por nomes recategorizados, 1,2% por pronomes plenos e 2,3% por pronomes nulos.

Neste contexto, temos resultados que mostram o uso dos nomes repetidos gerando efeito de penalidade sobre os pronomes nulos, conforme aconteceu quando há mais de 100 (cem) palavras entre o antecedente e a retomada. Há, aqui, indícios do uso do nome repetido gerando efeito de penalidade também sobre o pronome pleno.

Ao compararmos a frequência de uso das expressões correferenciais entre si, encontramos 71,9% das retomadas efetuadas por nomes repetidos e 28,1% por nomes recategorizados, diferença não significativa ( $p$ -valor = 0,85), 98,3% por nomes repetidos e 1,7%, diferença significativa por pronomes plenos ( $p$ -valor = 0,03) e, por fim, 96,7% por nomes repetidos e 3,3% por pronomes nulos, também significativa ( $p$ -valor = 0,05).

Em (137), temos a introdução do referente por meio do uso do nome *Adriana Ferrari* que é recategorizado para *diretora do Sibi/USP*, não ocorrendo outra retomada no mesmo parágrafo. Quando o referente é retomado, no

primeiro centro anafórico do parágrafo seguinte, o usuário optou pela repetição do nome diretora do Sibi/USP.

(137) [...] De acordo com Adriana Ferrari, diretora do Sibi/USP<sup>i</sup>, o orçamento anual aprovado para as bibliotecas da USP está em torno de R\$ 25 milhões, sendo que R\$ 22 milhões são gastos com a renovação de assinaturas de periódicos. Parece ser um número alto a princípio, mas é preciso levar em conta, por exemplo, que a assinatura de uma das principais revistas internacionais de psicologia custa aproximadamente US\$ 7.500. Todavia, grande parte dos usuários das bibliotecas da Universidade nem sabe da existência e do valor desse acervo de periódicos. A USP dispõe de aproximadamente 1,7 milhão de livros.

Para divulgar esse acervo e sua importância, a nova diretoria do Sibi tem como uma de suas propostas de trabalho um projeto eficiente na área de marketing. [...]

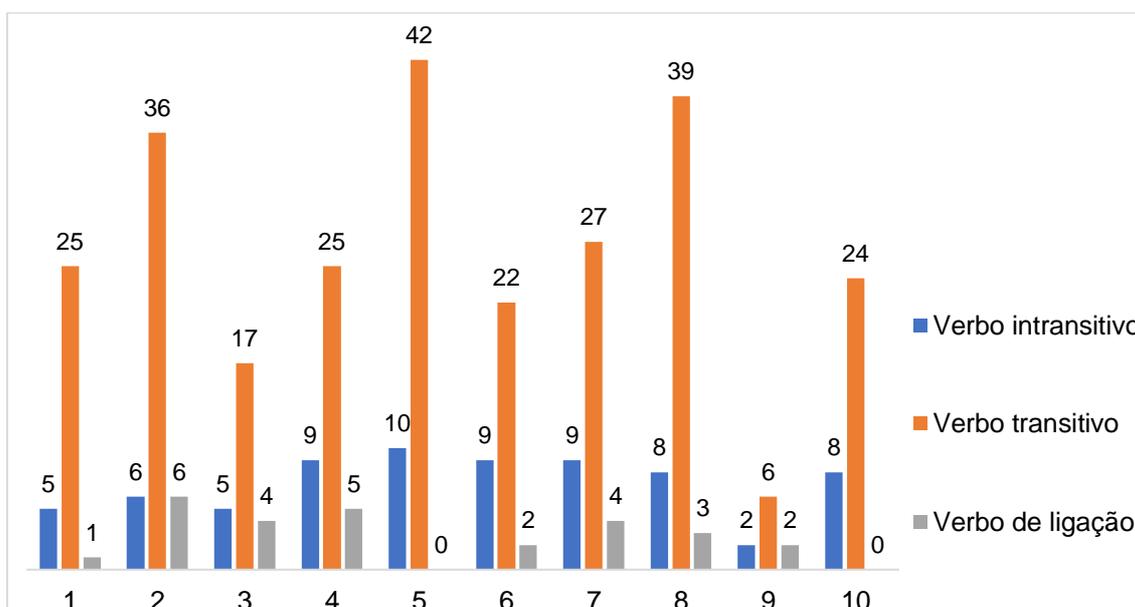
(JO-IF-JORUSP-cul-02set02\_02)

Acreditamos que a disposição textual em parágrafos subsequentes tende a distanciar os antecedentes de suas expressões correferentes, o que explica a alta frequência do uso de nomes repetidos neste contexto. Entretanto, podem ocorrer casos em que o antecedente esteja próximo ao final do parágrafo anterior, propiciando o surgimento do pronome pleno ou do pronome nulo por conta da quantidade de palavras e caracteres entre o antecedente e a retomada.

Por conta disso, é possível encontrar na língua, casos em que a última aparição do antecedente em posição de sujeito está próxima do final do parágrafo e, quando retomado pela primeira vez no parágrafo subsequente, levando em consideração a quantidade de palavras e de caracteres entre o antecedente e a retomada, os pronomes podem aparecer.

Reportamos, a seguir, a quantidade de verbos, divididos em três subcategorias na sentença anterior à retomada em dados do *corpus* escrito Lácio-Ref.

**Gráfico 18** – Ocorrência verbos intransitivos, verbos transitivos e verbos de ligação nos textos do *corpus* Lácio-Ref



Para quantificar os tipos de verbos presentes nas sentenças que estão antes da sentença na qual a retomada ocorre, analisamos, previamente 05 (cinco) tipos de verbos, sendo (i) verbo intransitivo, (ii) verbo transitivo direto, (iii) verbo transitivo indireto, (iv) verbo bitransitivo e (v) verbo de ligação. Entretanto, a ocorrência dos verbos transitivos indiretos e dos verbos bitransitivos foi tão baixa nos dados do *corpus* por nós analisado que optamos por reportar os verbos transitivos diretos, os verbos transitivos indiretos e os verbos bitransitivos na macrocategorias de verbos transitivos, mostrando apenas em números a quantidade de cada um deles.

Mostramos no exemplo (138, grifo nosso), a posição do verbo analisado nesta categoria, pois, levando em consideração o que foi feito em pesquisas anteriores que tratam do processamento correferencial, como visto no capítulo anterior, analisamos o tipo de verbo que ocorre na última sentença antes da retomada, pois, é esperado que diversos fatores de ordens sintáticas e semânticas tenha efeito sobre a forma anafórica a ser usada e, por isso, deve ser observado o que pode acontecer entre o referente e a retomada.

(138) [...] É a quinta vez que a capela da pequena cidade de Mülen, com apenas 300 habitantes, recebe o músico brasileiro; para que ele; grave um de seus discos. [...]

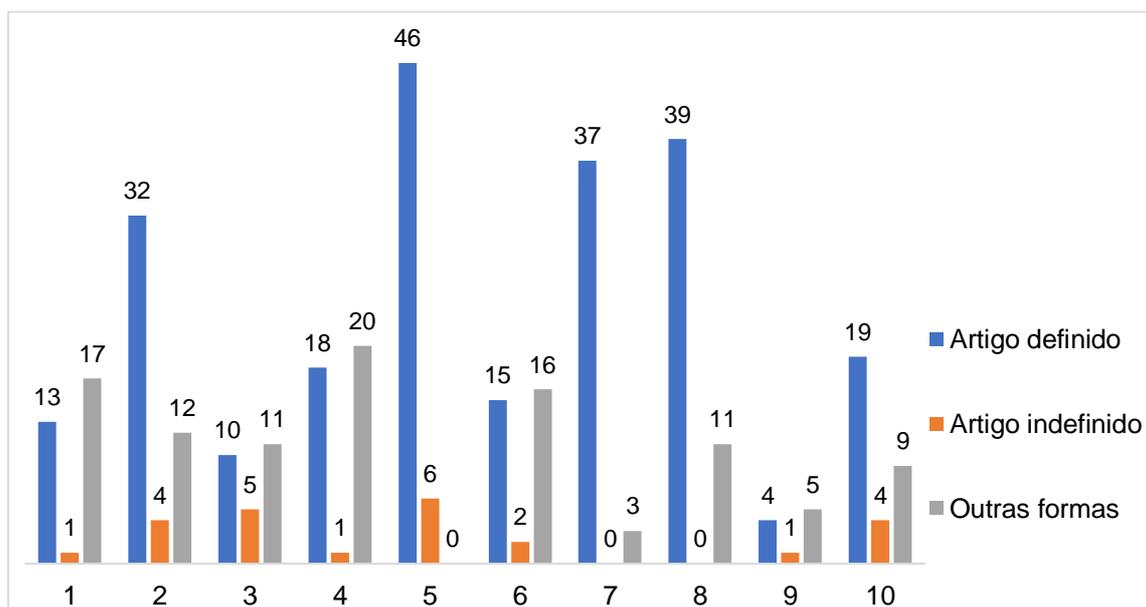
(JO-IF-JORUSP-cul-03jun02\_02)

A soma da quantidade de verbos na sentença anterior à retomada mostrou 361 ocorrências, sendo 71 por verbos intransitivos, 263 por verbos transitivos (254 transitivos diretos, 07 transitivos indiretos e 02 bitransitivos) e 27 por verbos de ligação, ou seja, 19,7% com verbos intransitivos, não significativa ( $p$ -valor = 0,58), 72,9% com verbos transitivos (96,5% transitivos diretos, 2,7% transitivos indiretos e 0,8% bitransitivos), também não significativa ( $p$ -valor = 0,91) e 7,4% com verbos de ligação, igualmente não significativa ( $p$ -valor = 0,15).

Com estes resultados, não buscamos encontrar variação significativa entre os tipos de verbo, mas apontar a ocorrência mais usual para então, levar à discussão sobre o padrão usual da língua em busca de parâmetros que desempenhem efeitos durante a resolução da correferência anafórica.

E, para finalizar a exposição de nossos resultados seguindo a sequência de nosso formulário, mostramos a seguir a quantidade de ocorrências dos artigos definidos e indefinidos precedendo o antecedente.

**Gráfico 19** – Ocorrência artigos definidos, indefinidos ou outras formas de definitude precedendo o referente nos textos do *corpus* Lácio-Ref



Para contabilizar a quantidade de artigos precedentes ao antecedente, por não ser de nosso interesse a definitude de nomes e demais formas linguísticas, atemo-nos a observar somente se os nomes eram precedidos de (i) artigos definidos, (ii) artigos indefinidos ou (iii) se não eram precedidos nem por um, nem por outro, como acontece quando o antecedente é um pronome pleno, nulo ou indefinido, por exemplo, ou quando são nomes próprios que ora são

precedidos de artigo, ora não e, por não ser objeto de nosso estudo, mas por fazer parte dos fatores que podem influenciar no processamento correferencial, analisamos três subcategorias, sendo, portanto, (i) o artigo definido que precede o antecedente, (ii) o artigo indefinido precedendo o antecedente e (iii) outras formas de definitude ou indefinitude que não são o artigo definido ou o artigo indefinido.

Com a soma dos valores da quantidade de antecedentes precedidos por artigos definidos, artigos indefinidos ou outras formas nos 10 (dez) textos, verificamos 361 ocorrências, sendo 233 (64,5%) dos antecedentes precedidos por artigos definidos, com  $p$  não significativo igual a 0,81, 24 por artigos indefinidos (6,7%), também com  $p$  não significativo (0,26), e 104 por outras formas (28,8%), da mesma forma com  $p$  não significativo igual a 0,56.

Observamos que a quantidade de antecedentes que são definidos antes de serem retomados é bem maior do que os antecedentes vagos introduzidos no texto, pois se compararmos apenas a quantidade de antecedentes precedidos por artigos definidos e por artigos indefinidos temos 257 ocorrências no total, sendo 233 antecedentes precedidos por artigos definidos e 24 por artigos indefinidos, implicando em 86,7% dos antecedentes precedidos por artigos definidos e 13,3% por artigos indefinidos.

Sobre as outras formas, podemos indicar que, quando retomados os textos analisados, a grande maioria delas também poderiam ser classificadas como expressões definidas ao longo do texto, mas que, por não ser de nossa alçada tratar aqui de questões semânticas ligadas à definitude, reduzimos as possibilidades à classificação explicada anteriormente.

Assim como a categoria de análise anterior, acreditamos que esta categoria também não tenha influência direta com a distribuição que buscamos explicar, mas decidimos levar em consideração por ser algo praticado na maioria das pesquisas citadas no capítulo anterior e, portanto, observamos a presença ou não de artigo precedendo o referente, conforme descrito na seção anterior.

Em (139), o referente *os pássaros* é introduzido com o artigo definido fazendo parte do nome<sup>196</sup>, sendo a forma antecedente da retomada feita pelo uso do pronome nulo no segundo centro anafórico. No terceiro e no quarto centro

---

<sup>196</sup> Nesta tese usamos os termos nome, repetido ou recategorizado, como sinônimo do termo sintagma nominal.

anafórico, os antecedentes não têm nenhum artigo compondo o referente por se tratar de pronome nulo.

(139) [...] Os pássaros, por exemplo, não exibem plumagens coloridas. Depois de  $\emptyset_i$  terem sido abertos para estudo,  $\emptyset_i$  parecem um tanto amassados e  $\emptyset_i$  têm, no lugar dos olhos, chumaços de algodão. [...]

(JO-IF-JORUSP-cul-02set02\_04)

E, uma vez expostos todos os resultados acima, damo-nos por satisfeitos em reportar as quantidades de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em posição de sujeito em textos escritos do *corpus* Lácio-Ref.

Na busca de indícios concretos de como estas expressões correferenciais são distribuídas em português brasileiro, buscamos elencar padrões já analisados em outras pesquisas psicolinguísticas sobre esta área como, por exemplo, o fato de sujeito retomar preferencialmente sujeito, o tipo de verbo da sentença anterior à sentença na qual há a retomada, a presença ou não de artigo definido e a distância entre o antecedente em quantidade de palavras e em quantidade de caracteres.

Com nossos resultados, esperamos que nossas interpretações levem, a partir dos dados da língua usada em contexto autêntico, a uma discussão ampla e esclarecedora sobre fatores que podem estar envolvidos na resolução da correferência anafórica e, assim, contribuir com pesquisas que tratam da correferência anafórica.

Levando em consideração os resultados encontrados e com interpretações depreendidas por meio e a partir das análises acima, empreendemos na próxima seção uma discussão com achados anteriores.

#### 4.5 Discussão

A análise dos resultados acima leva a interpretações que corroboram algumas pesquisas e outras não dentre as citadas nos capítulos anteriores e, a fim de buscar o máximo de correlação com cada uma delas, buscaremos mostrar em que concordamos ou não.

Porém, é necessário atentarmo-nos para o fato de que nossa pesquisa é documental e, portanto, descritiva, o que não nos dá indícios do que ocorre

durante o processamento tal como em pesquisas experimentais, mas nos dá, por meio de nossos resultados, indícios do que pode ou não ser percebido quando a língua é usada.

Desta forma, tratamos aqui da correferência anafórica e buscamos explicações de como estão distribuídos, em português brasileiro, os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos quando desempenham a função de expressões correferenciais na posição de sujeito.

Decidimos mostrar nossas interpretações seguindo os três pontos do capítulo anterior, a saber (i) teorias e hipóteses sintáticas sobre o processamento anafórico, (ii) teorias e hipóteses semânticas sobre o processamento e (iii) teoria pragmática sobre o processamento anafórico, interligando estas teorias com as pesquisas sobre o processamento anafórico em português brasileiro<sup>197</sup>.

No que tange especificamente a *pro*, ou seja, a categoria vazia que estudamos neste trabalho como pronome nulo (a contraparte do pronome foneticamente realizado) e com características -anafórico/+pronominal, é interessante uma pequena observação sobre a questão de o Português Brasileiro ser considerada uma língua, atualmente, parcialmente *pro-drop* (HOLMBERG; NAYADU; SHEEHAN, 2009).

Nossos resultados da quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos nos levam a crer que, de fato, o PB é uma língua parcialmente *pro-drop*, pois o que esperamos em uma língua com este parâmetro em um nível intermediário é a frequência de pronomes plenos e de pronomes nulos tão próxima que parece ser uma escolha feita ao acaso por parte do usuário.

Em nossos resultados, podemos ver que, de todas as 361 ocorrências de nomes e pronomes funcionando como expressão correferencial na posição de sujeito, 177 foram retomadas feitas somente por pronomes e, destas retomadas, 54 foram feitas por pronomes plenos e 123 por pronomes nulos, implicando em 69,5% das retomadas por pronomes nulos e 30,5% por pronomes plenos, pois não há diferença significativa ( $p$ -valor = 1,18) entre o uso do pronome nulo com

---

<sup>197</sup> Seguiremos, durante a discussão, a sequência apresentada no capítulo anterior, mas não achamos necessário subdividir o texto em categorias, já que estas podem ser facilmente revisitadas no capítulo anterior.

relação ao pronome pleno, quando levada em consideração a variabilidade dos textos analisados.

Assim, podemos dizer que há uma frequência ainda maior de pronomes nulos em textos escritos, o que apontaria para a efetivação do parâmetro *pro-drop*, mas que esta frequência não é tão forte a ponto de excluir os pronomes plenos, o que apontaria para a marcação de sujeitos, fazendo do PB uma língua não *pro-drop*.

Como acreditamos que o PB dá ao usuário opções de expressões correferenciais, vemos o pronome pleno como uma opção para o pronome nulo e, com o passar do tempo, por fatores não totalmente claros para nós, os usuários do PB têm preferido preencher o sujeito por um pronome foneticamente realizado. Assim, corroboramos Homlberg, Nayudu e Sheehan (2009) e afirmamos nossa crença de que o português brasileiro é uma língua parcialmente *pro-drop*, pelo menos na língua escrita, apesar de este não ser o objeto de nosso estudo<sup>198</sup>.

Para dar prosseguimento à nossa discussão, é preciso trazer os resultados sobre o uso das macrocategorias de nomes e pronomes funcionando como expressões correferenciais. Quando analisadas as macrocategorias, levamos em consideração o que foi proposto por Grosz e Sidner (1986) e por Grosz, Joshi e Weinstein (1995) quando da postulação da Teoria da Centralização e, por isso, buscamos contabilizar a quantidade de nomes e pronomes em posição de sujeito que retomavam um referente também em posição de sujeito.

Segundo proposto na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), buscamos averiguar se o uso dos pronomes é mais frequente do que o uso dos nomes para retomar uma entidade previamente introduzida no texto, pois, para os autores, quando há um centro anafórico a ser preenchido, este será preferencialmente preenchido por um pronome que é o indício de que o falante continua a falar da mesma coisa.

Rastreamos, durante a coleta de dados no *corpus* escrito Lácio-Ref, 361 retomadas na posição de sujeito, conforme proposto por Grosz e Sidner (1986)

---

<sup>198</sup> Acreditamos que as pesquisas sobre o parâmetro *pro-drop* estão mais ligadas à área da Sociolinguística, mas como faz parte de nossa fundamentação teórica e temos resultados que mostram a frequência de pronomes ao longo do texto, tivemos a possibilidade de dar evidências sobre o parâmetro *pro-drop* e o estágio no qual o Português Brasileiro se encontra com relação ao uso de pronomes plenos e de pronomes nulos.

e por Grosz, Joshi e Weinstein (1995), sendo 184 por nomes e 177 por pronomes, implicando em 51% das retomadas por nomes e 49% por pronomes, sem diferença significativa na frequência de uso ( $p$ -valor = 0,12)

Assim, percebemos que, pelo menos em textos escritos, os pronomes são pouco menos frequentes do que os nomes e, assim, temos um indício bastante curioso: se os pronomes são preferenciais, por que os nomes repetidos ainda são mais frequentes?

Para responder a esta pergunta, o primeiro ponto que precisamos levar em consideração é a vertente da língua na qual desenvolvemos nossa pesquisa: a língua escrita.

Tocamos neste ponto pelo fato de termos chamado a atenção anteriormente que as pesquisas psicolinguísticas parecem partir de possibilidades percebidas na língua falada, mas transposta para a língua escrita, o que pode levar a comportamentos específicos e distintos da fala.

Nas pesquisas psicolinguísticas revisitadas ao longo dos capítulos anteriores, nenhuma deixa claro os critérios elencados para a construção dos *designs* experimentais, mas sempre é relatado pelos pesquisadores que, de alguma maneira, a fala leva à construção das sentenças experimentais ao se referirem às possibilidades usadas pelos falantes.

Acreditamos que, aqui, reside um problema, pois, apesar de parâmetros serem respeitados tanto na língua escrita quanto na língua falada, há diferenças entre elas e, por tratarmos, nesta tese, de uma pesquisa em *corpora* linguísticos tendo como categorias de análise fatores elencados em pesquisas psicolinguísticas, o que foi postulado na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995) não foge a este problema.

O reflexo de que isto pode ser um problema em pesquisas psicolinguísticas é o fato de que, segundo nossos resultados, os nomes repetidos são mais frequentes do que os pronomes como formas correferentes em textos escritos, isto é, retomando uma entidade previamente introduzida e sendo retomada em um centro anafórico.

O que podemos inferir é que, tal como os nomes, os pronomes são as formas mais esperadas para retomar uma entidade previamente introduzida no texto, mas não a forma preferencial, pelo menos em textos escritos.

Além disso, temos de fazer uma observação: as pesquisas psicolinguísticas que motivaram nosso estudo buscam indícios que se apliquem

à língua como um todo em sentenças, no máximo em grupos de sentenças que chamam de nível textual/discursivo, como no caso da Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), mas acreditamos que, quando analisado um texto inteiro, com uma cadeia (cor)referencial em uso, temos muito mais indícios do que de fato acontece e, a primeira revelação importante que temos é: pronomes, em geral, são tão frequentes quanto nomes quando funcionam como formas correferentes.

Uma vez discutida a questão da frequência de pronomes no texto, continuamos nossa exposição dando prosseguimento à discussão que o resultado da quantidade de ocorrência destas formas nos levou a ter, pois, seguindo o que fora postulado na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995), Gordon, Grosz e Guilliom (1993) postularam que, se os pronomes devem retomar uma entidade previamente introduzida, é esperado que haja penalização dos nomes e, daí, surge o conceito do efeito de Penalidade do Nome Repetido.

Para Gordon, Grosz e Guilliom (1993), o nome ao ser repetido na posição de sujeito para retomar a entidade introduzida pelo mesmo nome é mais difícil de processar e, por isso, o pronome se torna a forma mais rápida de ser compreendida, surgindo daí o conceito do efeito de Penalidade do Nome Repetido.

O efeito de Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) tem sido amplamente pesquisado em português brasileiro e corroborado em diversas pesquisas psicolinguísticas (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; QUEIROZ, 2009; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; GONDIM, 2017) nas quais é afirmado o fato de que o uso dos pronomes plenos gera efeito de penalidade sobre os nomes repetidos.

Levando em consideração tal afirmação, em nossas análises, buscamos observar a frequência de ocorrências dos nomes repetidos comparando com a frequência de ocorrência dos pronomes plenos, buscando mais evidências sobre a distribuição das formas correferentes em textos escritos.

Porém, antes de apresentar os resultados encontrados para estas duas formas, temos de voltar à divisão de nossa macrocategoria dos nomes, pois podemos ter duas interpretações possíveis de nossos resultados: (i) nomes – incluindo nomes repetidos e nomes recategorizados – *versus* pronomes plenos

e (ii) nomes repetidos – excluindo nomes recategorizados – *versus* pronomes plenos.

Para (i) temos 238 retomadas no total, sendo 184 retomadas efetuadas por nomes e 54 por pronomes plenos, implicando em 77,3% das retomadas na posição de sujeito por nomes e 22,7% por pronomes plenos, sem diferença significativa ( $p$ -valor = 1,34). Para (ii) temos 150 retomadas no total, sendo 96 retomadas efetuadas por nomes e 54 por pronomes plenos, implicando em 64% das retomadas na posição de sujeito por nomes repetidos e 36% por pronomes plenos, também sem diferença significativa ( $p$ -valor = 1,29).

Ao levarmos em consideração que o português brasileiro é uma língua que admite a retomada do sujeito por meio do pronome nulo, tal qual o espanhol argentino, consideramos o que Gerlomi-Lezama (2008) denominou como Penalidade do Pronome Pleno.

Segundo este autor, em línguas em que o parâmetro do sujeito nulo existe, este será mais fácil de processar, ou seja, de ser compreendido, do que o pronome pleno. Este efeito foi corroborado em PB por algumas pesquisas (MAIA; CUNHA-LIMA, 2011, 2012, 2014; MAIA, 2013; LIMA, 2015) nas quais o uso do pronome nulo gerou efeito de penalidade sobre o nome repetido.

É interessante comentar o fato de que, quando buscados estes efeitos, os pesquisadores precisam analisar o tempo de processamento de nomes repetidos *versus* pronome plenos, para encontrar o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM 1993), e o nome repetido *versus* o pronome pleno *versus* o pronome nulo, para encontrar o efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008), pois o uso do pronome nulo gera efeito de penalidade sobre o uso do nome repetido, sendo, inclusive, lido mais rapidamente do que o pronome pleno, quando comparado o tempo de leitura das três formas correferentes.

Em nosso estudo, consideramos, portanto, que o uso do pronome nulo poderia ocorrer frequentemente, sendo uma das formas correferentes viáveis em português brasileiro, pelo fato de termos uma língua *pro-drop* e, em textos escritos, o uso desta forma correferencial ser bastante produzida.

Mais uma vez, analisamos nossos resultados, a fim de encontrar efeitos de penalidade e, retomamos a divisão exposta logo acima e comparamos (i) nomes *versus* pronome nulo e (ii) nomes repetidos *versus* pronome nulos tendo encontrado para (i) 307 retomadas no total, sendo 184 retomadas efetuadas por

nomes e 123 por pronomes nulos, implicando em 60% das retomadas na posição de sujeito por nomes e 40% por pronomes plenos, sem diferença significativa (p-valor = 1,35). Para (ii) temos 219 retomadas no total, sendo 96 retomadas efetuadas por nomes e 123 por pronomes nulos, implicando em 43,9% das retomadas na posição de sujeito por nomes repetidos e 56,1% por pronomes nulos, também sem diferença significativa (p-valor = 1,31)

Em busca de melhor compreender a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em PB e pautados em pesquisas psicolinguísticas sobre a correferência anafórica, achávamos estranho o fato de o nome ser sempre penalizado e por não ter sido reportado, nas pesquisas citadas, como a forma mais fácil de ser compreendida.

Logo começamos a associar a retomada do nome com a distância entre o referente e sua expressão correferente, mas vimos que, de acordo com Leitão e Simões (2011), quando o antecedente está afastado do correferente em até 38 sílabas há o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), sem ter sido reportado pelos autores a quantidade de caracteres sem e com espaço.

Lima (2015) mostrou que, havendo entre o antecedente e o referente uma sentença interveniente – o autor não reportou a quantidade de palavras e caracteres que distanciavam o referente da expressão correferencial, surgirá preferencialmente o pronome pleno e, sem esta sentença, surgirá preferencialmente o pronome nulo tendo, portanto, encontrado os dois efeitos citados acima, a depender do contexto, mas, estranhamente, os resultados apontaram para a não influência da distância no tipo de retomada. E, contrariando as hipóteses do autor, não houve contexto em que o nome repetido fosse a forma preferencial.

Considerando ainda estranho o fato de o nome não surgir como forma preferencial de uso e por termos encontrado mais nomes do que pronomes surgindo como forma correferente, de maneira geral, optamos por contabilizar a quantidade de palavras entre o referente e a retomada, algo ainda não feito nas pesquisas psicolinguísticas e, como forma de confirmação, analisamos também a quantidade de caracteres com e sem espaço entre o referente a retomada.

Com os resultados da distância feita por palavras e por caracteres temos algo interessante: quanto mais perto o referente estiver da retomada em quantidade de palavras, mais favorável é o uso do pronome nulo e, quanto mais

longe o referente estiver da retomada em quantidade de palavras, mais favorável é o uso do nome repetido, o que pode ser confirmado por meio do valor de significância encontrado através de rodadas estatísticas no software PAST ( $p$ -valor = 0,0001).

Para confirmar nossa interpretação, nossos resultados mostram que, quando entre o antecedente e a retomada houve até 10 (dez) palavras, 96% das retomadas foram feitas por pronomes nulos, enquanto houve apenas 4% por nomes repetidos. Quando analisadas comparativamente estas formas por meio de análise de variância, não encontramos diferenças significativas, mas podemos apontar uma possibilidade de caso em que o pronome é mais naturalmente usado na língua.

Curiosamente, encontramos que, quando entre o antecedente e a retomada há mais de 100 (cem) palavras, não houve nenhuma retomada por pronome pleno ou por pronome nulo, pois todas as ocorrências foram feitas por nomes, 39 por nomes repetidos e 21 por nomes recategorizados, implicando em 65% das retomadas efetuadas por nomes repetidos e 35% por nomes recategorizados. Esta pode ser mais uma evidência da influência da distância, em quantidade de palavras entre o antecedente e a retomada, em textos escritos em PB, já que há diferença significativa para o cruzamento entre a distância e a expressão anafórica usada ( $p$ -valor = 0,0001)

Como comprovação de que a distância em quantidade de palavras tem influência direta em como estabelecemos a correferência, buscamos indícios, em quantidade de caracteres do que pode acontecer, pois a quantidade de caracteres difere bastante em espaço, conforme mostramos na seção anterior.

Para a quantidade de caracteres sem espaço, nossos resultados reforçam que, quando menor a distância entre o referente e a retomada, mais frequente será o uso do pronome nulo. De todas as retomadas na distância de até 10 (dez) caracteres sem espaço, encontramos a ocorrência de 87,1% das retomadas feitas por pronomes nulos e 12,9% por pronomes plenos, sem diferença significativa ( $p$ -valor = 1,19), 89,7% por pronomes nulos e 10,3% por nomes recategorizados, também sem diferença significativa ( $p$ -valor = 1,24) e 98,4% por pronomes nulos e 1,6% por nomes repetidos, igualmente sem diferença significativa ( $p$ -valor = 1,39).

O mesmo padrão é seguido quando levada em consideração a quantidade de caracteres com espaço quando há, entre o antecedente e a retomada até 10

caracteres, tendo sido encontradas as mesmas quantidades de retomadas quando há até 10 caracteres com espaço, reforçando que há efeito de penalização sobre o nome repetido em pequena distância em quantidade de palavras, em quantidade de caracteres sem espaço e em quantidade de caracteres com espaço.

O efeito de distanciamento em quantidade de caracteres também comprova que, assim como ocorre em quantidade de palavras, quanto mais distantes em quantidade de caracteres sem espaço, mais frequente é o uso do nome repetido, já que de todas as retomadas feitas na distância de mais de 100 (cem) caracteres sem espaço, 59,3% das retomadas foram efetuadas por nomes repetidos e 40,7% por nomes recategorizado, sem diferença significativa ( $p$ -valor = 1,24) , 92,5% por nomes repetidos e 7,5% por pronomes plenos, também sem diferença significativa (1,29) e, por fim, 93,5% por nomes repetidos e 6,5% por pronomes nulos, igualmente sem efeito significativo ( $p$ -valor = 1,31).

E, quando levada em consideração a mesma distância em quantidade de caracteres com espaço, encontramos 53,7% das retomadas efetuadas por nomes repetidos, 38,4% por nomes recategorizados, 4,3% por pronomes plenos e 3,6% por pronomes nulos.

Comparar os resultados acima serve como uma boa comprovação de que a distância em quantidade de caracteres tem papel decisivo na forma correferente usada, pois levando em consideração a quantidade de caracteres sem espaço, o que torna o antecedente e a forma correferente mais próximos, há 6,5% das retomadas por pronomes nulos, enquanto que, no contexto em que foram contabilizados os caracteres com espaço, o que afasta o antecedente da forma correferente, há 3,6% das retomadas por pronomes nulos.

Por conta destes resultados, temos indícios que nos levam a acreditar que quanto maior a quantidade de palavras, de caracteres sem espaço e de caracteres com espaço, mais propício é o uso do nome repetido.

É interessante também observar como estão distribuídos os pronomes plenos e os nomes recategorizados, pois nossos resultados apontam que há distâncias em que o aparecimento de uma forma é tão frequente que parece seguir um padrão de uso (+pronome nulo/-nome repetido até 10 palavras, -pronome nulo/+nome repetido em mais de 100 palavras), mas que há formas que servem para serem usadas ocasionalmente em distâncias intermediárias: o pronome pleno e o nome recategorizado.

Temos esta convicção pelo fato de que encontramos, em nossos resultados, que, no intervalo de 11 (onze) a 50 (cinquenta) palavras e, depois, no intervalo de 51 (cinquenta e uma) a 100 (cem) palavras, quando o antecedente vai sendo afastado de seu referente, o uso do pronome pleno e do nome recategorizado se torna mais frequente do que o uso do pronome nulo.

No intervalo de 11 (onze) a 50 (cinquenta) palavras, de todas as retomadas feitas nesta distância, 26% foram feitas por nomes repetidos, 30,6% por nomes recategorizados, 25% por pronomes plenos e 18,4% por pronomes nulos. E, no intervalo de 51 (cinquenta e uma) a 100 (cem) palavras, de todas as retomadas feitas nesta distância, 48,9% foram feitas por nomes repetidos, 41,9% por nomes recategorizados, 2,2% por pronomes plenos e 7% por pronomes nulos.

Podemos perceber, com estes resultados, que a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em Português Brasileiro parece estar diretamente ligada a distância em quantidade de palavras e de caracteres entre o antecedente e a retomada.

Vemos que, quando o antecedente está entre 11 (onze) e 50 (cinquenta) palavras de distância da expressão correferencial, o uso do pronome pleno se torna mais frequente do que quando entre 51 (cinquenta e uma) e 100 (cem) palavras, acontecendo o inverso para o uso do nome recategorizado, que é menos frequente entre 11 (onze) e 50 (cinquenta) palavras do que entre 51 (cinquenta e uma) e 100 (cem) palavras, mostrando, assim, que o usuário da língua pode tê-los como opções para o pronome nulo e para o nome repetido, respectivamente.

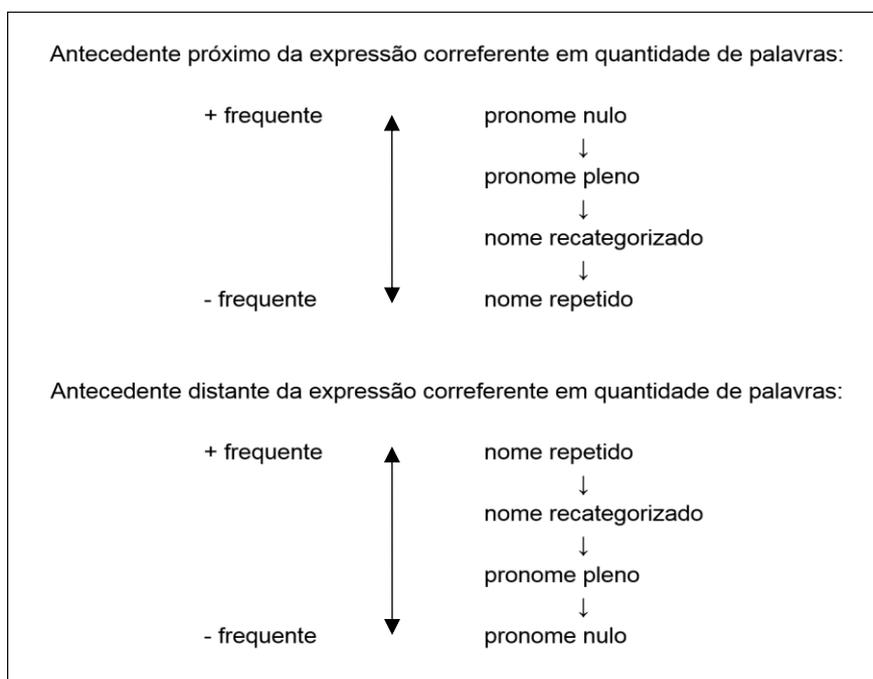
Este padrão é notado também quando levados em consideração a quantidade de caracteres (12,9% das retomadas feitas por nomes recategorizados e 24,2% por pronomes plenos no intervalo de 11 a 50 caracteres sem espaço e 13,6% das retomadas feitas por nomes recategorizados e 19,7% por pronomes plenos no mesmo intervalo de caracteres com espaço, enquanto há 25,5% das retomadas feitas por nomes recategorizados e 41,2% por pronomes plenos no intervalo de 51 a 100 caracteres sem espaço e 20% das retomadas feitas por nomes recategorizados e 44% por pronomes plenos nos mesmos intervalos de caracteres com espaço), reforçando nossas afirmações,

já que quanto mais próximo for o antecedente da retomada, mais frequente é o uso do pronome, neste caso, do pronome pleno<sup>199</sup>.

Acreditamos nisso pelo fato de que, se o uso do pronome nulo é mais frequente em distâncias curtas, é esperado que, ao ser afastado de seu antecedente, o pronome nulo seja menos usado e comece a ser substituído por sua forma opcional, o pronome pleno, no caso e, que, ao ser ainda mais afastado, o nome comece a ser preferencialmente usado. Como a distância pode não ser tão longa em quantidade de palavras, ao invés de usar o nome repetido, o usuário recai em opções para os nomes, como os nomes recategorizados.

Vejamos a distribuição destas formas correferentes no quadro 14.

**Quadro 14** – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em Português Brasileiro de acordo com a quantidade de palavras entre o antecedente a retomada em textos escritos



<sup>199</sup> Ao optarmos por quantificar a distância entre o antecedente e a retomada por caracteres sem ou com espaço, buscávamos apenas a confirmação de nossos resultados, mas encontramos um problema digno de ser pesquisado com mais afinco em pesquisas futuras, afinal, qual é a quantidade de caracteres em que os pronomes penalizam os nomes ou vice-versa? Acreditamos que, se construídos designs experimentais que busquem rastrear a quantidade de caracteres exatos e comparadas as retomadas por nomes repetidos, pronomes plenos e pronomes nulos, algum contexto mostrará o efeito da Penalidade do Nome Repetido, levando o pronome pleno a penalizar o nome repetido e, talvez, até mesmo o pronome pleno, sendo necessário adicionar informações ao conceito inicial desta penalidade no que diz respeito ao PB. Porém, estes são indícios que pensamos baseados na pesquisa de *corpus* escrito, o que pode não ser comprovado em teste laboratoriais.

Mesmo não tratando aqui sobre efeitos de penalidades, uma vez que a penalidade é um fenômeno exclusivamente cognitivo, acreditamos que o padrão de distribuição encontrado poderia apontar para efeito de penalidade dos nomes repetidos sobre os pronomes nulos, o que levaria a um novo efeito de penalidade.

Cabe ressaltar que estamos prevendo uma penalidade existente na língua a partir da frequência de nomes repetidos, nomes recategorizados, pronome plenos e pronomes nulos, o que necessita de comprovação em pesquisas experimentais para corroborar nossos achados, uma vez que efeitos de penalidade só podem ser percebidos em tarefas de processamento, não em produto linguístico e, por isso, hipotetizamos ser possível que tal fenômeno possa acontecer, sendo interessante, portanto, buscar evidências deste possível efeito em testes laboratoriais.

Ainda em busca de evidências sobre os efeitos encontrados em pesquisas psicolinguísticas que utilizaram de métodos experimentais, analisamos dois efeitos que devem ser levados em consideração quando se buscam efeitos de penalidade entre as formas correferentes: o efeito de Paralelismo Sintático e o efeito de Proeminência Sintática.

De acordo com o previsto por Gordon e Hendrick (1997), o sujeito é a posição sintática mais proeminente da sentença e, por isso, há maior tendência de ser retomado em centros anafóricos em posição de sujeito do que em posições não proeminentes, como a posição de objeto, por exemplo.

Chambers e Smith (1998) previram que, se o sujeito é a posição sintática mais proeminente, seria esperado que o centro anafórico na posição de sujeito retomasse mais antecedentes em posição proeminente do que em outras posições sintáticas, havendo, assim, um efeito de paralelismo entre as posições sintáticas preferenciais, tendo sido previsto, enfaticamente, que sujeito retoma sujeito.

Nossos resultados mostram indícios de que o efeito de Paralelismo Sintático (CHAMBERS; SMITH, 1998) pode ser percebido na língua escrita, já que das 361 retomadas analisadas por nós ao longo dos textos, 77,8% das ocorrências mostram a retomada entre sujeito/sujeito e 22,2% mostram outras posições sintáticas sendo retomadas na posição de sujeito, sem diferença significativa ( $p$ -valor = 0,38).

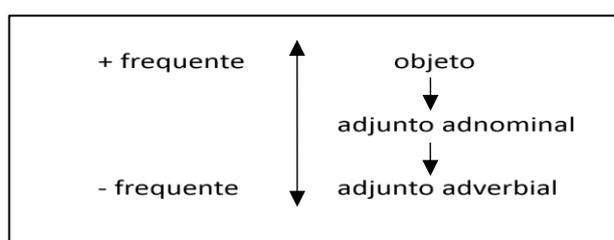
Estes resultados servem não só para corroborar pesquisas que tratam do efeito de Paralelismo Sintático (GORDON; HENDRICK, 1997; LEITÃO, 2005a), mas também daquelas que tratam do efeito de Proeminência Sintática, pois analisamos quais são as posições sintáticas que são retomadas posteriormente por uma expressão correferencial em posição de sujeito.

Analisamos somente as demais posições pelo fato de que, por termos encontrado indícios da ação do efeito de Paralelismo Sintático (CHAMBERS; SMITH, 1998), quando consideramos as retomadas feitas na posição de sujeito correlacionadas a uma entidade previamente introduzida também na posição de sujeito deve se sobressair como a forma mais proeminente sintaticamente, tornou-se redundante mostrar a quantidade de sujeitos que foram retomados também em posição de sujeito.

Interessou-nos, a partir de então, analisar se havia algum padrão de proeminência para as demais posições sintáticas, já que Gordon e Hendrick (1997) citam apenas que a posição sintática de sujeito é mais proeminente do que a posição de objeto e, quando analisamos os dados, encontramos retomadas também em posições de adjunto adnominal e de adjunto adverbial. Nossos resultados mostram que, de fato, um antecedente em posição de sujeito é preferencialmente retomado também em posição de sujeito, mas que, se o antecedente não estiver em posição de sujeito, pode acontecer de ele ser introduzido em outras posições sintáticas, tendo sido encontradas 50% de ocorrências do antecedente em posição sintática de objeto (p-valor = 0,99), 38,7% em posição adjunto adnominal (p-valor = 0,31) e 11,3% em posição de adjunto adverbial (p-valor = 0,10).

Mesmo não sendo este o objetivo central de nosso trabalho, acreditamos que estes resultados levam a mais um padrão que merece ser observado em pesquisas futuras, pois parece haver uma gradação de posições sintáticas a serem retomadas em posição de sujeito, como pode ser visto no quadro 15.

**Quadro 15** – Escala de posições sintáticas que introduzem o referente quando não há Efeito de Paralelismo Sintático na escrita



Se, em pesquisas futuras esta gradação for encontrada e, por sua vez, ocorrerem efeitos das penalidades supracitadas, teríamos novos indícios de que os efeitos de Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e de Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) não são exclusivos da retomada sujeito/sujeito, como previsto por Grosz e Sidner (1986) e por Grosz, Joshi e Weinstein (1995).

Como acreditamos que os efeitos de penalidade não são excludentes entre si, mas que dependem do contexto no qual ocorrem, achamos possível que a Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e a Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) possam também ser encontradas quando o antecedente está em uma posição menos proeminente e sem paralelismo sintático. Esta opinião nos leva a crer que, em contexto inverso, ou seja, o antecedente estando em posição de sujeito e sendo retomado em posições não proeminentes também é possível e pode mostrar efeitos de penalidade entre as formas correferenciais. Tal achado levaria a reformulações sobre os conceitos destas penalidades e, em PB, temos indícios de que o sujeito, ao ser retomado em posição de objeto gera o efeito da Penalidade do Nome Repetido (LEITÃO, 2005; QUEIROZ, 2009).

Na busca por contextos que possam vir a priorizar o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos, optamos por separar o texto em blocos que seguissem o padrão estudados em pesquisas psicolinguísticas em PB e analisamos a ocorrência destas formas em (i) orações subordinadas (MORGADO, 2011; LUEGI, 2012; LUEGI; MAIA, 2012; TEIXEIRA, 2014; TEIXEIRA, FONSECA; SOARES, 2014; HORA, 2014) (ii) orações coordenadas justapostas (LEITÃO, 2005; LEITÃO; MAIA, 2005; QUEIROZ, 2009; LEITÃO; SIMÕES, 2009; LEITÃO; RIBEIRO; MAIA, 2012; TEIXEIRA, 2014; SIMÕES, 2014; GONDIM, 2017)<sup>200</sup> (iii) orações coordenadas em períodos distintos (LEITÃO; PEIXOTO; SANTOS, 2012; MAIA; CUNHA-LIMA, 2011; MAIA, 2013; MAIA; CUNHA-LIMA, 2014; LIMA, 2014; LIMA, 2015; ALMOR *et al.*, 2017) e (iv) a primeira retomada em parágrafo subsequente.

---

<sup>200</sup> É preciso ressaltar que os estudos empreendidos por Leitão (2005), Leitão e Maia (2005), Queiroz (2009), Leitão e Simões (2009), Leitão, Ribeiro e Maia (2012), Teixeira (2014), Simões (2014) e Gondim (2017) tem estímulos experimentais tanto em orações coordenadas justapostas quando em orações coordenadas articuladas por conectivos e agrupamos, neste estudo, como orações coordenadas justapostas, entretanto, cada um destes estudos e exemplos de estímulos experimentais podem ser visto no capítulo anterior.

É importante notar que, das pesquisas acima, apesar de serem citados os contextos no qual a retomada ocorre, se dentro de oração coordenada ou de oração subordinada, de maneira geral os autores mostram como a estrutura sintática tem influência durante o processamento da anáfora, sendo feitas observações de que no contexto pesquisado ocorreu um efeito de penalidade e não outro.

Em contrapartida, encontramos indícios estatísticos de que o tipo de oração e o surgimento do nome repetido, do nome recategorizado, do pronome pleno ou do pronome nulo a ser usado no local da retomada são fatores que estão intimamente ligados.

Como nosso estudo não nos leva a inferir sobre como a mente do usuário da língua processa informações, podemos somente prever que o usuário faça dois cálculos sintáticos para efetuar a escolha da forma correferencial a ser usada: (1) a distância em quantidade de palavras entre o antecedente e a retomada e (2) o tipo de sentença na qual a retomada é efetuada.

Para (1) encontramos indícios de que, em textos escritos, quanto mais próximo o antecedente da retomada, mais frequente é o uso do pronome nulo enquanto que, quanto mais distante em quantidade de palavras o antecedente estiver da retomada, mais frequente é o uso do nome repetido.

Já para (2), temos que, em (i), orações subordinadas, de todas as retomadas feitas, 8,5% foram efetuadas por nomes repetidos (p-valor = 0,45), 1,6% por nomes recategorizados (p-valor = 0,41) , 20,4% por pronomes plenos (p-valor = 0,51) e 69,5% por pronomes nulos (p-valor = 0,82). Em (ii), orações coordenadas justapostas, de todas as retomadas feitas, 3,4% foram efetuadas por nomes repetidos (p-valor = 0,52), 2,3% por nomes recategorizados (p-valor = 0,68), 28,4% por pronomes plenos (p-valor = 0,47) e 65,9% por pronomes nulos (p-valor = 0,49). Em (iii), orações coordenadas em períodos distintos, de todas as retomadas feitas, 22,4% foram efetuadas por nomes repetidos (p-valor = 0,53), 48% por nomes recategorizados (p-valor = 0,43), 12,4% por pronomes plenos (p-valor = 0,56) e 17,2% por pronomes nulos (p-valor = 0,80). E, em (iv), parágrafos subsequentes, de todas as retomadas feitas, 77,8% foram retomadas efetuadas por nomes repetidos (p-valor = 0,82) e 22,2% por nomes recategorizados (p-valor = 0,57) e nenhuma por pronomes plenos ou por pronomes nulos nos textos analisados.

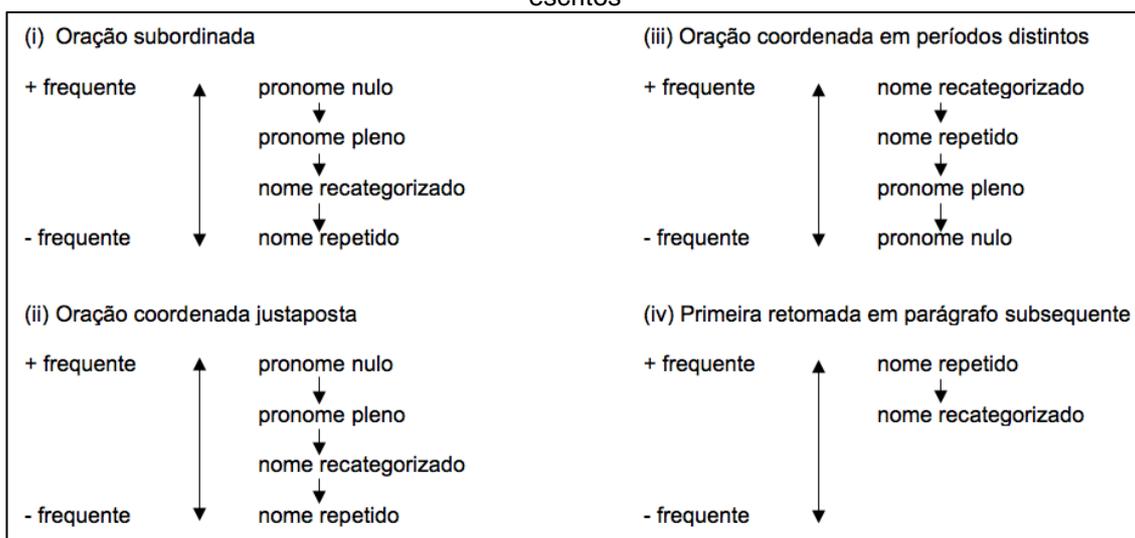
Estes resultados nos mostram que, em (i), os pronomes nulos são mais frequentes, em (ii) os pronomes nulos também são mais frequentes, em (iii) os nomes recategorizados são mais frequentes e, em (iv), um contexto não antes pesquisado, os nomes repetidos são mais frequentes.

Traçando um paralelo com a quantidade de palavras, estes resultados, de maneira geral, não são inesperados, visto que em (i), geralmente, o antecedente está próximo da retomada em quantidade de palavras, o que também ocorre em (ii), mas que, levando em consideração que chamamos de orações coordenadas justapostas aquelas que estão no mesmo período ligadas por vírgulas ou por conjunções, pode ocorrer de o antecedente estar bastante ou medianamente próximo da retomada, o que explica a alta frequência de pronomes nulos e o aumento da frequência de pronomes plenos quando comparado a (i). Já em (iii), como há tendência de o antecedente estar medianamente distante da retomada, o aumento da frequência de nomes recategorizados fica explicado, assim como a frequência alta de pronomes plenos também, já que os pronomes nulos diminuem em quantidade de ocorrências e o nome repetido tem aumento de frequência, indicando o efeito que a distância entre palavras gera ao processamento correferencial. Em (iv), contexto que não havia sido considerado em pesquisas psicolinguísticas que tratam do processamento correferencial, temos a primeira retomada do parágrafo subsequente sendo feita preferencialmente por nome repetido, mas também sendo feita por nome recategorizado, o que acontece porque, em geral, o antecedente fica mais distante da retomada quando feita a última referência em um parágrafo e retomado pela primeira vez no parágrafo seguinte.

Assim, encontramos o que parece seguir outro padrão de retomadas, diretamente ligado à quantidade de palavras entre o antecedente e a retomada, mas não podemos deixar de comentar que acreditamos ter encontrado indícios de como nossa mente processa a informação durante o processamento correferencial, mas tais fatores carecem de mais indícios por meio de pesquisas experimentais.

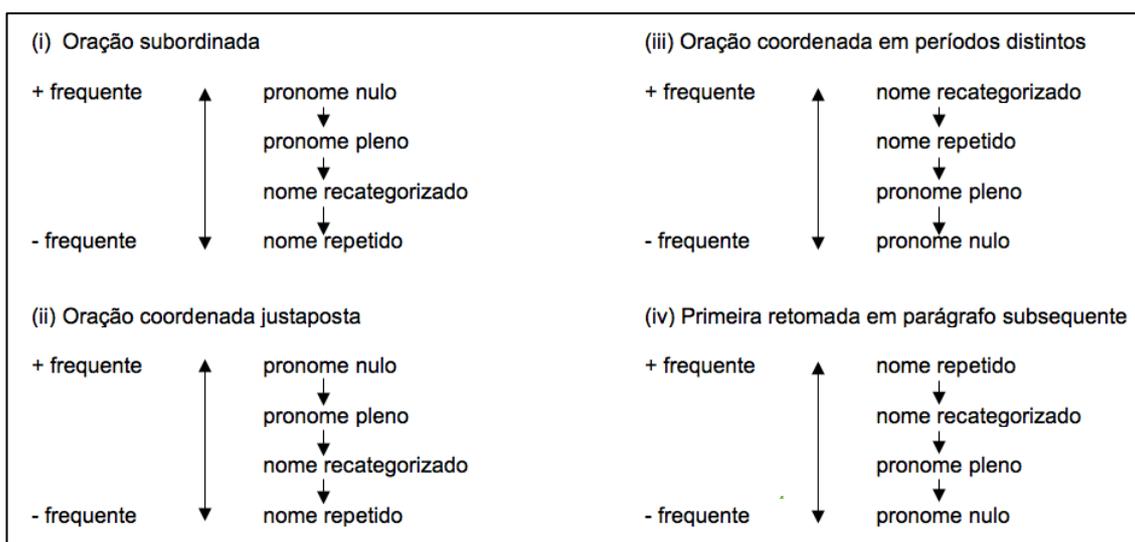
Logo, se estivermos certos, temos um padrão de retomadas que interliga o tipo de oração ou segmento textual no qual a retomada ocorre e a quantidade de palavras, gerando a distribuição conforme no quadro abaixo.

**Quadro 16** – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB de acordo com a disposição textual em textos escritos



Porém, a fim de utilizar as quatro formas de maneira mais bem distribuídas, apesar de nossos resultados, que podem ser mera coincidência, ou até mesmo por questões do gênero analisado, acreditamos que (iii) e (iv) estão com a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos de acordo com os dados e não com todas as possibilidades que há na língua, propomos que a melhor distribuição nestes contextos seria como disposto no quadro 17.

**Quadro 17** – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB de acordo com a disposição textual em textos escritos (modificado)



Optamos por fazer esta ressalva porque estamos mostrando o que acreditamos ser um fator importante a ser considerado em pesquisas futuras, pois é preciso considerar a distância em palavras entre o antecedente e a retomada bem como o tipo de oração ou segmento textual no qual será efetuada a retomada e, levando em consideração estes fatores, acreditamos que, em (ii) a ocorrência do pronome pleno deve ser maior do que do pronome nulo e que, em (iv), apesar de não terem sido encontradas ocorrências de pronomes plenos e de pronomes nulos, estes podem ocorrer em dado contexto.

Além disso, é preciso frisar que estamos fazendo uma escala de aparição das formas correferenciais por nós estudadas a depender do contexto sintático e que, se da forma mais frequente para a menos frequente uma delas não aparecer, o usuário terá a disposição a forma seguinte.

Com isto, queremos dizer que, se em uma oração subordinada, por exemplo, o usuário preferiu a forma correferencial do pronome pleno, em algum momento da resolução correferencial, ele pode ter cogitado antes o uso de um pronome nulo, mas isto é algo que supomos e que ficará a cargo de pesquisas futuras comprovar se isto ocorre ou não.

A seguir, decidimos apenas averiguar a frequência de dois fatores que, nas pesquisas psicolinguísticas sobre o processamento correferencial ocorrem, mas que não ficam claras as motivações de uso: o tipo de verbo e a presença do artigo.

Averiguamos o tipo de verbo que ocorre na sentença anterior à sentença na qual há a retomada e analisamos os verbos divididos em três subcategorias: verbos intransitivos, verbos transitivos e verbos de ligação. Encontramos a ocorrência de 19,7% com verbos intransitivos, não significativa ( $p$ -valor = 0,58), 72,9% com verbos transitivos (96,5% transitivos diretos, 2,7% transitivos indiretos e 0,8% bitransitivos), também não significativa ( $p$ -valor = 0,91) e 7,4% com verbos de ligação, igualmente não significativa ( $p$ -valor = 0,15).

Com isso, apenas gostaríamos de comentar que as pesquisas feitas em PB, por serem a maioria com verbos transitivos diretos (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; MAIA; CUNHA-LIMA, 2011, 2014, 2014; MAIA, 2013; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; TEIXEIRA, 2013; SIMÕES, 2014; LIMA, 2015; GONDIM, 2017) na sentença anterior à sentença na qual há a retomada, mostram um padrão autêntico da língua.

De qualquer maneira, gostaríamos de frisar que não acreditamos que o tipo de verbo – verbo intransitivo, verbo transitivo direto ou verbo de ligação – seja um dos fatores mais importantes durante o processamento correferencial e que gere efeitos de penalização entre as formas correferentes.

E, por fim, averiguamos que os antecedentes tendem a ser introduzidos por artigos definidos, por artigos indefinidos ou por outras formas, conforme explicado na seção anterior e encontramos 64,5% dos antecedentes precedidos por artigos definidos, 6,7% por artigos indefinidos e 28,8% por outras formas, o que, de maneira geral, também é comumente feito nas pesquisas psicolinguísticas em PB sobre o processamento correferencial.

E, no que tange às teorias e hipóteses semânticas sobre o processamento anafórico, gostaríamos de afirmar que compartilhamos da crença de que o processamento correferencial passar por dois estágios, conforme proposto por Van Gompel, Liversedge e Pearson (2004). Para nós, o processamento correferencial depende, no mínimo, de dois estágios, estando o estágio voltado para a identificação do antecedente, o que envolve mais questões de cunho sintático, para nós e, um estágio subsequente, no qual a forma semântica da anáfora é verificada, estando envolvidas, aqui, questões de número, de gênero e de animacidade/inanimacidade, por exemplo.

Além desta teoria, também compartilhamos da crença de que, dentre as opções que o usuário da língua dispõe para efetuar a correferência, há formas que são mais frequentes do que outras, o que desemboca nos achados de Almor (1999). Assim como este autor, acreditamos que existem formas que demandam maior esforço para ser ligado ao seu antecedente e, por isso, podem gerar prejuízos para a compreensão global do texto.

Por fim, acreditamos ter indícios para corroborar a Teoria da Acessibilidade (ARIEL, 1991), pois, para a autora, há um padrão universal, que pode variar entre as línguas, para como as formas linguísticas são acessadas na mente do usuário, variando de formas mais acessíveis a formas menos acessíveis.

Na escala proposta por Ariel (1991), os pronomes nulos, que ela chama de *zeros*, são altamente acessíveis enquanto os nomes, a depender de como usados no texto, ficam cada vez menos acessível.

Em nosso trabalho, associamos esta acessibilidade ao fator de distanciamento entre os antecedentes e as retomadas em quantidade de

palavras e não analisamos todas as formas propostas pela autora, mas, dentre as formas pesquisadas neste trabalho, temos fortes indícios de que Ariel (1991) estava certa.

Acreditamos ter encontrado indícios do que a autora chama de *padrão universal de acessibilidade*, pois, analisando dados de *corpus* escritos, vimos que os pronomes nulos retomam antecedentes mais acessíveis, os pronomes plenos e os nomes recategorizados estão em nível intermediário e os nomes repetidos são utilizados para trazer de volta à mente<sup>201</sup> do interlocutor os referentes que estão em baixa acessibilidade, ou seja, mais distante do antecedente.

Assim, buscando retomar os resultados encontrados nesta pesquisa em uma ampla discussão com as pesquisas que nortearam o nosso trabalho, esperamos ter trazido luz sobre como ocorre a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em Português Brasileiro quando funcionam como expressões correferenciais em dados de textos escritos.

Esperando que mais indícios possam ser encontrados e, então, buscando dar mais robustez aos nossos resultados e ao que foi discutido nesta seção, trataremos, no próximo capítulo desta tese, sobre como estão distribuídos os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos quando funcionam como expressões correferenciais na posição de sujeito em Português Brasileiro a partir da coleta em dados de fala.

---

<sup>201</sup> Ao longo de nosso trabalho, em alguns momentos, citamos termos como *memória* e *mente*, mas fazemos tais usos de maneira genérica, pois não é nosso objetivo, aqui, buscar indícios quantitativos do que, de fato, acontece no que tange à memória durante o processamento correferencial. Entretanto, acreditamos que, ao tratar da distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial em uma língua, torna-se quase impossível não interligar as questões linguísticas à memória. Este fato, por exemplo, acontece em nosso trabalho no que diz respeito à quantidade exata de palavras e de caracteres que fazem o usuário da língua retomar um antecedente por nomes. Acreditamos, pois, que há questões que extrapolam o âmbito linguístico e que ficam a cargo da memória durante o processamento correferencial, mas atentamo-nos apenas aos fatores linguísticos, o que explica a ausência de um capítulo que trate de memória em nosso trabalho.

## 5 EVIDÊNCIAS SOBRE O PROCESSAMENTO CORREFERENCIAL EM DADOS DE *CORPUS* DE FALA

A maneira como as expressões correferenciais são usadas, em especial, quando a retomada do antecedente é feita por meio de nomes repetidos, pronomes plenos e pronomes nulos, tem levado diversos pesquisadores (GORDON; GROSZ; GUILLIOM, 1993; GORDON; HENDRICK, 1997; CHAMBERS; SMITH, 1998; GERLOMINI-LEZAMA, 2008, entre outros) da área da Psicolinguística a investigar como os usuários da língua fazem a escolha de um pronome pleno ao invés de um nome repetido durante tarefas de leitura, por exemplo.

Dispondo de evidências baseadas em experimentos de leitura, tem-se afirmado que os nomes repetidos são mais lentamente processados do que os pronomes plenos tanto em línguas não *pro-drop*, como o inglês (GORDON; GROSZ; GUILLIOM, 1993) quanto em línguas *pro-drop*, como o português brasileiro (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; LIMA, 2015; GONDIM, 2017). Entretanto, também há evidências de que os nomes repetidos são mais lentamente processados do que os pronomes nulos em línguas *pro-drop*, como o espanhol argentino (GERLOMINI-LEZAMA, 2008) e o português brasileiro (MAIA; CUNHA-LIMA, 2011, 2012; MAIA, 2014; LIMA, 2015).

O Português Brasileiro é uma língua na qual as pesquisas desenvolvidas sobre o processamento correferencial levam ao encontro tanto da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) quanto da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008), tendo sido apontado por Lima (2015) que uma penalidade não exclui a outra, mas que não co-ocorrem no mesmo contexto, indicando, assim, para uma possível influência do contexto durante a resolução correferencial, o que também foi cogitado por Almor *et al.* (2017). Além disso, é importante levar em consideração que a maneira como as penalidades e usos de formas correferentes são processadas depende de fatores de diversas ordens (GONDIM, 2017).

O fato é que, em nenhuma pesquisa realizada com participantes sem patologias até agora sobre o processamento correferencial, o nome surge como

a forma mais fácil de ser processada em línguas como o Inglês, o Português Brasileiro e o Espanhol Argentino<sup>202</sup>, por exemplo.

Contudo, nossa pesquisa em dados de *corpus* escrito traz evidências sobre um possível padrão de distribuição pautada no uso, buscamos mais embasamento para nossos objetivos e hipóteses<sup>203</sup>, descrevendo, neste capítulo, os procedimentos para a coleta dos dados em *corpus* de fala, os resultados e as discussões relativas ao uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos.

### 5.1 Procedimentos para a obtenção de dados empíricos no *corpus* de fala

Foi utilizado para o levantamento das ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em posição de sujeito, o material disponível no projeto PORCUFORT<sup>204</sup>.

O Projeto Oral Culto Fortaleza (PORCUFORT) foi idealizado e coordenado pelo professor doutor José Lemos Monteiro, em parceria com o CNPQ, com o objetivo de organizar um banco de dados da variedade urbana falada pelos fortalezenses.

O *corpus* PORCUFORT<sup>205</sup> é constituído por dados de 73 (setenta e três) informantes estratificados por sexo, faixa etária e tipo de registro<sup>206</sup>. Sendo os dois primeiros critérios de coleta do *corpus* não relevantes para nossa pesquisa,

---

<sup>202</sup> Na verdade, desconhecemos pesquisas em outras línguas que apontem para os nomes como a forma menos penalizada, pois há consenso geral de que o uso do pronome sempre gera efeito de penalidade sobre o nome. Decidimos não apresentar pesquisas em diversas línguas para não ficarmos andando em círculos sobre o mesmo objeto de estudo e resultados parecidos e, por isso, optamos por nos concentrar nas línguas de onde foram originados os conceitos da Penalidade do Nome Repetido e da Penalidade do Pronome Pleno, ou seja, o inglês e o espanhol argentino.

<sup>203</sup> Os objetivos e hipóteses tanto da pesquisa em dados de *corpus* escrito quanto em dados de *corpus* de fala são os mesmos e, por isso, foram reportadas no capítulo anterior como comuns a toda a pesquisa desenvolvida neste trabalho.

<sup>204</sup> Durante o período de levantamento dos *corpora* de língua falada para a seleção do *corpus* a ser usado em nossa pesquisa, analisamos o *Corpus* Brasileiro, idealizado e coordenado pelo pesquisador Tony Berber Sardinha e tentamos ter acesso ao C-ORAL BRASIL, coordenado pelos pesquisadores Tommaso Raso e Heliana Mello, mas não obtivemos sucesso. Além disso, pesquisamos por potenciais *corpus* pelo site do projeto LINGUATECA, mas os *corpora* dispunham somente de sentenças, o que não atenderia nossos objetivos. Por fim, tivemos acesso aos dados dos *corpora* NORPUFORT e PORCUFORT, por meio de colegas e professores em contato com a professora Aluiza Alves de Araújo, tendo este último atendido aos nossos critérios que nortearam a seleção tanto do *corpus* escrito quanto do *corpus* de fala.

<sup>205</sup> O *corpus* PORCUFORT foi digitalizado e está sob a responsabilidade da professora Aluiza Alves de Araújo da Universidade Estadual do Ceará – UECE. O *corpus* pode ser encontrado por meio do site <http://www.uece.br/posla/index.php/projetos-de-pesquisa/linha-02> para buscar o e-mail da professora e, assim, solicitar o acesso.

<sup>206</sup> Para mais informações, ver Araújo, Viana e Pereira (2018).

pois não atrelamos o tipo de expressão correferencial que o usuário da língua faz a questões ligadas a sexo ou idade, detemo-nos ao terceiro critério: o tipo de registro.

Os registros que compõe o *corpus* PORCUFORT foi dividido em três tipos de registro: (i) diálogo entre informante e documentador – DID, (ii) diálogo entre dois documentadores – D2 e, (ii) elocução formal – EF.

No *corpus* PORCUFORT, há, 31 inquéritos classificados como DID, 14 inquéritos D2 e 19 inquéritos EF. Para a coleta destes dados, os pesquisadores selecionaram sujeitos<sup>207</sup> que fossem fortalezenses natos, filhos de pais cearenses, residentes em Fortaleza e que não tivessem feito longas viagens para fora da capital cearense, além de possuírem o nível superior completo.

A partir dos critérios estabelecidos, os pesquisadores chegaram ao *corpus* dividido pelos três critérios elencados por eles a fim de coletar um *corpus* que obedecesse aos padrões da Sociolinguística Variacionista, chegando ao resultado final conforme mostrado no quadro 18.

**Quadro 18** – Estratificação do *corpus* PORCUFORT <sup>208</sup>

Faixa Etária	Sexo					
	Masculino			Feminino		
	Tipo de Registro					
	DID	D2	EF	DID	D2	EF
I (22-35)	6	4	3	4	8	4
II (36-55)	6	4	5	5	4	3
III (56 em diante)	6	3	2	3	3	0
<b>Total</b>	<b>18</b>	<b>12</b>	<b>10</b>	<b>12</b>	<b>14</b>	<b>07</b>

O tipo de registro (i) é resultado de uma entrevista feita com um informante que conduzia determinadas perguntas a fim de registrar as respostas dadas naturalmente por meio do sujeito entrevistado, o tipo (ii) é resultado de conversas entre informantes com o mesmo grau de intimidade, como amigos ou familiares, fazendo destes inquéritos os mais importantes dos três registros coletados

<sup>207</sup> Optamos por não adentrar detalhadamente no perfil de cada sujeito, pois acreditamos que não é de basilar importância para nossos resultados, mas decidimos seguir o mesmo padrão levado em consideração para os textos selecionados na pesquisa com os dados do *corpus* escrito, descrito no capítulo anterior.

<sup>208</sup> Para mais informações, ver Araújo, Viana e Pereira (2018).

(ARAÚJO; VIANA; PEREIRA, 2018), enquanto o tipo (iii) é resultado é o registro mais formal dos três.

Como, por questões já explicadas no capítulo anterior, em nosso trabalho, buscamos analisar as retomadas feitas por meio de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito em textos escritos e falados em Português Brasileiro, necessitávamos selecionar os registros mais formais possíveis, a fim de manter um padrão com as pesquisas psicolinguísticas<sup>209</sup> que nortearam nosso trabalho e com os dados do *corpus* escrito, levando-nos, assim, a selecionar o tipo de registro de elocuições formais.

Segundo Araújo, Viana e Pereira (2018, p. 185)

o EF é o tipo de registro mais formal dentre os três que constituem o *corpus* PORCUFORT. É composto, geralmente, por aulas, exposições e conferências nos quais os informantes mantêm um nível muito elevado de monitoramento da própria fala.

Além disso, Araújo, Viana e Pereira (2018) afirmam que as gravações transcritas no tipo de registro EF do *corpus* PORCUFORT foram feitas, na maioria das vezes, em salas de aula ou auditórios universitários, o que levou com que os registros das falas dos informantes apresentassem o tempo mínimo de duração de 30 minutos.

A fim de compor nossos textos para análise ainda levando em consideração o padrão feito para a seleção dos textos analisados do *corpus* escrito, buscamos analisar 10 (dez) textos falados por informantes que tivessem perfis de jovens adultos, o que ficaria próximo do perfil buscado para compor os sujeitos participantes de pesquisas experimentais sobre o processamento correferencial e, por isso, selecionamos os textos que compõe a faixa etária I (25 – 35).

Reiteramos que o quesito idade não é um fator relevante em nossa pesquisa, mas buscamos um padrão de escolaridade, pois necessitávamos de informantes graduandos ou graduados e, como descrito acima, o *corpus* PORCUFORT, sobretudo para o tipo de registro EF, teve informantes graduados.

Levando isso em consideração e pelo fato de não termos os 10 (dez) textos necessários retirando somente os inquéritos dos informantes da faixa

---

<sup>209</sup> Ver capítulo 2 e capítulo 3.

etária I, decidimos completar a quantidade com inquéritos dos informantes da faixa etária II, ficando, portanto, nossa análise debruçada em 07 (sete) textos da faixa etária I e 03 (três) textos da faixa etária II, dos quais 05 (cinco) textos são registro de homens e 05 (cinco) de mulheres<sup>210</sup>.

Mostramos, no quadro abaixo, as normas de transcrição adotadas para a realização das transcrições para o PORCUFORT e que são encontradas ao longo de todos os inquéritos que compõem o *corpus*.

**Quadro 19** – Normas de transcrição adotadas no PORCUFORT

Ocorrências	Sinais	Exemplificação
Incompreensão de palavras ou segmentos	( )	Do nives de rensa ( ) nível de renda nominal
Hipótese do que se ouviu	(hipótese)	(estou) meio preocupado (com gravador) o
Truncamento (havendo homografia, usa-se acento indicativo da tônica e/ou timbre)	/	E comé/e reinicia
Entonação enfática	Maiúscula	Porque as pessoas reTÊM moeda
Prolongamento de vogal e consoante (como s, r)	:: podendo aumentar para :::: ou mais	Ao emprestarmos éh:: ... dinheiro
Silabação	-	Por motivo tran-sa-ção
Interrogação	?	E o Banco... Central... certo?
Qualquer pausa	...	São três motivos... ou três razoes ... que fazem com que se retenha moeda ... existe uma ... retenção
Comentários descritivos do transcritor	((minúscula))	((tossiu))
Comentários que quebram a sequência temática da exposição: desvio temático	- - - -	... a demanda de moeda - - vamos dar casa essa notação - - demanda de moeda por motivo ...
Superposição, simultaneidade de vozes	Ligando as linhas	a. na casa de sua irmã b. [sexta-feira? a. fazem LÁ b. [cozinham lá
Indicação de que a fala foi tomada ou interrompida em determinado ponto. Não no seu início, por exemplo.	(...)	(...) nós vimos que existem...

<sup>210</sup> Mesmo não levando em consideração critérios de sexo e de idade, optamos por balancear os registros analisados em busca de obter resultados fieis ao que é usado na língua com relação aos nomes repetidos, aos nomes recategorizados, aos pronomes plenos e aos pronomes nulos funcionando como expressões correferenciais.

Citações literais de textos, durante a gravação	"entre aspas"	Pedro Lima... ah escreve na ocasião. "O cinema falado em língua estrangeira não precisa de nenhuma baRREIra entre nós"...
<p>Iniciais maiúsculas: só para nomes próprios ou para siglas (USP, etc)</p> <p>2. Fáticos: ah, éh, ahn, ehn, uhn, tá (não por está: tá? Você está brava?)</p> <p>3. Nomes de obras ou nomes comuns estrangeiros são grifados.</p> <p>4. Números por extenso.</p> <p>5. Não se indica o ponto de exclamação (frase exclamativa)</p> <p>6. Não se anota o <i>cadenciamento da frase</i>.</p> <p>7. Podem-se combinar sinais. Por exemplo: oh:::... (alongamento e pausa)</p> <p>8. Não se utilizam sinais de pausa, típicas da língua escrita, como ponto e vírgula, ponto final, dois pontos, vírgula. As reticências marcam qualquer tipo de pausa.</p>		

(Fonte: Araújo; Viana; Pereira, 2018)

Levando em consideração estes sinais que surgiram ao longo do texto e os critérios explicados acima, selecionamos 10 (dez) textos para serem analisados levando em consideração as mesmas categorias de análise do *corpus* escrito, descrito no capítulo anterior.

Entretanto, como preenchemos o mesmo formulário tanto para os dados do *corpus* escrito como para os dados do *corpus* de fala, optamos por, para os dados analisados neste *corpus*, excluir as categorias de quantidade de caracteres sem e com espaço entre o antecedente e a retomada.

Tomamos esta decisão, levando em consideração as normas de transcrição do PORCUFORT, conforme exposto no quadro acima, por dois motivos: (i) a contagem de caracteres parece ser assertiva em dados de língua escrita, talvez por este motivo sejam computadas as quantidades de caracteres sem e com espaço entre o antecedente e a retomada em algumas pesquisas psicolinguísticas, enquanto nas pesquisas com dados de fala, em geral, há somente a computação da quantidade de ocorrências do objeto de estudo e, (ii) os sinais utilizados na transcrição do PORCUFORT fazem o resultado não ser fiel à quantidade de caracteres sem e com espaço entre o antecedente e a retomada.

Além disso, acreditamos que excluir a categoria de análise da quantidade de caracteres sem e com espaço entre o antecedente e a retomada não gera prejuízo para nossos resultados, pois esta categoria funciona mais como uma confirmação para a categoria de quantidade de palavras entre o antecedente e a retomada do que como um resultado a ser analisado de maneira assertiva por si só.

Ainda para chegar aos nossos resultados, tivemos de fazer pequenas alterações na disposição textual, uma vez que os textos não tem pontuação final nem são paragrafados, seguindo de forma ininterrupta.

Como nossas subcategorias de análise, conforme descrito no capítulo anterior, tratam sobre fatores de ordem sintática ligados tanto à estrutura interna quanto externa da sentença e em parágrafo subsequente, tivemos de, ao ler os textos com transcrição de fala, fazer pontuação e paragrafação.

Segundo Jubran *et al.* (2002), quando o pesquisador trabalha com *corpus* de fala, em muitos casos, é preciso reorganizar o discurso oral para que se possa, a partir de então, buscar evidências do objeto de estudo. De acordo com os autores, o discurso pode ser reorganizado, desde que seja levado em consideração (i) o tópico discursivo, (ii) a hierarquia discursiva e (iii) a linearidade discursiva.

A possibilidade de reorganização do discurso de fala em *corpus* orientada por Jubran *et al.* (2002) não fica distante do que é descrito no Projeto Norma Linguística Urbana Culta (NURC) ao apontar que, para compreender e trabalhar com o discurso oral, é preciso: a) identificar e delimitar as unidades tópicas, b) caracterizar as relações de interdependência hierárquica e sequencial entre os tópicos, e, c) detectar traços reveladores da estrutura interna das unidades tópicas.

A fim de facilitar a reestruturação também em sentenças e parágrafos, levando em consideração as questões discursivas apontadas acima, trabalhamos com a noção de parágrafo, formado por sentenças, de acordo com a classificação de Garcia (2004, p.221).

O parágrafo é uma unidade de composição constituída por um ou mais de um período, em que se desenvolve determinada ideia *central*, ou *nuclear*, a que se agregam outras, *secundárias*, intimamente relacionadas pelo sentido e logicamente decorrentes dela. (Grifos do autor).

Uma vez feitas as devidas ressalvas e levando em consideração todas as demais categorias de análise que compõem nosso formulário para preenchimento de ocorrências de fatores ligados ao processamento correferencial e o que permeia o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferenciais na posição de sujeito em textos falados, apresentamos, na seção

a seguir, os resultados encontrados por meio de dados do *corpus* de fala PORCUFORT.

## 5.2 Resultados

Os resultados encontrados a partir da coleta de dados no *corpus* de fala PORCUFORT que serão reportados logo a seguir servem para, na próxima seção, levar a uma discussão embasada em nossa coleta feita por meio de formulários a fim de quantificar a ocorrência dos nomes repetidos, dos nomes recategorizados, dos pronomes plenos e dos pronomes nulos que funcionam como expressão correferente de um antecedente.

Assim como aconteceu com os dados do *corpus* escrito, os gráficos que trazem os resultados dos dados do *corpus* de fala foram computados por meio do preenchimento de formulários em documentos de Word e transpostos para a mesma planilha de Excel que os dados do capítulo anterior, configurando, assim, como os dados de 11 (onze) a 20 (vinte) da planilha, mas nomeados de acordo com o número do inquérito no PORCUFORT, conforme mostra o quadro 20.

**Quadro 20** – Exemplo da planilha do Excel do parâmetro de contabilização dos dados com ou sem efeito de Paralelismo Sintático

Entrevista	Nº	Parâmetro	Paralelismo sintático
Inquerito n° 35	11	Existe	65
		Não Existe	18
Inquerito n° 138	12	Existe	99
		Não Existe	20
Inquerito n° 25	13	Existe	179
		Não Existe	58
Inquerito n° 17	14	Existe	197
		Não Existe	49
Inquerito n° 38	15	Existe	137
		Não Existe	18
Inquerito n° 52	16	Existe	135
		Não Existe	29
Inquerito n° 19	17	Existe	331
		Não Existe	63
Inquerito n° 152	18	Existe	56
		Não Existe	19
Inquerito n° 18	19	Existe	209
		Não Existe	41
Inquerito n° 55	20	Existe	51
		Não Existe	16

Assim como reportado nos resultados dos dados do *corpus* escrito, traremos ao longo desta seção os gráficos com os resultados coletados para

cada categoria de análise dos dados do *corpus* de fala, a saber: (i) quantidade de nomes e de pronomes com função de expressão correferente na posição de sujeito, (ii) quantidade de nomes repetidos, de pronomes plenos e de pronomes nulos com função de expressão correferente na posição de sujeito, (iii) quantidade de nomes repetidos, de nomes recategorizados, de pronomes plenos e de pronomes nulos com função de correferente na posição de sujeito, (iv) quantidade de palavras entre o antecedente e a retomada, (v) ocorrência do efeito de Paralelismo Sintático, (vi) ocorrência de antecedentes em outras posições sintáticas quando não há o efeito de Paralelismo Sintático, (vii) ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações subordinadas, (viii) ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas justapostas, (ix) ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas em períodos distintos, (x) ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como primeira retomada em parágrafo subsequente, (xi) ocorrência do tipo de verbo da sentença anterior à sentença da retomada e (xii) ocorrência de artigo precedendo o termo antecedente.

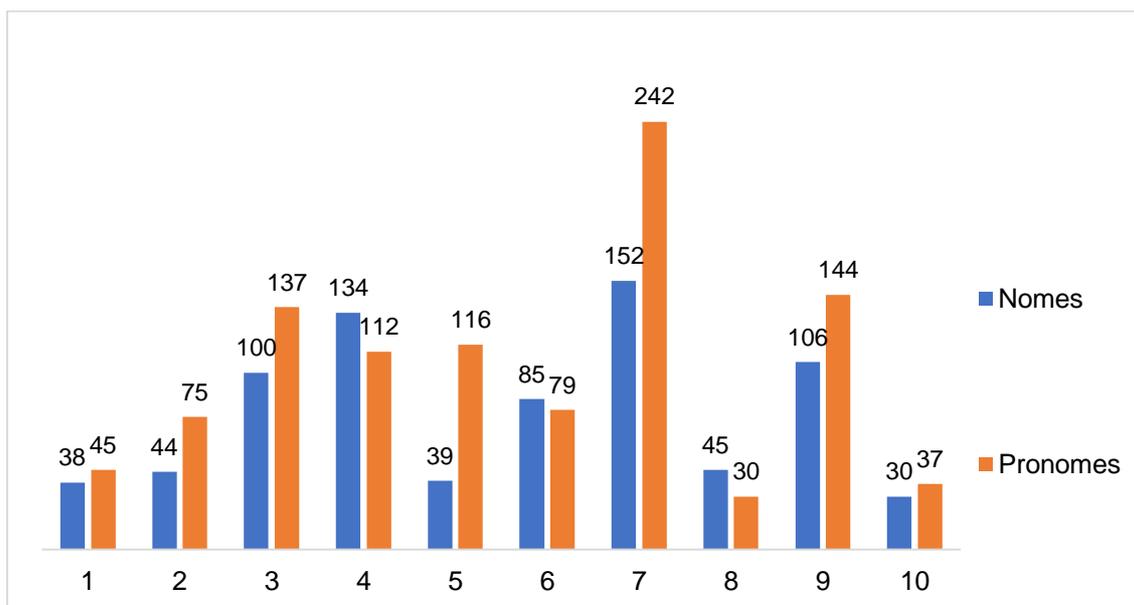
Antes de começar a reportar os gráficos e os respectivos resultados numéricos, gostaríamos ainda de reforçar que todos os dados tiveram os resultados quantificados por meio das mesmas contas matemáticas já explicadas no capítulo anterior e que a divisão das expressões correferenciais seguiram o mesmo padrão. Além disso, para buscar padrões de variação significativa, usamos, assim como fizemos com relação aos dados do *corpus* escrito, a análise estatística usando o *software* PAST, aplicando o Qui-quadrado quando há cruzamento de variáveis e o test-T para comparar formas entre si.

Durante a exposição dos resultados, usaremos exemplos extraídos dos inquéritos analisados, a fim de ilustrar o uso da(s) forma(s) correferente(s) em dado contexto, por ser um resultado bastante específico, como acontece quando levada em consideração a distância em quantidade de palavras entre o antecedente e a retomada e o tipo de sentença ou parágrafo. Por conta disso, optamos por não exemplificar os gráficos que trazem resultados próximos ao que foi encontrado no capítulo anterior, como ocorre nos gráficos que tratam sobre (i) o uso de nomes ou pronomes e suas subcategorias, (ii) paralelismo sintático

e (iii) proeminência sintática, (iv) tipo de verbo da sentença anterior e (v) tipo de artigo do sintagma nominal.

Reportamos, a partir de agora, os resultados, iniciando pela quantidade de ocorrência de nomes e pronomes na posição de sujeito em dados de fala, conforme o gráfico 20.

**Gráfico 20** – Quantidade de nomes e pronomes como expressões correferentes na posição de sujeito em textos do *corpus* PORCUFORT



A fim de condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 1790 retomadas na posição de sujeito, sendo 773 por nomes e 1017 por pronomes, implicando em 43,1% das retomadas efetuadas por nomes e 56,9% por pronomes, sem diferença significativa, sem diferença significativa (p-valor = 0,27).

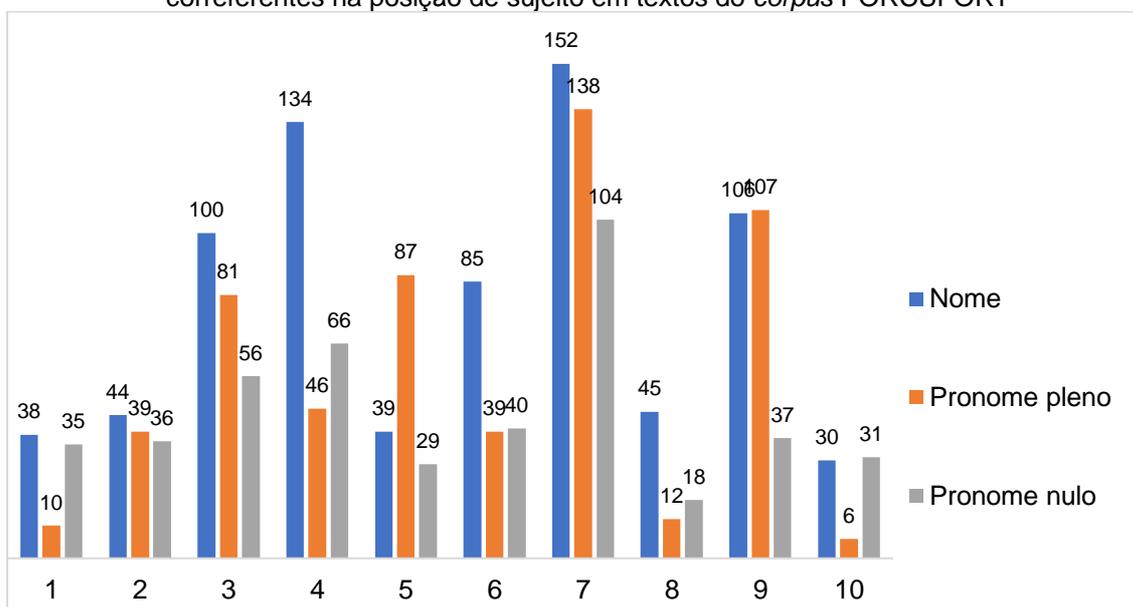
Podemos inferir, a partir dos resultados de ocorrência de uso de nomes e de pronomes na posição de sujeito em dados de fala, que os pronomes despontam como mais frequente do que os nomes, mas só temos indícios de uma possível preferência, sem indício de algum fator que inviabilize o aparecimento frequente de nomes com função correferencial na posição de sujeito em textos falados.

Ao olharmos os resultados individuais encontrados em cada inquérito e considerando a variabilidade de textos analisados por nós, podemos perceber que o uso de nomes e de pronomes com função correferencial na posição de

sujeito é muito variável de pessoa para pessoa, havendo sempre a preferência de uma forma ao invés de outra, mesmo que o uso de cada uma das formas esteja muito próximo da outra.

A fim de compreender melhor como os nomes e os pronomes se comportam em posição de sujeito e de analisar os dados baseados nas pesquisas psicolinguísticas feitas em PB, apresentamos abaixo o gráfico no qual a macrocategoria dos pronomes foi dividida em pronomes plenos e pronomes nulos e a macrocategoria de nomes renomeada para nomes repetidos.

**Gráfico 21** – Quantidade de nomes, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do *corpus* PORCUFORT



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 1790 retomadas na posição de sujeito, sendo 773 por nomes e 1017 por pronomes. Destas retomadas, foram feitas 565 por pronomes plenos e 453 por pronomes nulos, implicando em 43,1% das retomadas efetuadas por nomes, 31,6% por pronomes plenos e 25,3% por pronomes nulos.

Considerando os resultados mostrados no gráfico 21, optamos por destacar os resultados da categoria de pronomes, neste momento, já que os resultados da categoria de nomes não foram subdivididos, estando inalterado como no gráfico 20. Já a categoria de pronomes, por ter sido subdividida entre as duas formas correferentes possíveis – pronomes plenos e pronome nulos – traz resultados interessantes.

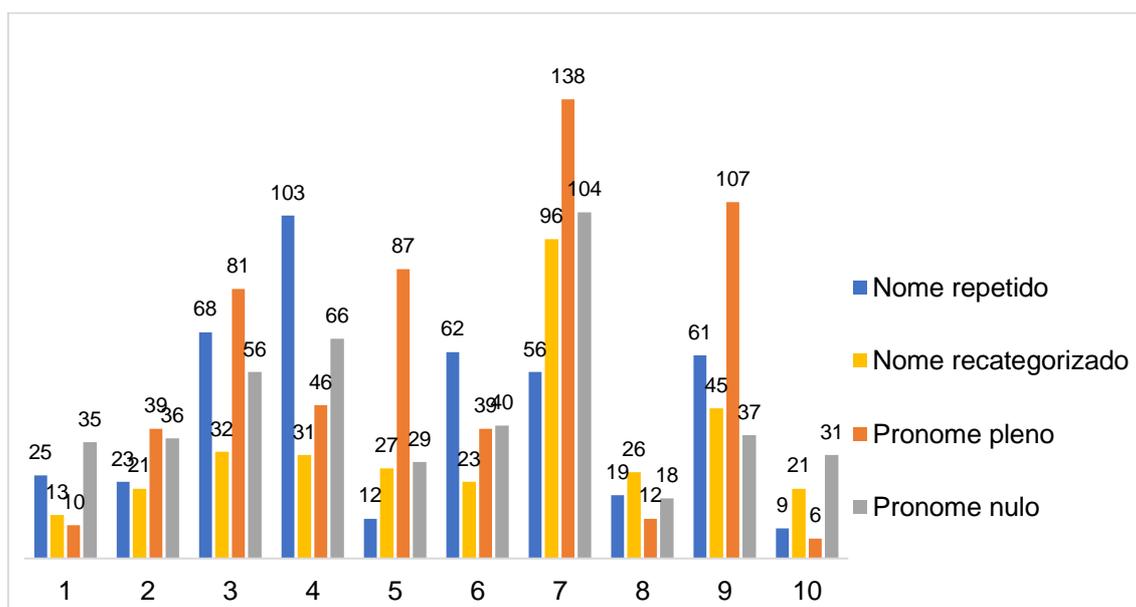
Percebemos, através destes resultados, que os pronomes plenos tendem a retomar mais antecedentes em textos falados, com 55,5% das retomadas por pronomes plenos, do que os pronomes nulos, com 44,5%, o que não aponta para diferença significativa, sem diferença significativa ( $p$ -valor = 0,22).

Temos indícios, portanto de que, em textos falados, de maneira geral, há uma tendência para a menor frequência do pronome nulo, enquanto o pronome pleno surge como a forma pronominal mais usada.

Desta forma, podemos perceber que, quando o texto é falado, os usuário tendem a marcar mais os sujeitos quando optam pelo uso de um pronome, pois se o pronome pleno for mesmo uma opção para o pronome nulo, este resultado aponta para a marcação de um sujeito foneticamente realizado como a forma preferível pelos usuários da língua, sobrepondo, em termos de frequência, o parâmetro do pronome nulo.

Entretanto, assim como a macrocategoria dos pronomes traz indícios de como o usuário faz uso do pronome pleno e do pronome nulo, buscamos evidências de como os nomes são distribuídos ao longo do texto em posição de sujeito. Por conta disto, o gráfico 22 mostra, em quantidade de ocorrências, como estão distribuídos os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos na posição de sujeito em textos falados em PB.

**Gráfico 22** – Quantidade de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes na posição de sujeito em textos do *corpus* PORCUFORT



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 1790 retomadas na posição de sujeito, sendo 773 por nomes e 1017 por pronomes. Das retomadas feitas por nomes, foram feitas 438 por nomes repetidos e 335 por nomes recategorizados, implicando em 24,4% das retomadas efetuadas por nomes repetidos, 18,7% por nomes recategorizados, 31,6% por pronomes plenos e 25,6% por pronomes nulos.

O uso dos nomes repetidos, apesar de serem mais frequentes do que o uso dos nomes recategorizados, não aponta para efeito significativo, pois, de todas as 773 retomadas feitas por meio de nomes, 56,6% foram feitas por nomes repetidos e 43,4% por nomes recategorizados, sem diferenças significativas (p-valor = 0,26). E, o uso dos pronomes plenos, apesar de ser mais frequente, também não aponta para efeitos significativo, pois, quando analisadas as ocorrências totais de pronomes plenos com relação aos nomes repetidos, 56,4% das retomadas foram feitas por pronomes plenos e 43,6% por nomes repetidos, também sem diferença significativa (p-valor = 0,25). Já quando analisadas as ocorrências dos pronomes plenos com relação aos nomes recategorizados, 62,8% das retomadas foram feitas por pronomes plenos e 37,2% por nomes recategorizados, igualmente sem diferença significativa (p-valor = 0,49).

Os gráficos acima mostram a quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito em textos falados do *corpus* PORCUFORT.

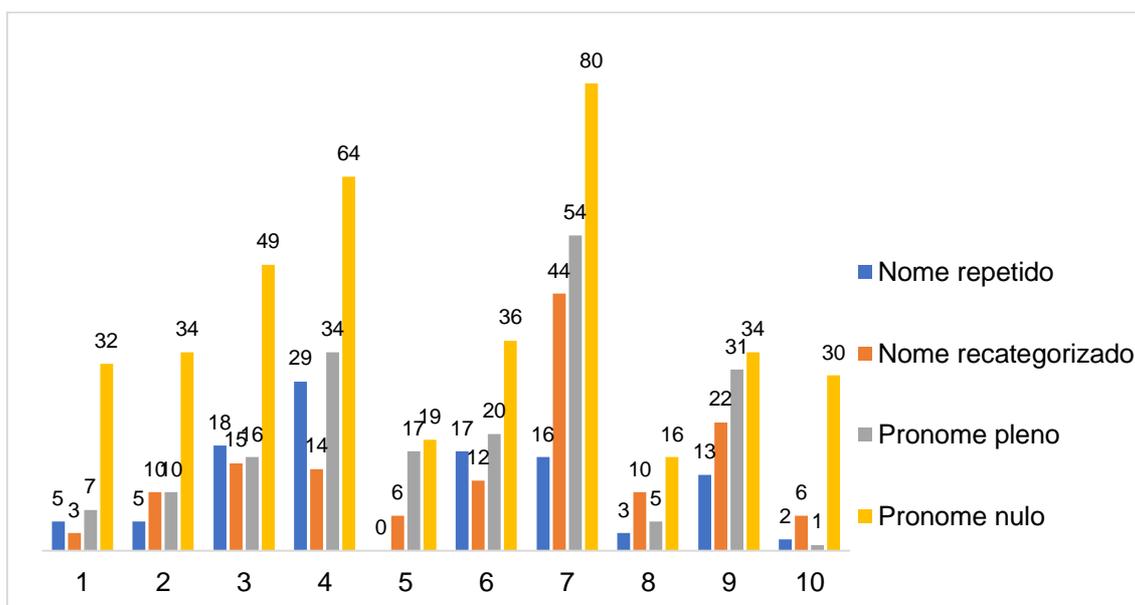
Damos sequência à descrição de nossos resultados seguindo a ordem das categorias de análise de nossos formulários e, por isso, apresentaremos abaixo uma sequência de 04 (quatro) gráficos que reportam a quantidade de palavras entre o antecedente e a retomada.

Para a coleta de dados nos textos falados do *corpus* PORCUFORT, seguimos as mesmas categorias de análise para os dados do *corpus* escrito Lácio-Ref e, por isso, as sequências de quantidade de palavras entre o antecedente e a retomada mostram as formas correferentes distanciadas de seus antecedentes em intervalos de 0 (zero) a 10 (palavras), 11 (onze) a 50 (cinquenta) palavras, 51 (cinquenta e uma) a 100 (cem) palavras e distantes a mais de 100 (cem) palavras.

Retomamos os intervalos de quantidade de palavras entre os antecedentes e as retomadas como forma de reaproximar os valores pesquisados, uma vez que não vimos necessidade de descrever por completo as categorias de análise neste capítulo, conforme explicado anteriormente<sup>211</sup>.

Feitos os comentários pertinentes sobre as categorias de análise que continuam a serem descritas nesta seção, mostramos abaixo os gráficos que trazem os resultados de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito quando há, entre o antecedente a retomada, até 10 (dez) palavras.

**Gráfico 23** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há até 10 palavras entre o antecedente a retomada em textos do *corpus* PORCUFORT



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas no intervalo de 0 (zero) até 10 (de) palavras, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos,

<sup>211</sup> A título de lembrete: optamos por, nos dados do *corpus* de fala, não reportar os resultados para as categorias de análise da quantidade de caracteres sem e com espaço entre o antecedente e a retomada, como foi feito no subcapítulo de resultados dos dados do *corpus* escrito. Decidimos não apresentar estes resultados para os dados do *corpus* de fala porque, conforme descrito na subseção 4.4, por haver caracteres introduzidos no texto transcrito a fim de fazer marcações específicas da fala, a quantidade de caracteres sem e com espaço entre o antecedente e a retomada não seria tão fiel quanto ocorreu na descrição dos dados do *corpus* escrito. Desta forma, deixamos de descrever uma de nossas categorias de análise, dando continuidade a todas as demais.

839 retomadas nesta distância, sendo 108 por nomes repetidos, 142 por nomes recategorizados, 195 por pronomes plenos e 394 por pronomes nulos, implicando em 12,8% por nomes repetidos, 16,9% por nomes recategorizados, 23,2% por pronomes plenos e 47,1% por pronomes nulos.

No exemplo (140)<sup>212</sup>, o referente *suicídio* é retomado duas vezes, servindo como exemplo para as duas formas correferentes mais frequentes quando há, entre o antecedente e a retomada, até 10 (dez) palavras, pois há a primeira retomada pelo pronome pleno, distanciada por 05 (cinco) palavras e, na sequência, pelo pronome nulo, distanciada por 02 (duas) palavras.

(140) [...]... o suicídio<sub>i</sub> é um fenômeno cosmopolita... Ele<sub>i</sub> existiu... E::... Ø<sub>i</sub> EXISTe... [...]  
(inq. 19 homem 41 anos ef)

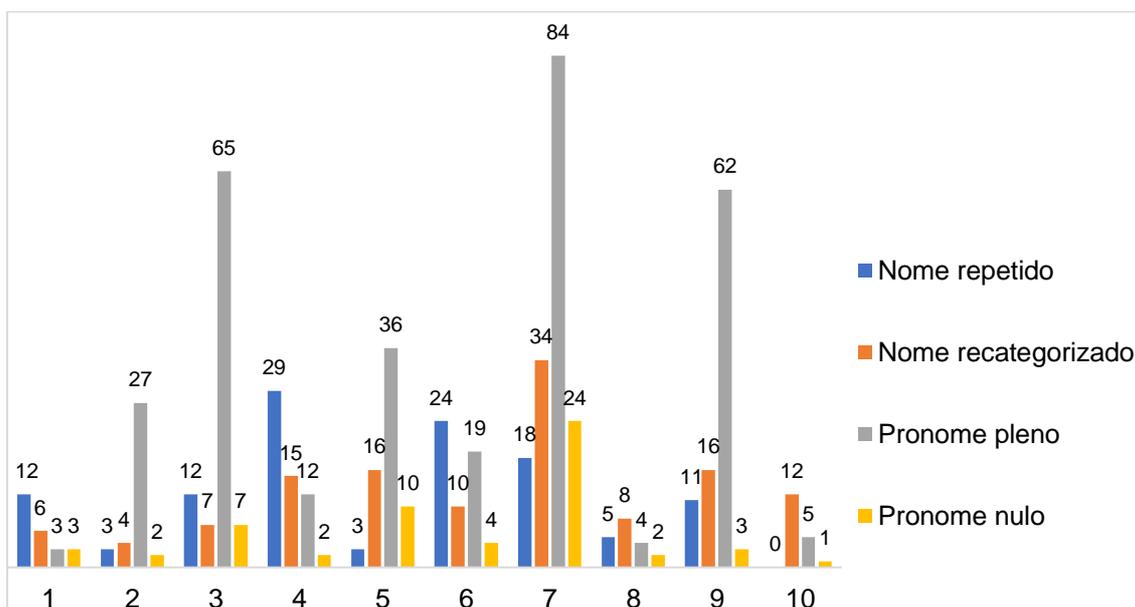
Com estes resultados, inferimos que, quando o antecedente e a expressão correferente estão próximas em uma quantidade de até 10 (dez) palavras, há tendência ao surgimento do pronome nulo, enquanto o nome repetido passa a ser menos frequente. Quando comparadas as formas entre si, tomando o pronome nulo como o mais frequente, encontramos 78,5% das retomadas efetuadas por pronomes nulos e 21,5% por nomes repetidos, sem diferença significativa (p-valor = 0,82); 73,6% das retomadas efetuadas por pronomes nulos e 26,4% por nomes recategorizados, também sem diferença significativa (p-valor = 0,87) e, 66,9% das retomadas efetuadas por pronomes nulos e 33,1% por pronomes plenos, igualmente sem diferença significativa (p-valor = 0,92).

No gráfico 24, veremos os resultados da frequência de uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito em textos falados quando há, entre o antecedente e a retomada, entre 11 (onze) e 50 (cinquenta palavras).

---

<sup>212</sup> Ao longo da exposição dos exemplos extraídos dos dados de fala, optamos por, nos casos em que analisamos questões que envolvem períodos e parágrafos distintos, colocar em nota de rodapé, a versão pontuada. A pontuação simples, isto é, o ponto final que separa períodos será representado por uma barra simples [/]. A separação indicada para mostrar o final de um parágrafo para o começo do seguinte será representada por duas barras [//], a fim de modificar somente o necessário do dado original, seguindo os critérios descritos na seção anterior.

**Gráfico 24** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 11 a 50 palavras entre o antecedente e a retomada em textos do *corpus* PORCUFORT



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas no intervalo de 11(onze) até 50 (cinquenta) palavras, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 620 retomadas nesta distância, sendo 117 por nomes repetidos, 137 por nomes recategorizados, 308 por pronomes plenos e 58 por pronomes nulos, implicando em 18,9% das retomadas efetuadas por nomes repetidos, 22,1% por nomes recategorizados, 49,6% por pronomes plenos e 9,4% por pronomes nulos.

Os resultados do uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito em textos falados mostram que, quando há, entre o antecedente e a retomada, de 11 (onze) a 50 (cinquenta) palavras, o pronome pleno passa a ser a forma mais frequente.

Analisando as retomadas feitas por pronomes plenos com relação às demais formas, encontramos que houve 72,5% das retomadas efetuadas por pronomes plenos e 27,5% por nomes repetidos, sem diferença significativa (p-valor = 0,86), 69,2% das retomadas efetuadas por pronomes plenos e 30,8% por nomes recategorizados, também sem diferença significativa (p-valor = 0,88) e, 84,2% das retomadas efetuadas por pronomes plenos e 15,8% por pronomes nulos, igualmente sem diferença significativa (p-valor = 1,41).

Estes resultados revelam que o uso do pronome pleno passa a ser a forma mais frequente no intervalo de 11 (onze) a 50 (cinquenta) palavras entre o antecedente e a retomada, também sem efeito significativa de uso sobre as outras formas correferentes, além de apontar para a queda no uso do pronome nulo quando a distância entre o antecedente e a retomada aumenta em quantidade de palavras.

Falamos que há queda no uso dos pronomes nulos pelo fato de que, se comparado com o uso dos pronomes plenos, neste contexto, há uma diferença de 68,4% de retomadas efetuadas a mais por pronomes plenos do que por pronomes nulos, apontando para uma frequência muito baixa destes pronomes.

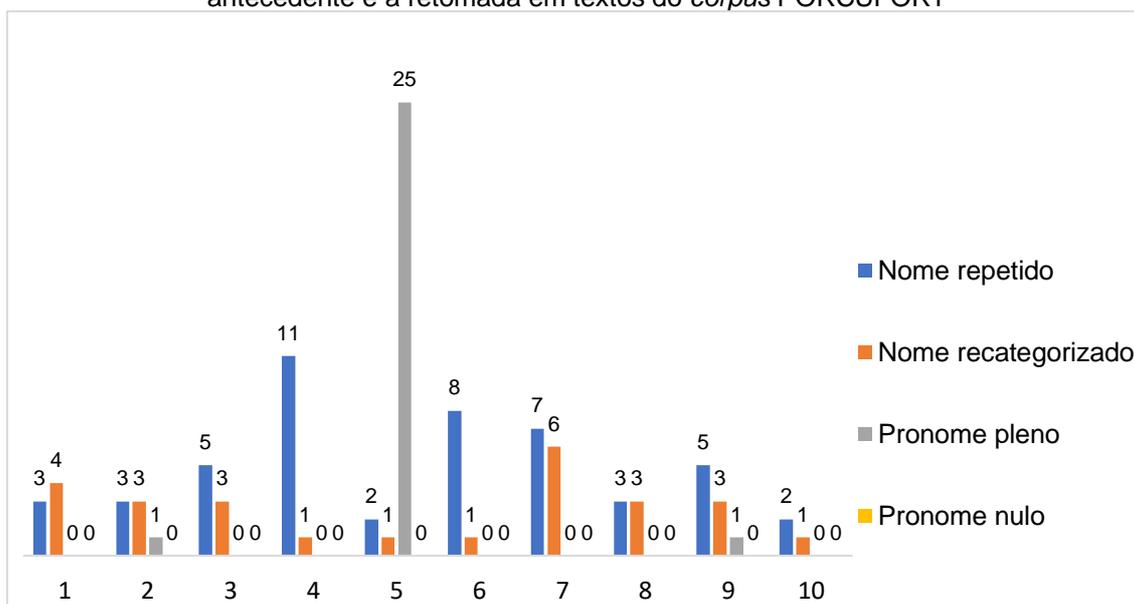
No exemplo (141), o referente *suicídio* é retomado por meio do uso do pronome pleno, quando há entre o antecedente e a retomada 12 (doze) palavras.

(141) o::... o suicídio::: eh::... como um fato histÓrico... que num é:: como:: um fato histórico ele; SOfre então transformações conforme a cultura e o TEMpo...

(inq. 19 homem 41 anos ef)

O gráfico abaixo mostra a quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito de textos falados quando há, entre o antecedente e a retomada, entre 51 (cinquenta e uma) e 100 (cem) palavras.

**Gráfico 25** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há de 51 a 100 palavras entre o antecedente e a retomada em textos do *corpus* PORCUFORT



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas no intervalo de 51 até 100 palavras, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 102 retomadas nesta distância, sendo 49 por nomes repetidos, 26 por nomes recategorizados e 27 por pronomes plenos, implicando em 48% das retomadas efetuadas por nomes repetidos, 25,5% por nomes recategorizados e 26,5% por pronomes plenos.

Os resultados mostrados no gráfico 25 mostram que o uso dos nomes repetidos, no intervalo de 51 (cinquenta e uma) a 100 (cem) palavras passa a ser mais frequente e, parece haver, entre os nomes recategorizados e os pronomes plenos uma competição, pois ambos podem ser usados nesta distância como forma correferente sem danos à compreensão.

É interessante notar que, na língua falada, na distância em que o nome recategorizado é usado como opção para o nome repetido, o pronome pleno também é usado com, aparentemente, a mesma função, o que pode ser um indício de que, na fala, qualquer centro anafórico na posição de sujeito, independente da distância em quantidade de palavras, pode ser preenchido pelo pronome pleno.

Encontramos, neste contexto, 65,3% das retomadas efetuadas por nomes repetidos e 34,7% por nomes recategorizados, sem diferença significativa (p-valor = 0,64). Já nas retomadas entre nomes repetidos e pronomes plenos, encontramos 64,4% das retomadas efetuadas por nomes repetidos e 35,6% por pronomes plenos, também sem diferença significativa (p-valor = 0,61). E, quando comparadas as retomadas feitas entre nomes repetidos e pronomes nulos, encontramos 100% das retomadas efetuadas por nomes repetidos, o que não nos permitiu analisar se há diferença estatisticamente significativa entre estas formas.

No exemplo (141), o referente *arte barroca* é retomado pelo uso do pronome pleno após o uso de 04 (quatro) palavras e, posteriormente, pelo uso do nome recategorizado *barroco* quando há, entre o antecedente e a retomada, 55 (cinquenta e cinco) palavras.

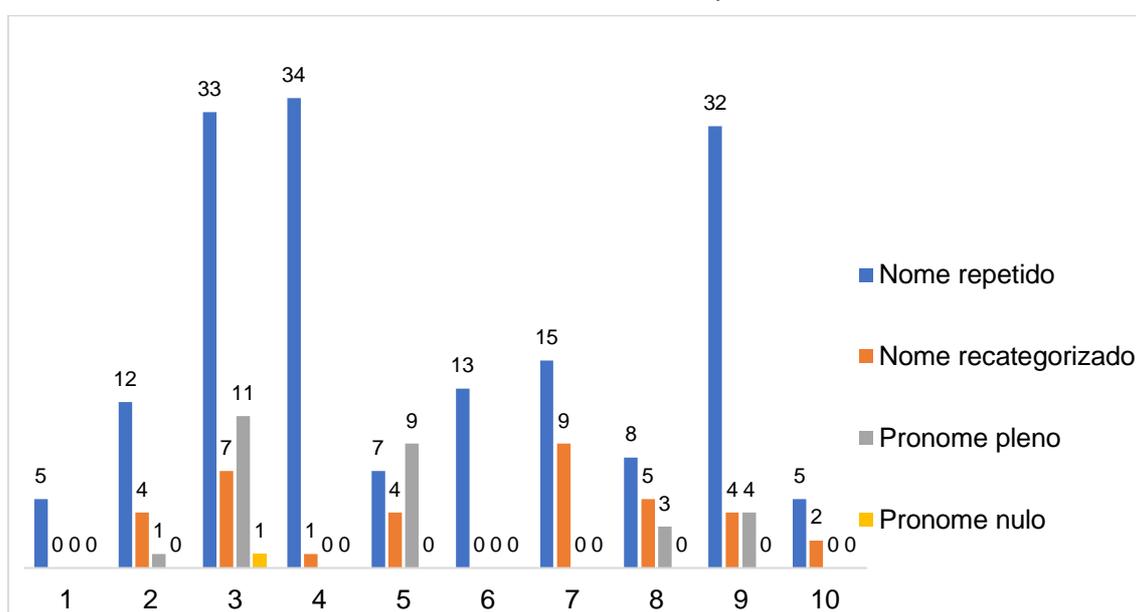
(141) a arte baRROca:... que exprime essa contradição... ela vai... oscilar da mesma forma entre o clássico... /tá? ou seja o PAgão... e o medieval... né? que vai se apresentar:: na figura do cristão... /tá?... daí porque:: a questão de se colocar como uma arte da indisciPLIna... se nós

compaRARmos aos dois movimentos que integram a era CLÁssica... ou seja o ClassiCISmo e o Arcadismo...o Barroco representa um desvio da orientação clássica...

(inq. 35 mulher 30 anos ef)

O gráfico 26 mostra a quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferente em posição de sujeito de textos falados quando há, entre o antecedente e a retomada, mais de 100 (cem) palavras.

**Gráfico 26** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos como expressões correferentes quando há mais de 100 palavras entre o antecedente a retomada em textos do *corpus* PORCUFORT



Ao condensar os resultados acima descritos, optamos por somar todas as ocorrências de retomadas encontradas quando há mais de 100 palavras entre o antecedente e a retomada, tendo sido encontradas, na soma de ocorrência dos 10 (dez) textos, 229 retomadas nesta distância, sendo 164 por nomes repetidos, 36 por nomes recategorizados, 28 por pronomes plenos e 01 por pronome nulo, implicando em 71,7% das retomadas efetuadas por nomes repetidos, 15,7% por nomes recategorizados, 12,2% por pronomes plenos e 0,4% por pronomes nulos.

Os resultados acima mostram que, mais uma vez, o uso do nome repetido é mais frequente em longa distância entre o antecedente e a retomada e, por conta disso, analisamos a ocorrência das formas entre si em busca de efeitos estatisticamente significantes, tendo encontrado 82% das retomadas feitas por

nomes repetidos e 18% por nomes recategorizados, sem diferença significativa (p-valor = 0,58), 85,4% das retomadas feitas por nomes repetidos e 14,6% por pronomes plenos, também sem diferença significativa (p-valor = 0,44) e, 99,4% das retomadas feitas por nomes repetidos e 0,6% por pronomes nulos, igualmente sem diferença significativa (p-valor = 0,33).

No exemplo (142), o referente *Marx* é retomado por pronome nulo após o uso de 06 (seis) palavras e, posteriormente, é retomado pela repetição do nome após o uso de 130 (cento e trinta) palavras.

(142) Marx<sub>i</sub> vai se contrapor a esse (conhecimento) Ø<sub>i</sub> vai dizer o que o pensamento ((ruído)) ele é fruto da nova concepção histórica da nova concepção material... é a partir de como nós vivemos de como nós (fazemos) o nosso dinheiro como nós nos organizamos que nós podemos (então) refletir a nossa ação ENTÃO a consciência o pensamento a representação é um erro posterior e no primeiro momento nós temos ESSE momento de contato com a realidade... NÓS quer dizer as pessoas vivas distintas na sua (consecutude) na sua materialidade essas pessoas na sua materialidade vivENDO trava-LHANDo movendo relações sociais né? ((funga)) eh eh dis/ distintas nas diversas classes sociais essas pessoas é que formam a consciência então a consciência é um elemento secundário com cer/ eh:: considerando que o primeiro elemento secundário ( )... elemento secundário de ser é:: não prioritário... ENTÃO Marx<sub>i</sub> é a existência que determina a consciência (então) eu só posso pensar em cima de coisas que existe e não em coisas que eu imagino que SÃO como predizia a filosofia alemã hegeliana não é a imagem de uma única imagem que seja mas é o homem real concreto (né?)

(inq. 19 homem 41 anos ef)

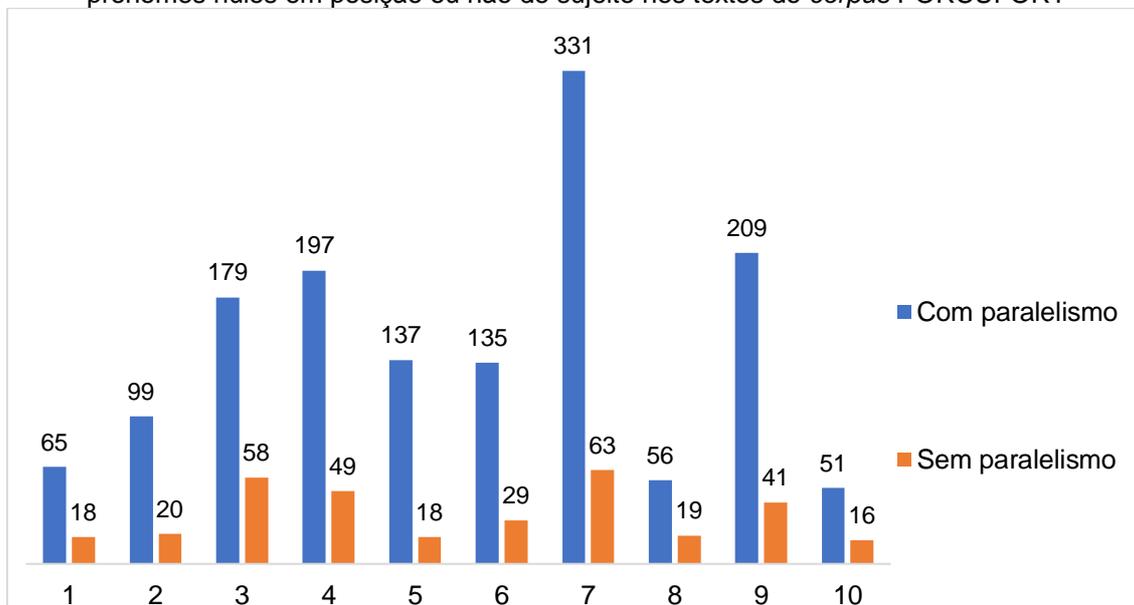
Uma vez que findamos a exposição dos resultados com relação à quantidade de palavras entre o antecedente e a retomada, reforçamos que, diferentemente dos resultados do *corpus* escrito, não apresentaremos os resultados da quantidade de caracteres sem e com espaço entre o antecedente e a retomada por acreditarmos que os resultados não seriam tão fiéis como reportados no capítulo anterior.

A fim de analisar uma possível influência da distância no tipo de forma correferente usada nos textos de fala, aplicamos o Qui-quadrado para que fossem analisadas todas as distâncias, sendo apontado pelo *software* PAST, efeito significativo no cruzamento das duas variáveis (p-valor = 0,0001). Este resultado nos leva a interpretar que pode haver efeito direto entre a distância e o tipo de retomada usada pelos falantes.

Dando prosseguimento à descrição de nossos resultados, mostramos abaixo o gráfico com a quantidade de ocorrências de nomes repetidos, nomes

recategorizados, pronomes plenos e pronomes nulos na posição de sujeito, observando se o antecedente estava em posição também de sujeito ou não a fim de averiguar se há ou não o efeito de paralelismo sintático em textos falados em PB.

**Gráfico 27** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em posição ou não de sujeito nos textos do *corpus* PORCUFORT

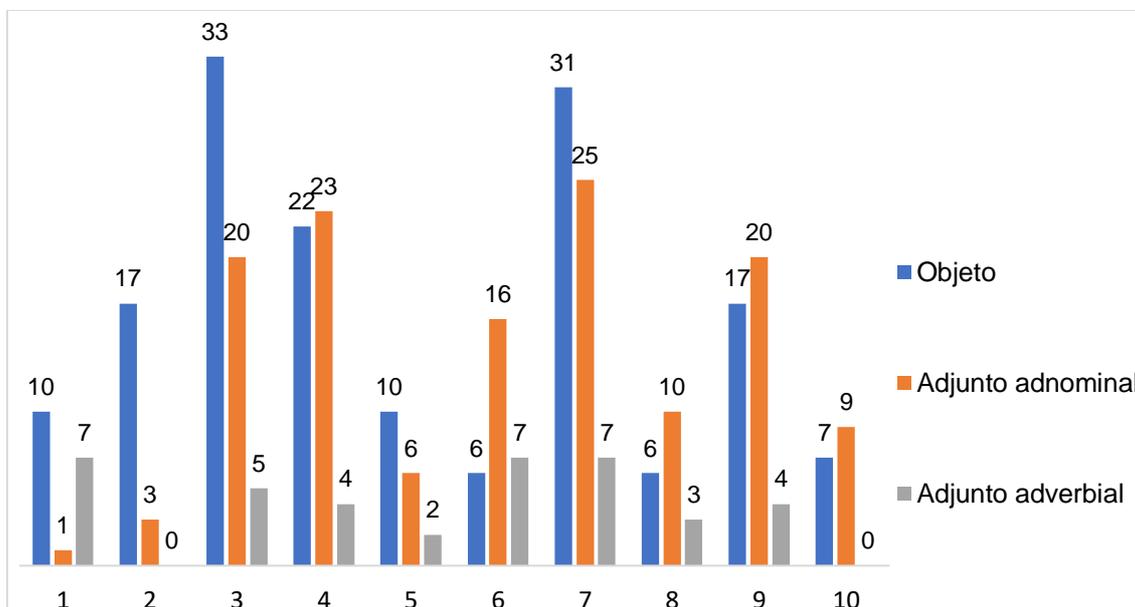


Como o efeito de paralelismo sintático trata da posição do antecedente com relação à retomada, observamos se o antecedente era retomado na posição de sujeito ou não, havendo o efeito de paralelismo quando o antecedente é retomado por uma expressão correferencial estando ambos em posição de sujeito e, consideramos sem paralelismo quando o antecedente está em outra posição que não seja a de sujeito da sentença.

Somando todas as ocorrências contabilizadas, a fim de averiguar o efeito de paralelismo sintático, temos 1790 retomadas em posição de sujeito, sendo 1459 retomando um antecedente também em posição de sujeito e 331 retomando antecedente em outra posição sintática, implicando em 81,5% de ocorrências sujeito/sujeito e 18,5% outras posições sintáticas/sujeito, sem diferença significativa ( $p$ -valor = 1,06). Com estes resultados, temos indícios de que há grande tendência para o termo antecedente em posição de sujeito ser retomado por uma expressão correferencial também em posição de sujeito, indicando uma possível influência do efeito de Paralelismo Sintático em textos falados em PB.

O gráfico 28 mostra a ocorrência das posições sintáticas mais frequentes em que o antecedente ocorre quando não está em posição sujeito.

**Gráfico 28** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em posições sintáticas que não de sujeito nos textos do *corpus* PORCUFORT



O gráfico 28 mostra as entradas de antecedentes que não estavam em posição de sujeito, tendo sido recorrente o fato de os antecedentes ocorrerem em posição de objeto, de adjunto adnominal ou adjunto adverbial<sup>213</sup>.

Somando todas as ocorrências contabilizadas, a fim de averiguar a frequência de em quais posições sintáticas os referentes tendem a ocorrer quando não há efeito de paralelismo sintático, temos 331 retomadas em posição sintática que não de sujeito, sendo 159 retomadas de um antecedente em posição de objeto, 133 em posição de adjunto adnominal e 39 em posição de adjunto adverbial, implicando em 48% de ocorrências objeto/sujeito, sem diferença significativa, também sem diferença significativa (p-valor = 1,05), 40,1% adjunto adnominal/sujeito, igualmente sem diferença significativa (p-valor = 0,49) e 11,9% adjunto adverbial/sujeito (p-valor = 0,17). Esses valores de *p* implicam que as diferenças que as diferenças de ocorrência nas posições de objeto/sujeito, adjunto adnominal/sujeito e adjunto adverbial/sujeito não são significativas.

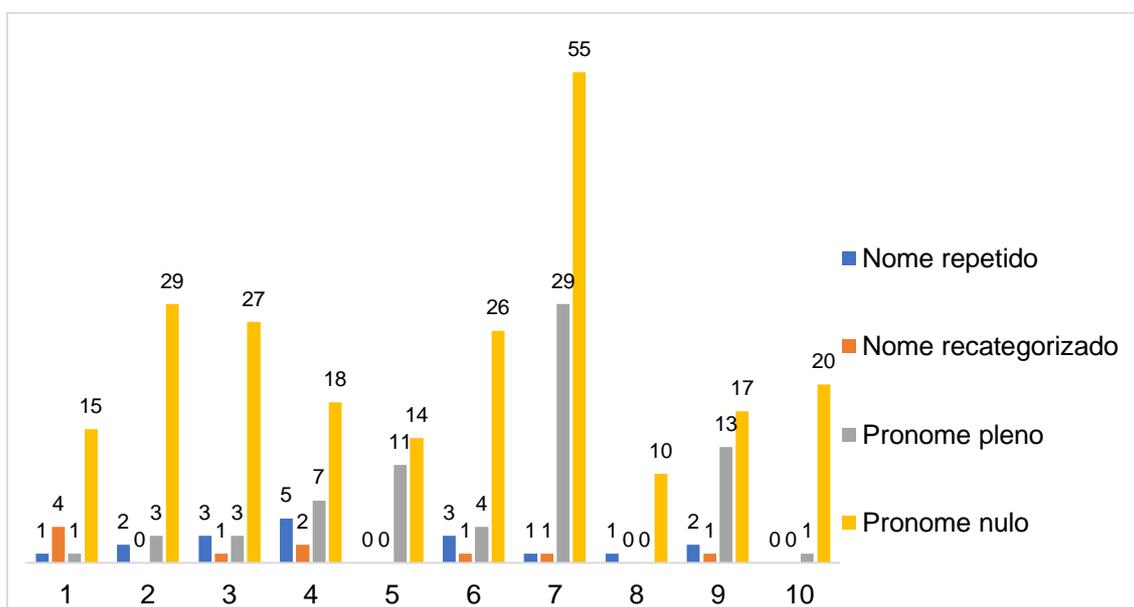
<sup>213</sup> As ressalvas feitas para a categoria de análise de ocorrências dos antecedentes quando em outras posições sintáticas que não fossem a de sujeito foram mantidas sem alteração ao que foi reportado no capítulo anterior para os resultados dos dados do *corpus* escrito.

Com os resultados acima, podemos notar que, em textos falados, é viável que um antecedente que não esteja em posição de sujeito seja retomado por nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial em posição de sujeito, havendo a tendência de que, quando o antecedente não estiver em posição de sujeito, esteja em posição de objeto, de adjunto adnominal ou de adjunto adverbial.

Passamos agora para os resultados dos dados sobre o tipo de sentença na qual a retomada é efetuada, levando em consideração que a ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos foi observada quanto ao tipo de sentença na qual ocorre, sendo em (i) orações subordinadas, (ii) orações coordenadas justapostas<sup>214</sup>, (iii) orações coordenadas em períodos distintos<sup>215</sup> e (iv) em parágrafo subsequente<sup>216</sup>.

O gráfico abaixo mostra a ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito de orações subordinadas.

**Gráfico 29** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações subordinadas nos textos do *corpus* PORCUFORT



<sup>214</sup> São orações coordenadas dispostas uma ao lado da outra estando ligadas por vírgulas e/ou conjunções.

<sup>215</sup> São orações coordenadas que, apesar de estarem uma ao lado da outra, são separadas por ponto final.

<sup>216</sup> Levamos em consideração, nesta subcategoria de análise, como é feita a primeira retomada em parágrafo subsequente.

Somando as quantidades totais de retomadas em orações subordinadas dos 10 (dez) textos, houve 331 retomadas, sendo 18 por nomes repetidos, 10 por nomes recategorizados, 72 por pronomes plenos e 231 por pronomes nulos, implicando em 5,4% das retomadas efetuadas por nomes repetidos, 3,1% por nomes recategorizados, 21,7% por pronomes plenos e 69,8% por pronomes nulos.

Ao compararmos as formas entre si, observamos que o pronome nulo é a forma mais frequente e, assim, comparamos a frequência deste com as demais formas, tendo ocorrido, portanto, 92,7% das retomadas efetuadas por pronomes nulos e 7,3% por nomes repetidos, sem diferença significativa (p-valor = 0,63), 95,9% por pronomes nulos e 4,1% por nomes recategorizados, também sem diferença significativa (p-valor = 0,61) e, por fim, 76,3% por pronomes nulos e 23,7% por pronomes plenos, igualmente sem diferença significativa (p-valor = 0,79).

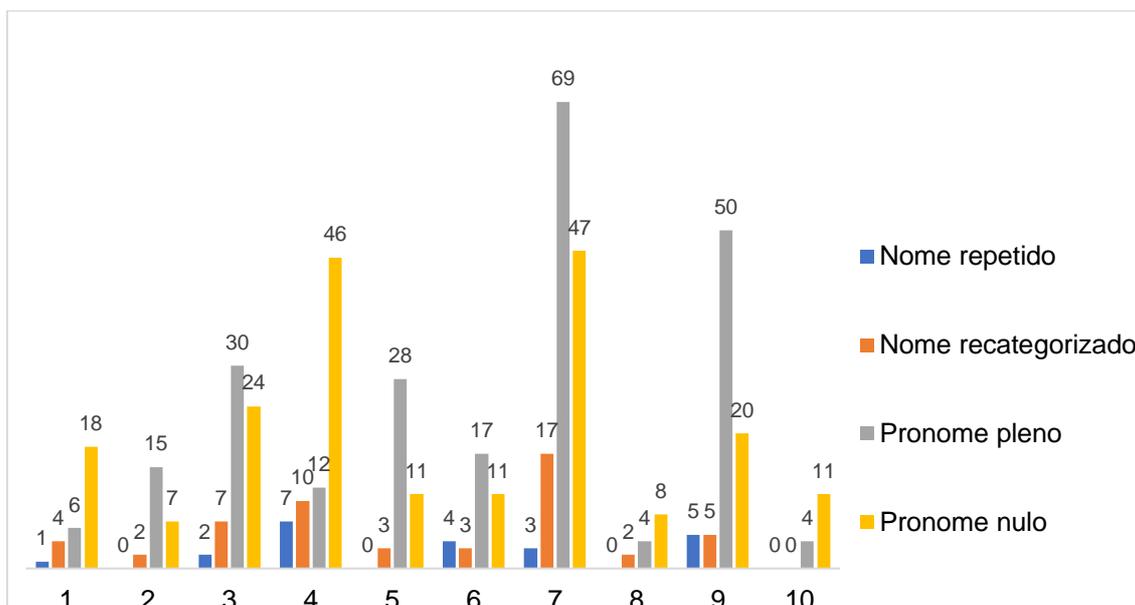
No exemplo (143), o referente *pidgin* é retomado na posição de sujeito da oração subordinada por meio do pronome nulo em duas ocorrências seguidas.

(143) as características do pidgini... quanto às características do pidgini... primeiramente nós já vimos que s/... que  $\emptyset_i$  não seria uma... língua materna né? que  $\emptyset_i$  seria uma segunda língua... certo?...

(inq. 138 mulher 31 anos ef)

O gráfico abaixo mostra os resultados da ocorrência das expressões anafóricas na posição de sujeito de orações subordinadas justapostas, ou seja, orações subordinadas ligadas entre si por vírgulas e/ou conjunções, mas estando no mesmo período.

**Gráfico 30** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas justapostas nos textos do *corpus* PORCUFORT



Somando as quantidades totais de retomadas em orações coordenadas justapostas dos 10 (dez) textos, houve 513 retomadas, sendo 22 por nomes repetidos, 53 por nomes recategorizados, 235 por pronomes plenos e 203 por pronomes nulos, implicando em 4,2% das retomadas efetuadas por nomes repetidos, 10,3% por nomes recategorizados, 45,8% por pronomes plenos e 39,7% por pronomes nulos.

Ao compararmos as formas entre si, observamos que o pronome pleno passa a ser a forma mais frequente e, assim, encontramos 90,3% das retomadas efetuadas por pronomes plenos e 9,7% por nomes repetidos, sem diferença significativa ( $p$ -valor = 0,78), 81,6% por pronomes plenos e 18,4% por nomes recategorizados, também sem diferença significativa ( $p$ -valor = 0,84) e, por fim, 51,9% por pronomes plenos e 48,1% por pronomes nulos, também sem diferença significativa ( $p$ -valor = 0,20).

No exemplo (144), o referente *ideias* reforçado sequencialmente pelo pronome pleno *elas* que é retomado novamente usado na oração coordenada em período distinto e, posteriormente, preenchendo o sujeito da oração coordenada justaposta, ilustrando os resultados encontrados no gráfico 30.

(144) [...] como é que as iDEias; elas elas; podem ser des::coLAdas da sua base material [/] elas; partem da base material mas elas; ao longo do caminho [...]

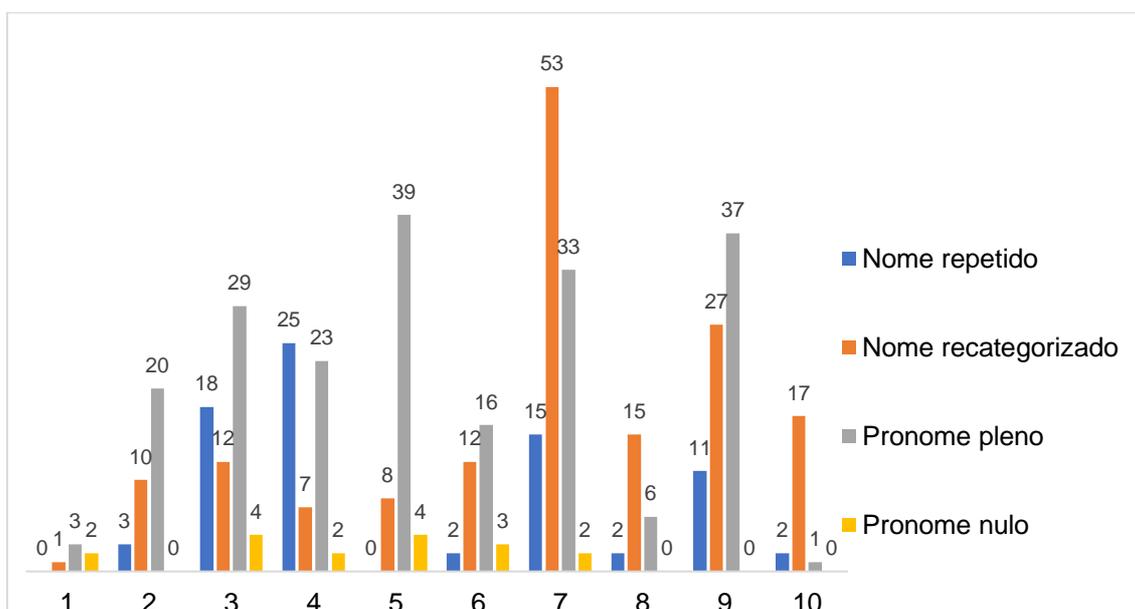
(inq. 18 mulher 42 anos ef)

Os resultados acima mostram que em orações coordenadas justapostas, os usuários do PB, na fala, fazem preferencialmente a retomada por meio dos pronomes plenos, mas o uso dos pronomes nulos, neste contexto, parece ser uma opção viável, já que os resultados são muito próximos em termos de frequência. Já os nomes recategorizados, apesar de mais frequentes do que os nomes repetidos, são menos usados do que os pronomes, o que indica a preferência do uso de pronomes para preencher sujeitos de orações coordenadas justapostas.

No gráfico 31, veremos os resultados da ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito de orações coordenadas em períodos distintos.

**Gráfico 31** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em orações coordenadas em períodos distintos nos textos do *corpus*

PORCUFORT



Somando as quantidades totais de retomadas em orações coordenadas em períodos distintos dos 10 (dez) textos, houve 464 retomadas, sendo 78 por nomes repetidos, 162 por nomes recategorizados, 207 por pronomes plenos e 17 por pronomes nulos, implicando em 16,8% das retomadas efetuadas por nomes repetidos, 34,9% por nomes recategorizados, 44,6% por pronomes plenos e 3,7% por pronomes nulos.

Ao compararmos as formas entre si, observamos que o pronome pleno continua a ser, neste contexto, a forma preferencialmente usada, sendo seguida pelo nome recategorizado, o que, mais uma vez dá indício de como a distância em quantidade de palavras entre o antecedente e a retomada pode ter influência sobre o uso da expressão correferente na posição de sujeito em textos escritos. Encontramos 72,7% das retomadas efetuadas por pronomes plenos e 27,3% por nomes repetidos, sem diferença significativa (p-valor = 0,61), 56,1% por pronomes plenos e 43,9% por nomes recategorizados, também sem diferença significativa (p-valor = 0,93) e, por fim, 92,4% por nomes recategorizados e 7,6% por pronomes nulos, igualmente sem diferença significativa (p-valor = 0,44).

No exemplo (145), referente *novo governante* é retomado na posição de sujeito da oração coordenada em período distinto por meio do uso do pronome pleno, conforme esperado ao levar em consideração os resultados do gráfico 31.

(145) E Esse NOvo governante<sub>i</sub> vai modificar tudo...[/] ele<sub>i</sub> deCREta então que:: é proibido o que ANtes era permitido e vice-versa...

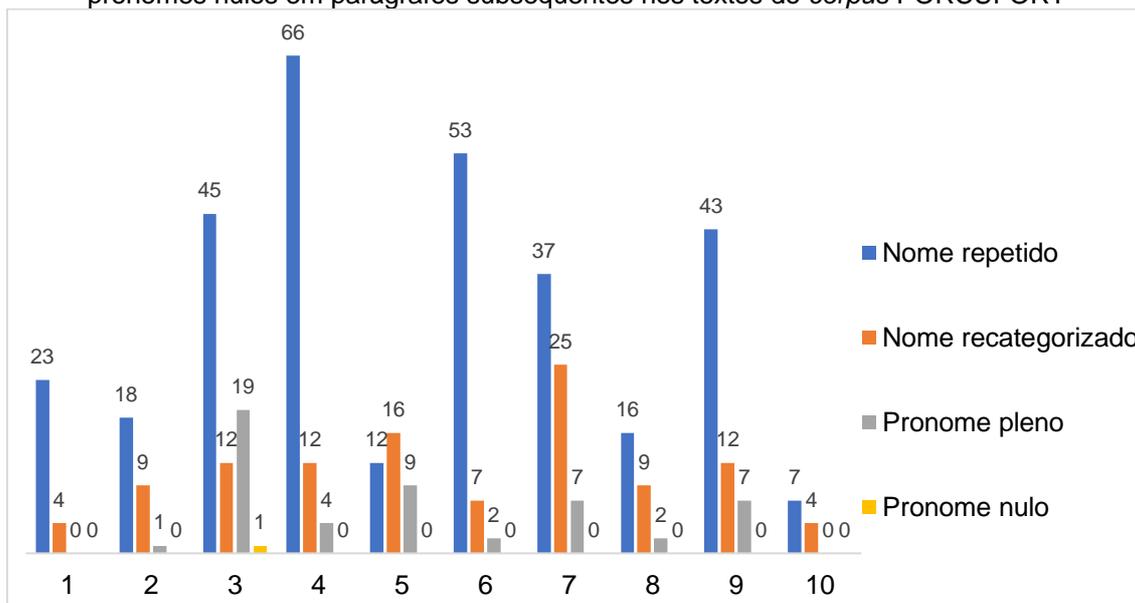
(inq. 35 mulher 30 anos ef)

Inferimos, por meio destes resultados que, assim como cogitamos anteriormente, o pronome pleno e o nome recategorizado parecem funcionar como opções que o usuário da língua tem a sua disposição para retomar um antecedente quando, em quantidade mediana de palavras entre o antecedente e a retomada, o contexto não é tão próximo que faça emergir o pronome nulo nem é tão distante que propicie o aparecimento do nome repetido.

Outra questão importante que inferimos a partir destes resultados é que o uso do pronome pleno pode ser tão frequente em posição de sujeito de orações coordenadas de períodos distintos a ponto de fazer com que o uso do pronome nulo quase não surja neste contexto.

Mostramos, a seguir, o gráfico com a quantidade de ocorrências das expressões anafóricas na posição de sujeito quando a retomada ocorre em parágrafo subsequente.

**Gráfico 32** – Ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em parágrafos subsequentes nos textos do *corpus* PORCUFORT



Somando as quantidades totais de retomadas que funcionam como a primeira retomada em parágrafo subsequente dos 10 (dez) textos, houve 482 retomadas, sendo 320 por nomes repetidos, 110 por nomes recategorizados, 51 por pronome pleno e 01 por pronome nulo, implicando em 66,4% das retomadas efetuadas por nomes repetidos, 22,8% por nomes recategorizados, 10,5% por pronomes plenos e 0,3% por pronomes nulos.

Ao compararmos a frequência das expressões correferenciais entre si, encontramos 74,4% das retomadas efetuadas por nomes repetidos e 25,6% por nomes recategorizados, sem diferença significativa ( $p$ -valor = 0,63), 86,2% por nomes repetidos e 13,8% por pronomes plenos, também sem diferença significativa ( $p$ -valor = 0,34) e, por fim, 99,6% por nomes repetidos e 0,4% por pronomes nulos, igualmente sem diferença significativa ( $p$ -valor = 0,29).

No exemplo (145) o referente *espigões* é recategorizar para esses espigões que surge pela repetição deste nome na primeira vez em que o antecedente é retomado na posição de sujeito de parágrafo subsequente.

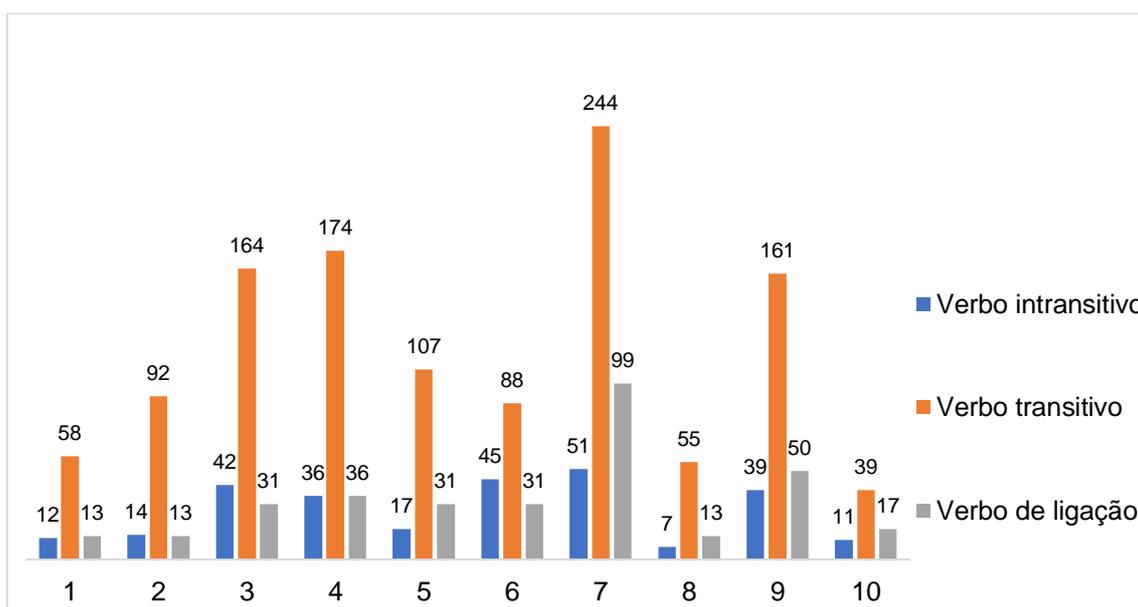
(145) então Uma das causa/ tem uma outra causa aí mais complexa... é está relacionada com os espigÕES; esses e/ espigões;... construído aqui na::... aqui em Fortaleza na nas PRAIA/ pra... con/ conter a... a maRÉ também a:: as ONda/ as altas ONda/ né?...[//] então esses espigÕES a... DIZEM os especialistas... que esses espigÕES... MODifiCaram a aÇÃO das correntes marí/... marítimas... né?...

(inq. 52 homem 39 anos ef)

A fim de encontrar possíveis efeitos estatisticamente significativos, cruzamos, no software PAST, o tipo de retomada com a disposição textual, tendo encontrado efeito significativo ( $p$ -valor = 0,0001), o que pode ser interpretado como um forte indício da influência que o tipo de oração ou o fato de estar em parágrafo seguinte tem sobre a escolha da forma correferente a ser usada na posição de sujeito em textos falados em PB.

Mostramos abaixo a frequência do tipo de verbo da sentença anterior à sentença da retomada.

**Gráfico 33** – Ocorrência verbos intransitivos, verbos transitivos e verbos de ligação nos textos do corpus PORCUFORT



Para descrever a quantidade de verbos que ocorreram na sentença anterior à sentença da retomada, retornamos à classificação que propomos. Conforme explicado no capítulo anterior, quantificamos os verbos em três tipos: (i) verbos intransitivos, (ii) verbos transitivos, englobando os verbos transitivos diretos, os verbos transitivos indiretos e os verbos bitransitivos e, (iii) verbos de ligação.

Durante a descrição dos resultados mostrados abaixo, para (ii) há os devidos comentários sobre a quantidade de ocorrências de verbos transitivos diretos, transitivos indiretos e verbos bitransitivos, o que foi feito, a fim de facilitar a leitura dos dados.

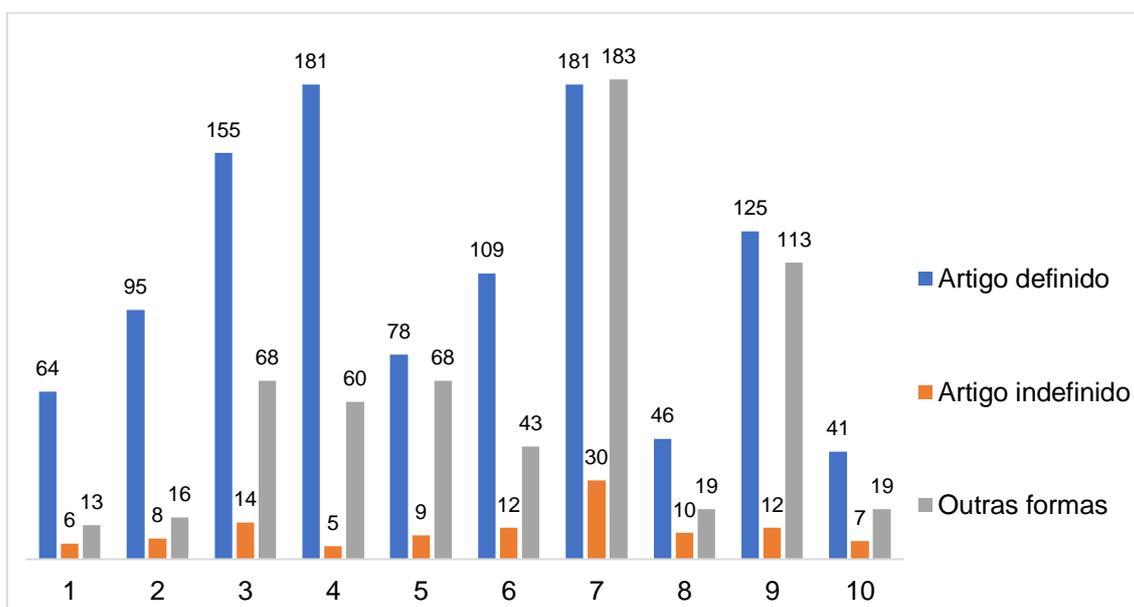
A soma da ocorrência da quantidade de verbos na sentença anterior à retomada mostrou 1790 ocorrências, sendo 274 por verbos intransitivos, 1182

por verbos transitivos (1115 transitivos diretos, 61 transitivos indiretos e 06 bitransitivos) e 334 por verbos de ligação, implicando em 15,3% das sentenças anteriores à retomada com verbos intransitivos, sem diferença significativa (p-valor = 0,37), 66% com verbos transitivos (94,3% transitivos diretos, 5,1% transitivos indiretos e 0,6% bitransitivos), sem diferença significativa (p-valor = 0,66) e 18,7% com verbos de ligação, também sem diferença significativa (p-valor = 0,22).

Os resultados acima mostram que, quando observado o tipo de verbo da sentença anterior à sentença na qual ocorre a retomada, os verbos transitivos diretos são mais frequentes na fala, sendo seguidos pelo uso dos verbos de ligação e, menos frequentemente, pelo uso de verbos intransitivos.

Mostramos, abaixo, os resultados de nossa última categoria de análise nos dados de *corpus* de fala sobre o tipo de artigo que precede o antecedente.

**Gráfico 34** – Ocorrência artigos definidos, indefinidos ou outras formas de definitude precedendo o antecedente nos textos do *corpus* PORCUFORT



Na contabilização do tipo de artigo que precede o antecedente, seguimos o mesmo que fora proposto para o *corpus* escrito e que está explicado no capítulo anterior, isto é, observamos somente se havia (i) artigo definido ou (ii) artigo indefinido diante do antecedente, ficando todas as outras possibilidades de precedente quantificadas em (iii) outras formas de (in)definitude.

Com a soma dos valores da quantidade de antecedentes precedidos por artigos definidos, artigos indefinidos ou outras formas nos 10 (dez) textos temos

1790 ocorrências, sendo 1075 dos antecedentes precedidos por artigos definidos, sem diferença significativa ( $p$ -valor = 0,46), 113 por artigos indefinidos e 602 por outras formas, implicando 60% dos antecedentes precedidos por artigos definidos, também sem diferença significativa ( $p$ -valor = 0,44), 6,4% por artigos indefinidos e 33,6% por outras formas, igualmente sem diferença significativa ( $p$ -valor = 0,59).

Os resultados acima mostram que há uma tendência de o antecedente ser bem definido no texto ao ser, na maioria, precedido por artigo definido, enquanto os antecedentes precedidos por artigo indefinido têm baixa quantidade de ocorrência, o que mostra a necessidade de definir bem o termo que será retomada ao longo do texto na fala.

Na busca de indícios de como estas expressões correferenciais são distribuídas em português brasileiro, buscamos elencar padrões já analisados em pesquisas psicolinguísticas sobre esta área como, por exemplo, o fato de sujeito retomar preferencialmente sujeito, o tipo de verbo da sentença anterior à sentença na qual há a retomada, a presença ou não de artigo definido e a distância entre o antecedente em quantidade de palavras. Aproveitamos para reiterar que não mostramos os resultados da quantidade de caracteres sem e com espaço entre o antecedente e a retomada pelo fato de acreditarmos que estes dados não seriam tão fiéis quanto ocorre em dados escritos, conforme justificamos anteriormente neste capítulo.

Depois de analisarmos 1790 ocorrências de retomadas feitas por nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito, temos dados robustos de como os fatores analisados parecem influenciar no tipo de expressão correferencial usada na retomada.

Desta forma, esperamos que nossas interpretações levem, a partir dos dados autênticos da língua falada, a uma discussão ampla e esclarecedora sobre fatores que podem estar envolvidos na resolução da correferência anafórica e, assim, contribuir com esta área de pesquisa.

Tendo à nossa disposição os resultados acima e com interpretações depreendidas por meio e a partir das análises feitas, empreendemos na próxima seção uma discussão com achados anteriores.

### 5.3 Discussão

A análise dos resultados encontrados a partir da contabilização da ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito de textos de *corpus* de fala nos levou a interpretações que corroboram algumas pesquisas e outras não das que foram citadas no capítulo anterior.

Tentando buscar o máximo de correlação com cada uma destas pesquisas, buscaremos detalhar em quais aspectos concordamos ou não com o que tem sido feito ao longo dos últimos anos em pesquisas psicolinguísticas sobre a resolução correferencial.

Aproveitamos para, novamente, advertir para o fato de que nossa pesquisa é documental e, portanto, descritiva, não contendo os mesmos procedimentos que as pesquisas laboratoriais, mas que nos dá, por meio de resultados de ocorrência/frequência, indícios do que de fato acontece quando a língua é usada em contextos reais.

Tratamos neste trabalho sobre a correferência anafórica e buscamos explicações de como estão distribuídos os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos com função correferencial na posição de sujeito e, mais especificamente, neste capítulo, debruçamo-nos em fatores que emergem quando a língua é falada.

Seguindo o desenvolvimento da discussão feita no capítulo anterior e tentando padronizar ao máximo nosso texto, seguiremos a exposição de nossa discussão a partir dos pontos norteadores de nosso trabalho, a saber: (i) teorias e hipóteses sintáticas sobre o processamento anafórico, (ii) teorias e hipóteses semânticas sobre o processamento anafórico e (iii) teoria pragmática sobre o processamento anafórico, interligando estas teorias a nossos resultados sobre o processamento anafórico em PB.

Sobre a categoria vazia *pro*, a qual estudamos nesta tese sob o nome de pronomes nulo (a parte oposto ao pronome foneticamente realizado) que tem as características -anafórico/+pronominal, encontramos um resultado, no mínimo interessante, no que concerne à questão de ser o Português Brasileiro, no estágio atual, uma língua parcial *pro-drop* (HOLMBERG; NAYADU; SHEEHAN, 2009).

Os resultados que encontramos a partir da quantidade de ocorrências de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito de textos falados indica que, quando levada em consideração a fala, o pronome pleno é preferencialmente usado.

Das 1790 ocorrência de nomes e pronomes funcionando como expressão correferencial na posição de sujeito em textos do *corpus* de fala, 1017 foram retomadas feitas somente com uso de pronomes e, destas retomadas, 565 foram feitas por pronomes plenos e 452 por pronomes nulos, implicando em 55,5% das retomadas efetuadas por pronomes plenos, sem diferença significativa ( $p$ -valor = 0,58) e 44,5% por pronomes nulos, também sem diferença significativa ( $p$ -valor = 0,54).

Com estes resultados, encontramos indícios de que, em textos falados, há maior frequência de pronomes plenos, o que poderia apontar para possível variação do parâmetro *pro-drop* em PB. Entretanto, é preciso frisar que esta diferença não é estatisticamente significativa a ponto de excluir o uso dos pronomes nulos, o que apontaria de fato para a perda deste parâmetro, tornando o Português Brasileiro uma língua não *pro-drop*.

Por acreditarmos que o usuário da língua tem a sua disposição opções de expressões correferenciais, entendemos o pronome pleno como uma opção para o pronome nulo. Nossos resultados nos levaram a inferir que, na fala, os usuários da língua têm preferência pelo preenchimento do centro anafórico na posição de sujeito por um pronome foneticamente realizado, ou seja, pelo pronome pleno.

De modo geral, os nossos resultados mostram que os pronomes plenos, na fala, parecem ser usados de modo aleatório, pois levando em consideração questões como quantidade de palavras entre o antecedente e a retomada ou o tipo de oração/parágrafo no qual a retomada é efetuada, o pronome pleno surge em qualquer contexto.

Levando em consideração estes fatos, corroboramos Homlberg, Nayudu e Sheehan (2009) pelo fato de acreditarmos que, na língua falada, o parâmetro *pro-drop* parece estar fadado ao desaparecimento quando levado em consideração os pronomes correferentes, pois o pronome pleno está superando, em termos de frequência, o uso do pronome nulo. Porém, como o pronome nulo, mesmo em menor quantidade, ainda configura o quadro de formas pronominais

correferenciais e é usado pelos falantes, acreditamos que o PB é, de fato, uma língua parcialmente *pro-drop*, apesar de este não ser o objeto de nosso estudo.

Dando prosseguimento à nossa discussão, trataremos a partir de agora o uso das macrocategorias de nomes e pronomes com função correferencial em textos falados em Português Brasileiro.

Quando analisadas estas duas macrocategorias, levamos em consideração as afirmações de Grosz e Sidner (1986) e de Grosz, Joshi e Weinstein (1995) ao postularem a Teoria da Centralização e, por isso, contabilizamos a quantidade de ocorrências de nomes e pronomes em posição de sujeito que retomavam um referente também em posição de sujeito.

Conforme Grosz e Sidner (1986) e Grosz, Joshi e Weinstein (1995), os pronomes são as formas que tendem a retomar um antecedente, pois o uso de um nome retomando o mesmo nome antecedente, tende a deixar o processamento mais cognitivamente custoso. Com isso, Grosz e Sidner (1986) e Grosz, Joshi e Weinstein (1995) afirmam que, em centros anafóricos na posição de sujeito, as formas que devem retomar o antecedente são os pronomes por serem os elos naturais de manutenção do referente na mente do usuário da língua.

Ao quantificarmos as ocorrências de nomes e pronomes com função correferente na posição de sujeito de textos falados, encontramos 1790 ocorrências, sendo 773 das retomadas sendo efetuadas por nomes e 1017 por pronomes, implicando em 43,1% das retomadas feitas por nomes, sem diferença significativa ( $p$ -valor = 0,65) e 56,9% por pronomes, também sem diferença significativa ( $p$ -valor = 0,74).

Com estes resultados, percebemos que, em textos falados, os pronomes são mais frequentes do que os nomes, o que pode ser um indicativo de que o que fora proposto por Grosz e Sidner (1986) e Grosz, Joshi e Weinstein (1995) pode ser correto, mostrando que os pronomes são os elos naturais para manutenção de um referente na mente do falante. Entretanto, é necessário observar que, mesmo os pronomes sendo as formas preferenciais, os nomes aparecem como também formas legítimas de retomadas de antecedentes, pois a diferença percentual no valor de  $p$  de 0,09 na quantidade de ocorrências efetuadas por nomes e pronomes não aponta para um efeito no qual o pronome seja a forma privilegiada para preencher um centro anafórico na fala em PB.

Diante de um resultado, de certa forma, inesperado de acordo com a postulação feita por Grosz e Sidner (1986) e Grosz, Joshi e Weinstein (1995), temos de observar diversos fatores, partindo, de antemão, da vertente da língua na qual encontramos estes resultados: a língua falada.

Para discutir sobre isto, é preciso lembrar o fato de que as pesquisas linguísticas, durante a construção dos *designs* experimentais, partem da pressuposição de como as pessoas falam e transpõem as crenças linguísticas obtidas por meio da vivência com a língua falada para a língua escrita.

É necessário reiterar que os postulados que compõem a Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995) partem da fala para experimentos de leitura e postulam informações que englobam a língua, ou seja, não há discriminação do que é fala e do que é escrita.

Um reflexo de que é preciso fazer esta distinção em pesquisas psicolinguísticas pode ser percebido a partir de nossos resultados da pesquisa da ocorrência das macrocategorias de nomes e pronomes com função correferencial na posição de sujeito pelo fato de tanto os nomes quanto os pronomes, e não somente estes, serem formas legítimas para a retomada de referentes.

Além disso, o que podemos afirmar, pautados nos indícios dados por meio dos resultados encontrados, é que os nomes e os pronomes são formas esperadas para a retomada de um antecedente, sendo que, na fala, os pronomes são mais frequentes do que os nomes quando preenchem um centro anafórico na posição de sujeito, retomando um antecedente também em posição de sujeito.

Mais uma vez gostaríamos de chamar a atenção para o fato de que as pesquisas psicolinguísticas em geral tendem a observar e explicar o uso de nomes e pronomes com função correferencial dentro das sentenças (nível intrassentencial) e/ou quando há o antecedente sendo retomado em uma outra sentença, geralmente próxima em quantidade de sílabas (nível interssentencial).

É comum, assim como fizeram Grosz e Sidner (1986) e Grosz, Joshi e Weinstein (1995), chamar o agrupamento de sentenças de segmento discursivo e, a partir de então, analisar as interações que levam ao uso de um nome ou de um pronome como forma correferente.

Porém, acreditamos que, quando analisado um texto inteiro, com uma cadeia (cor)referencial em uso, temos mais indícios do que acontece quando o falante faz uso de uma forma ou de outra e, partindo deste pressuposto, a primeira revelação importante dada por nossos resultados é: os falantes usam tanto nomes como pronomes para retomar antecedentes, sendo mais frequente o uso de pronomes do que nomes na língua falada.

O fato de, segundo Grosz e Sidner (1986) e Grosz, Joshi e Weinstein (1995), os pronomes serem as formas preferenciais para preencher centros anafóricos na posição de sujeito ao retomarem um antecedente também em posição de sujeito, levou Gordon, Grosz e Guillion (1993) a postularem que, se os pronomes devem retomar uma entidade previamente introduzida, é esperado que haja penalização dos nomes e surge, então, o conceito do efeito da Penalidade do Nome Repetido.

O efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) tem sido amplamente pesquisado em PB e corroborado em diversas pesquisas psicolinguísticas (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; QUEIROZ, 2009; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2015; GONDIM, 2017) nas quais os autores afirmam o fato de que os nomes repetidos são penalizados por pronomes plenos.

Entendendo que a análise de dados de *corpus* de língua falada não pode ser diretamente comparados com resultados encontrados em pesquisas experimentais por serem métodos de naturezas diferentes, usamos fatores elencados em pesquisas psicolinguísticas para, em textos autênticos, observar os indícios que poderiam ser encontrados a partir da frequência de uso das formas correferentes, sem tocar diretamente no que tange às penalidades por este ser um fenômeno exclusivamente cognitivo.

Tomando como base o conceito do efeito de Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993), conforme concebido por seus idealizadores, buscamos observar a frequência de ocorrências dos nomes repetidos comparando com a frequência da ocorrência de pronomes plenos, esperando indícios que nos levassem a compreender a distribuição destas formas em PB.

Porém, antes de prosseguir com nossa discussão, precisamos lembrar que, em nosso trabalho, partimos de duas macrocategorias de formas

correferentes: os nomes e os pronomes. Uma vez feita a subdivisão dos pronomes em pronomes plenos e em pronomes nulos, temos de fazer o mesmo para a macrocategoria dos nomes, pois, durante nossas análises, conforme explicado no capítulo anterior, encontramos dois tipos de nomes: os nomes repetidos e os nomes recategorizados.

Dito isto, voltamos aos nossos resultados explicando que, a macrocategoria dos nomes pode nos levar a duas interpretações de nossos resultados; (i) nomes – incluindo nomes repetidos e nomes recategorizados – *versus* pronomes plenos e (ii) nomes repetidos – excluindo nomes recategorizados – *versus* pronomes plenos.

Para (i) temos 1338 retomadas no total, sendo 773 retomadas efetuadas por nomes e 565 por pronomes plenos, implicando em 57,7% das retomadas feitas por nome, sem diferença significativa ( $p$ -valor = 0,74) e 42,3% por pronomes plenos, também sem diferença significativa ( $p$ -valor = 0,64). Para (ii) temos 1003 retomadas no total, sendo 438 por nomes repetidos e 565 por pronomes plenos, implicando em 43,6% das retomadas feitas por nomes, sem diferença significativa ( $p$ -valor = 0,65) e 56,4% por pronomes plenos, também sem diferença significativa ( $p$ -valor = 0,74).

Levando em consideração que o PB é uma língua que admite a retomada do sujeito por meio do pronome nulo, assim como acontece em espanhol argentino, levamos em consideração o que Gerlomini-Lezama (2008) denominou como Penalidade do Pronome Pleno.

Para este Gerlomini-Lezama (2008), em línguas em que o parâmetro do sujeito nulo existe, é mais fácil de o leitor/ouvinte compreender o pronome nulo do que o pronome pleno quando comparado ao tempo de processamento do pronome nulo em relação ao nome repetido, isto é, se comparados os tempos de processamento das retomadas nome repetido/pronome nulo e nome repetido/pronome pleno, está será mais difícil de processar em milésimos de segundos.

O efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) foi corroborado em PB por algumas pesquisas (MAIA; CUNHA-LIMA, 2011, 2012, 2014; MAIA, 2013; LIMA, 2015) em que foi encontrado o uso do pronome nulo penalizando o uso do nome repetido, sendo, inclusive, nestes contextos, mais facilmente processado do que o pronome pleno.

A fim de averiguar se, no todo textual, seria possível encontrar possíveis influências deste efeito, analisamos somente as retomadas feitas por nomes/pronomes nulos, levando em consideração as duas interpretações possíveis para os resultados encontrados para nós na macrocategoria de nomes.

Para (i) temos 1225 retomadas no total, sendo 773 retomadas efetuadas por nomes e 452 por pronomes nulos, implicando em 63,1% das retomadas feitas por nomes, sem diferença significativa ( $p$ -valor = 0,78) e 36,9% por pronomes plenos, também sem diferença significativa ( $p$ -valor = 0,61). Para (ii) temos 890 retomadas no total, sendo 438 por nomes repetidos e 452 por pronomes plenos, implicando em 49,2% das retomadas feitas por nomes, sem diferença significativa ( $p$ -valor = 0,69) e 50,8% por pronomes plenos, também sem diferença significativa ( $p$ -valor = 0,70).

A fim de encontrar mais evidências de como estão distribuídos os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos com função correferencial em PB, observamos se a distância seria um fator de influência no tipo de retomada e, por isso, analisamos se a quantidade de palavras entre o antecedente e a retomada traria novos indícios para este campo de pesquisa.

Tínhamos a necessidade de averiguar este fator pelo fato de já ter sido indicado que, levando em consideração o afastamento entre o antecedente e a retomada em até 38 sílabas há influência da distância no processamento correferencial (LEITÃO; SIMÕES, 2011), quando há uma sentença interveniente entre o antecedente e a retomada, não (LIMA, 2015).

Após a análise dos resultados expostos acima, temos considerações interessantes a fazer, pois temos indícios de que quanto mais perto estiver o referente em quantidade de palavras, mais favorável é o uso do pronome nulo e, quanto mais longe estiver o referente da retomada em quantidade de palavras, mais favorável é o uso do nome repetido.

A fim de confirmar nossa interpretação, dispomos dos resultados em dados de *corpus* de fala que nos mostram que, quando há, entre o antecedente e a retomada, até 10 (dez) palavras, 78,5% das retomadas foram feitas por pronomes nulos ( $p$ -valor = 0,84), e 21,5% por nomes repetidos ( $p$ -valor = 0,43), sem diferença significativa quando comparadas somente as retomadas entre si.

Já quando há, entre o antecedente e a retomada, mais de 100 (cem) palavras, o uso do pronome nulo cai drasticamente, tendo sido feitas 99,4% das retomadas por nomes repetidos ( $p$ -valor = 0,82) e 0,6% por pronomes nulos ( $p$ -valor = 0,42) nesta distância, novamente, sem resultados significativos quando comparadas as retomadas entre si.

Encontramos resultados estatisticamente significativos quando cruzamos o tipo de retomada com a distância na qual se encontra, o que nos deu mais indícios da influência da distância na forma correferente usada no texto falado.

Ainda tratando sobre a quantidade de palavras entre o antecedente e a retomada, observamos que os pronomes plenos surgiram em todos os intervalos analisados por nós, com exceção do intervalo de 0 (zero) a 10 (dez) palavras.

Apesar de não tratarmos aqui de uma pesquisa experimental, que nos dariam evidências sobre o que ocorre na mente do ouvinte, não podemos deixar de mostrar que acreditamos que os nomes repetidos e os pronomes nulos são as primeiras formas correferenciais que vêm à mente de um usuário da língua quando ele tem de preencher um centro anafórico e, a depender de fatores que estamos tentando identificar, os nomes recategorizados e os pronomes plenos surgem como opções para as duas primeiras formas.

Considerando esta possibilidade, podemos entender melhor a frequência do pronome pleno na língua falada, pois no intervalo de 11 (onze) a 50 (cinquenta) palavras e no intervalo de 51 (cinquenta e uma) a 100 (cem) palavras, o pronome pleno surge naturalmente, estando em concorrência com o nome recategorizado de maneira quase acirrada. Este mesmo comportamento acontece, também, quando o antecedente e a retomada estão próximos em até 10 (dez) palavras, contexto no qual o uso do pronome nulo se torna preferencial. Do mesmo modo ocorre quando há mais de 100 (cem) palavras entre o antecedente e a retomada, contexto no qual o uso do nome repetido gera efeito de penalidade sobre o pronome nulo, o pronome pleno é tão frequente quando o nome repetido.

Esta frequência parece sugerir que o pronome pleno funciona como uma possibilidade de uso em qualquer centro anafórico de texto falado, pois é como se, em qualquer distância, independentemente da quantidade de palavras, o pronome pleno pudesse aparecer naturalmente, sem causar danos à produção do texto.

Já o uso dos nomes recategorizados parece seguir um comportamento padrão, pois eles estão sempre seguindo os nomes repetidos em quantidade de ocorrências, mas sempre abaixo destes e concorrendo com o pronome pleno, pois quando há, entre o antecedente e a retomada, até 10 (dez) palavras, os nomes recategorizados são mais frequentes do que os nomes repetidos (16,9% das retomadas feitas por nomes recategorizados e 12,8% por nomes repetidos), sem diferença significativa ( $p$ -valor = 0,99). Quando há entre 11 (onze) e 50 (cinquenta) palavras, o mesmo comportamento se mantém (22,1% das retomadas feitas por nomes recategorizados e 18,9% por nomes repetidos), sem diferença significativa ( $p$ -valor = 0,98). Quando os nomes repetidos começam a ser as formas preferencialmente usadas para a retomada, o nome recategorizado surge claramente como um opção, quando há entre 51 (cinquenta e uma) e 100 (cem) palavras (48% das retomadas feitas por nomes repetidos e 25,5% por nomes recategorizados), sem diferença significativa ( $p$ -valor = 0,82).

O fato é que, levando em consideração os nomes recategorizados, não temos indícios de que, na língua falada, haja uma distância que propicie o surgimento desta forma como preferencial, mas podemos afirmar que, se a distância entre o antecedente e a retomada for mediana, entre 11 (onze) e 50 (cinquenta) palavras, o nome recategorizado é potencialmente uma forma viável de retomada.

É interessante comentar que, em nossa pesquisa, no intervalo de 51 (cinquenta e uma) a 100 (cem) palavras, não encontramos ocorrência do uso do pronome nulo nos textos falados. Isto poderia ser mais um reforço para a hipótese de que pode haver contextos em que o uso do nome repetido é tão frequente que pode gerar efeito de penalidade, mas isso carece de pesquisa experimental, sendo apenas uma pressuposição que fazemos, baseados em dados de frequência. Aqui, mais uma vez, reiteramos que nossa hipótese, mesmo existindo, não pode tomada à baila como algo concreto, pois pesquisa em dados de *corpora* linguístico é de natureza distinta de pesquisa experimental, além do fato de produto ser diferente de processamento e, por isso, não podemos julgar a frequência como parâmetro de comprovação de compreensão.

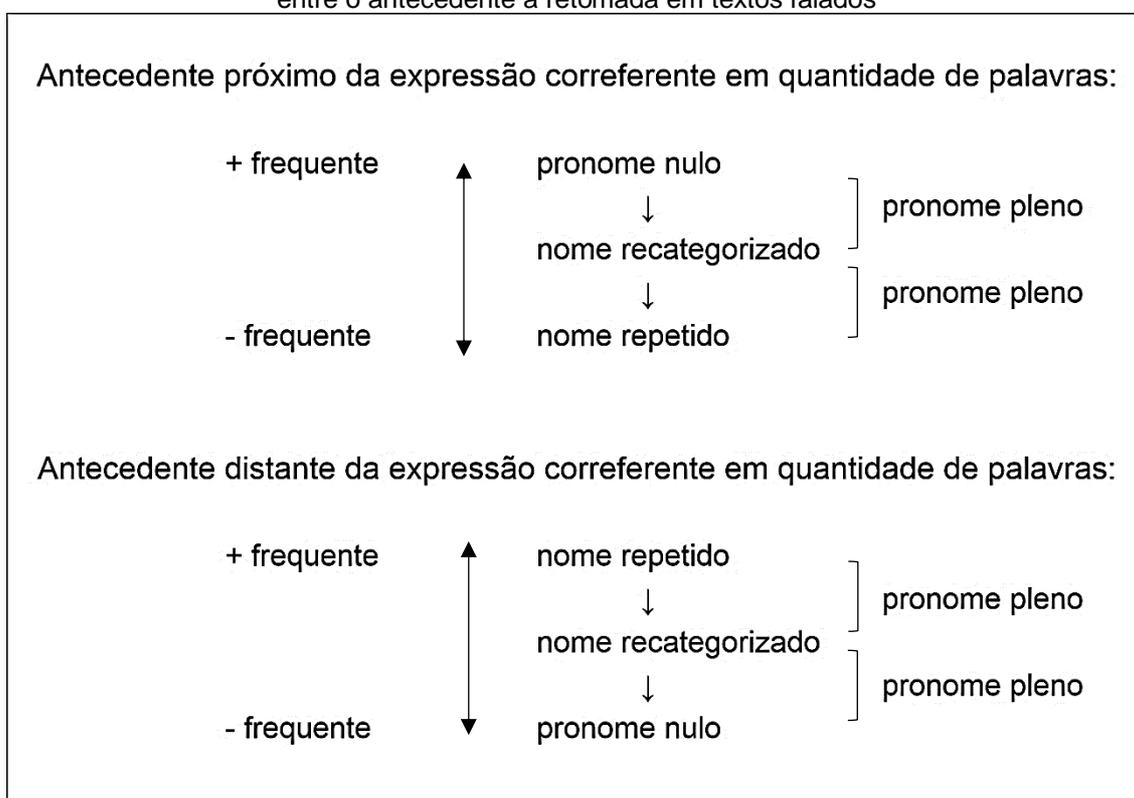
Por fim, levando em consideração a quantidade de palavras entre o antecedente e a retomada, o que podemos afirmar é que, em distâncias menores, o uso dos pronomes nulos é preferencial, mas não gera efeito de

penalidade sobre as demais formas, na fala; enquanto em distâncias maiores, o uso dos nomes repetidos é preferencial, ocorrendo efeito de penalidade sobre os pronomes nulos.

Reiteramos que, ao cruzar as todas as distâncias analisadas e o tipo de retomada, encontramos significância estatística, o que nos leva a afirmar que a distância pode ter influência direta no uso da forma correferente em textos falados em Português Brasileiro.

Após a análise conjunta de nossos resultados, encontramos o padrão de uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na língua falada mostrado no quadro 21.

**Quadro 21** - Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB falado de acordo com a quantidade de palavras entre o antecedente a retomada em textos falados



De fato, percebemos a partir de nossos resultados que as formas correferentes que sofrem influência direta da quantidade de palavras entre o antecedente e a retomada são os pronomes nulos (usados quando estão próximo do antecedente) e os nomes repetidos (usados quando estão distantes do antecedente), enquanto os nomes recategorizados tendem a ser usados em distâncias intermediárias e o pronome pleno paira sobre todas as formas, pois,

independente da distância em quantidade de palavras, o pronome pleno pode ser usado em qualquer distância na língua falada.

Dando continuidade à busca por evidências sobre os efeitos encontrados em pesquisas psicolinguística em que foram utilizados métodos experimentais, analisamos dois efeitos que devem ser levados em consideração quando se buscam efeitos de penalidade entre as formas correferentes: o efeito de Paralelismo Sintático (CHAMBERS; SMITH, 1998) e o efeito de Proeminência Sintática (GORDON; HENDRICK, 1997).

De acordo com Gordon e Hendrick (1997), o sujeito é a posição sintática mais proeminente da sentença e, por isso, há mais tendência de a retomada acontecer também em posição de proeminência, ou seja, de sujeito, do que em posições não proeminentes.

Pensando nisso, Chambers e Smith (1998) previram que, se o sujeito é a posição sintática mais proeminente, é esperado que o centro anafórico em posição de sujeito retome mais antecedente em posições proeminentes do que em posições sintáticas não proeminentes, criando, assim, o conceito do efeito de Paralelismo Sintático.

Nossos resultados mostram que, sim, o efeito de Paralelismo Sintático é percebido na língua falada, já que das 1790 retomadas analisadas por nós ao longo dos textos, 81,5% das ocorrências mostram a retomada sujeito/sujeito e 18,5% mostram a retomada entre um antecedente em posições sintáticas não proeminentes sendo retomados na posição de sujeito, sem diferença significativa ( $p$ -valor = 0,52).

Estes resultados, além de corroborar pesquisas que tratam do efeito de Paralelismo Sintático, servem também para que observemos em quais posições sintáticas os antecedentes tendem a aparecer quando estão em posições não proeminentes.

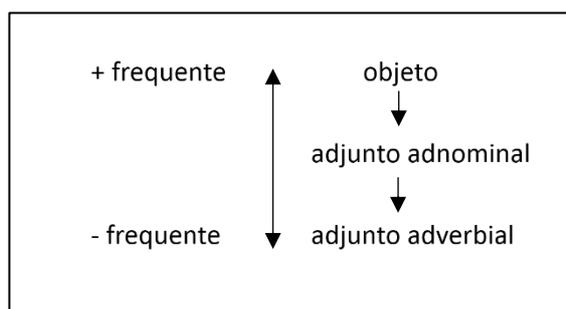
Para chegar aos resultados das posições não proeminentes, deixamos de lado as retomadas efetuadas entre sujeito/sujeito, por termos os resultados de Paralelismo Sintático que já tratam diretamente deste tipo de retomada e, reportá-los também na comparação de posições proeminentes e não proeminentes seria redundante, pois, conforme afirmam Gordon e Hendrick (1997) temos fortes indícios de que um antecedente em posição proeminente tende a ser retomado também em posição proeminente.

Feita esta observação, interessou-nos, então, analisar se havia algum padrão de proeminência para as demais posições sintáticas, uma vez que Gordon e Hendrick (1997) citam a posição de sujeito como proeminente e a posição de objeto como não proeminente.

Contudo, enquanto analisávamos nossos dados, percebemos que haviam retomadas de antecedentes que estavam em posição sintática de objeto, de adjunto adnominal e de adjunto adverbial<sup>217</sup> e, por isso, fizemos destas três posições as nossas subcategorias de análise, tendo encontrado 48% das ocorrências retomando antecedente em posição de objeto (p-valor = 0,66), 40,1% em posição de adjunto adnominal (p-valor = 0,65) e 11,9% em posição de adjunto adverbial (p-valor = 0,46) na língua falada. Esses valores de p implicam que as diferenças nas posições objeto/sujeito, adjunto adnominal/sujeito e adjunto adverbial/sujeito não são significativas.

Mesmo que este não seja o objetivo central de nosso estudo, acreditamos que estes resultados mostram um padrão que merece ser observado em pesquisas futuras, pois há uma possível gradação de posições sintática a serem retomadas por uma forma correferencial em posição de sujeito, conforme o quadro que segue.

**Quadro 22** – Escala de posições sintáticas que introduzem o referente quando não há Efeito de Paralelismo Sintático na fala



Tendo em vista esta possível gradação de possibilidades de posições sintáticas dos antecedentes e, se comprovado em pesquisas futuras que estas retomadas são possíveis, poderíamos ter a disposição novos indícios de que os efeitos de Penalidade do Nome Repetido e de Penalidade do Pronome Pleno não são exclusivos da retomada sujeito/sujeito, como previsto por Gordon, Grosz e Guillion (1995).

<sup>217</sup> Ver ressalvas sobre a subcategoria de adjunto adverbial no capítulo anterior.

Por acreditarmos que os efeitos de penalidade não são excludentes entre si, mas que dependem de contextos para serem percebidos na língua, achamos ser possível que tanto o efeito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) quanto o efeito da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) podem ser encontrados também quando o antecedente e a forma correferencial utilizada para a retomada não estiverem em paralelismo sintático.

É a possibilidade de mudança de uma posição mais proeminente sendo retomada por uma posição menos proeminente ou vice-versa que nos leva a crer que os resultados de Leitão (2005) e Queiroz (2009) são indícios a serem considerados em uma possível futura reformulação no conceitos dos efeitos de penalidade, conforme idealizados por Gordon, Grosz e Guilliom (1993) e, posteriormente, por Gerlomini-Lezama (2008).

Na busca por contextos que possam vir a priorizar o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos, optamos por separar o texto em blocos que seguissem o padrão estudados em pesquisas psicolinguísticas em PB e analisamos a ocorrência destas formas em (i) orações subordinadas (MORGADO, 2011; LUEGI, 2012; LUEGI; MAIA, 2012, TEIXEIRA, 2014; TEIXEIRA, FONSECA; SOARES, 2014; HORA, 2014) (ii) orações coordenadas justapostas (LEITÃO, 2005; LEITÃO; MAIA, 2005; QUEIROZ, 2009; LEITÃO; SIMÕES, 2009; LEITÃO; RIBEIRO; MAIA, 2012; TEIXEIRA, 2014; SIMÕES, 2014; GONDIM, 2017)<sup>218</sup> (iii) orações coordenadas em períodos distintos (LEITÃO; PEIXOTO; SANTOS, 2012; MAIA; CUNHA-LIMA, 2011; MAIA, 2013; MAIA; CUNHA-LIMA, 2014; LIMA, 2014; LIMA, 2015; ALMOR *et al.*, 2017) e (iv) a primeira retomada em parágrafo subsequente.

É importante reforçar que, dentre as pesquisas acima citadas, há relação entre a estrutura sintática e possibilidade da expressão correferente usada, mas não há indicação por parte dos autores de que a estrutura sintática seja fator determinante para o surgimento de uma forma correferencial ao invés de outra. O que ocorre, de maneira geral, são observações para os efeitos de penalidade supracitados que podem ou não ocorrer nos contextos pesquisados.

---

<sup>218</sup> Reiteramos que estes estudos tratam tanto de orações coordenadas justapostas quanto de orações coordenadas articuladas por conjunções, tendo sido agrupados, aqui, para traçarmos comparações. Para compreender melhor estes trabalho, ver o capítulo 03.

Em contrapartida, acreditamos que o tipo de oração ou a forma como as orações estão dispostas no nível textual, se em blocos de orações justapostas ou se em parágrafos subsequentes, e a retomada por nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos está intimamente ligado.

De fato, mesmo compartilhando da ideia de que o usuário da língua faz ligações sintáticas em um primeiro momento e depois verificações de cunho semântico, conforme postulam Van Gompel, Liversedge e Pearson (2004), acreditamos que, mesmo no momento dos cálculos sintáticos, o usuário analisa mentalmente (1) a distância em quantidade de palavras entre o antecedente e a retomada e (2) o tipo de sentença no qual a retomada é efetuada.

Para (1) já temos indícios de que, em textos falados, quanto mais próximo for o antecedente da retomada, maior é a frequência de uso dos pronomes nulos e que, quanto mais distante for o antecedente da retomada, mais frequente é o uso dos nomes repetidos.

Já para (2), temos indícios de que, em (i), orações subordinadas, de todas as retomadas feitas, 5,4% foram efetuadas por nomes repetidos (p-valor = 0,44), 3,1% por nomes recategorizados (p-valor = 0,43), 21,7% por pronomes plenos (p-valor = 0,52) e 69,8% por pronomes nulos (p-valor = 0,82). Em (ii), orações coordenadas justapostas, de todas as retomadas feitas, 4,2% foram efetuadas por nomes repetidos (p-valor = 0,43), 10,3% por nomes recategorizados (p-valor = 0,46), 45,8% por pronomes plenos (p-valor = 0,67) e 39,7% por pronomes nulos (p-valor = 0,63). Em (iii), orações coordenadas em períodos distintos, de todas as retomadas feitas, 16,8% foram efetuadas por nomes repetidos (p-valor = 0,49), 34,9% por nomes recategorizados (p-valor = 0,60), 44,6% por pronomes plenos (p-valor = 0,66) e 3,7% por pronomes nulos (p-valor = 0,43). E, em (iv), parágrafos subsequentes, de todas as retomadas feitas, 66,4% foram retomadas efetuadas por nomes repetidos (p-valor = 0,80), 22,8% por nomes recategorizados (p-valor = 0,53), 10,5% por pronomes plenos (p-valor = 0,46) e 0,3% por pronomes nulos (p-valor = 0,42) nos textos analisados.

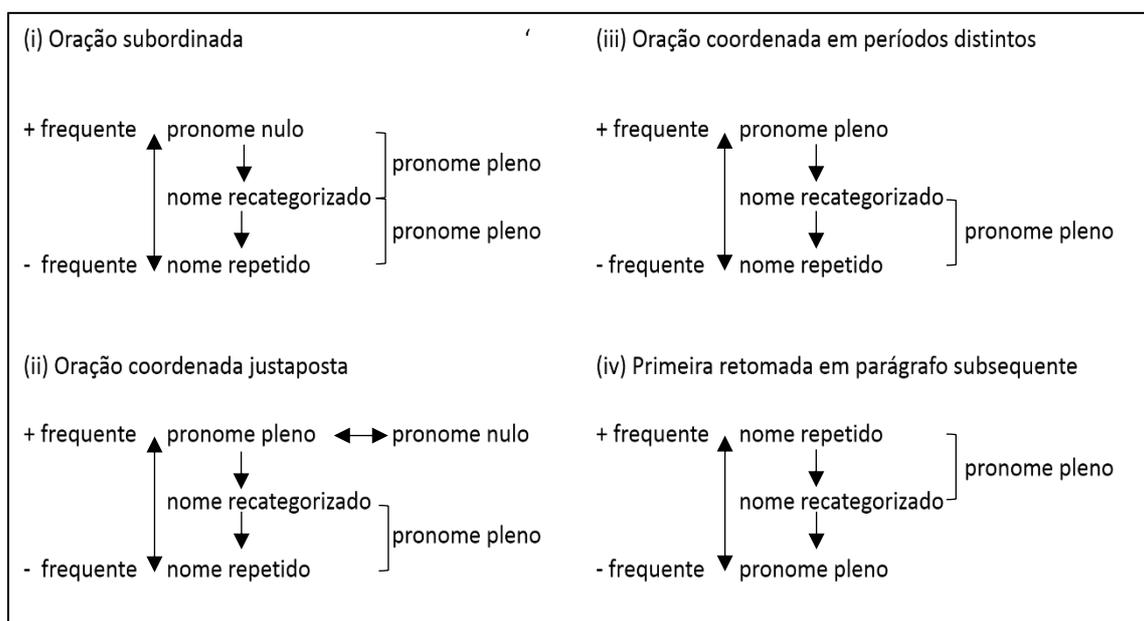
Estes resultados mostram que, em (i), os pronomes nulos são mais frequentes, em (ii), os pronomes nulos e os pronomes plenos podem ocorrer quase na mesma frequência, em (iii), os nomes recategorizados e os pronomes plenos concorrem, sendo mais comum o uso do pronome pleno do que do nome recategorizado, e, em (iv), contexto não pesquisado antes, os nomes repetidos são mais frequentes.

Os usos das formas correferentes na relação tipo de oração/quantidade de palavra pode nos dar informações importantes, pois, se considerarmos que em (i) há uma tendência de os antecedentes estarem próximos da retomada, temos a explicação do porquê de o uso do pronome nulo ser tão frequente neste contexto. Como (ii) e (iii) já apontam para contextos em que a quantidade de palavras entre o antecedente e a retomada tende a aumentar, a queda de frequência no uso do pronome nulo, o aumento na frequência do uso do nome recategorizado e a presença constante do pronome pleno ficam explicadas. Já em (iv), por ser um contexto no qual há tendência do longo distanciamento entre o antecedente e a retomada, fica explicada a alta frequência no uso dos nomes repetidos e a baixa frequência no uso dos pronomes nulos.

Estes resultados, estatisticamente significativos, reforçam nossa tese de que a quantidade de palavras e a disposição das orações/parágrafos no texto têm ligação direta com o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito em Português Brasileiro.

Se nossa concepção estiver certa, encontramos um padrão de uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferencial na posição de sujeito em Português Brasileiro. Este padrão seria, na fala, de acordo como mostrado no quadro a seguir.

**Quadro 23** – Distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB de acordo com a disposição textual em textos falados



Acreditamos, portanto, que, em (i), o uso frequente do pronome nulo é seguido pelo nome recategorizado e, em menor frequência, pelo nome repetido, sendo que, o pronome pleno pode ocorrer sem prejuízo à produção do texto. Em (ii), o pronome pleno surge com forma preferencial, podendo ocorrer quase na mesma frequência que o pronome nulo, enquanto os nomes recategorizados e os nomes repetidos são menos frequentes. Em (iii), tanto o pronome pleno quanto o nome recategorizado podem surgir quase na mesma frequência, enquanto o pronome nulo pode aparecer raramente, já o nome repetido ocorre com menor frequência. E, em (iv), o nome repetido é a forma preferencialmente usada, podendo ocorrer o surgimento do nome recategorizado ou mesmo do pronome pleno, enquanto o pronome nulo pode raramente ocorrer neste contexto.

Portanto, interligando a quantidade de palavras entre o antecedente e a retomada e o tipo de disposição textual das orações e parágrafos ao longo do texto, acreditamos encontrar um padrão de distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulo com função correferencial na posição de sujeito, na fala, em Português Brasileiro.

Além disso, é preciso frisar que estamos propondo uma distribuição gradual de aparição das formas correferenciais por nós estudadas a depender do contexto sintático e que, se da forma mais frequente para a menos frequente uma delas não aparecer, o usuário terá à disposição a forma seguinte.

Com isto, queremos dizer que, se em uma oração subordinada, por exemplo, o usuário preferiu a forma correferencial do pronome pleno, em algum momento da resolução correferencial, ele pode ter cogitado antes o uso de um pronome nulo, mas isto é algo que supomos e que ficará a cargo de pesquisas futuras comprovar se ocorre ou não.

A seguir, decidimos apenas averiguar a frequência de dois fatores que, nas pesquisas psicolinguísticas sobre o processamento correferencial ocorrem: o tipo de verbo e a presença do artigo.

Averiguamos o tipo de verbo que ocorre na sentença anterior à sentença na qual há a retomada e analisamos os verbos divididos em três subcategorias: verbos intransitivos, verbos transitivos e verbos de ligação. Encontramos a ocorrência de 15,3% das sentenças anteriores à retomada com verbos intransitivos, sem diferença significativa ( $p$ -valor = 0,60), 66% com verbos

transitivos (94,3% transitivos diretos, 5,1% transitivos indiretos e 0,6% bitransitivos), sem diferença significativa ( $p$ -valor = 0,52) e 18,7% com verbos de ligação, igualmente sem diferença significativa ( $p$ -valor = 0,65).

Com estes resultados, gostaríamos de comentar que as pesquisas feitas em PB (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; QUEIROZ, 2009; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; MAIA; CUNHA-LIMA, 2011, 2014, 2014; MAIA, 2013; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; TEIXEIRA, 2013; SIMÕES, 2014; LIMA, 2015; GONDIM, 2017), por serem a maioria com sentenças experimentais na quais aparecem verbos transitivos diretos na sentença anterior à sentença na qual há a retomada, mostram um padrão autêntico da língua.

De qualquer maneira, gostaríamos de frisar que não acreditamos que o tipo de verbo – verbo intransitivo, verbo transitivo direto ou verbo de ligação – seja um dos fatores mais importantes quando o usuário vai fazer uso de uma forma correferente.

Por fim, averiguamos que os antecedentes tendem a ser introduzidos por artigos definidos, por artigos indefinidos ou por outras formas, conforme explicado na seção anterior e encontramos 60% dos antecedentes precedidos por artigos definidos, sem diferença significativa ( $p$ -valor = 0,70), 6,4% por artigos indefinidos, também sem efeito significativo ( $p$ -valor = 0,44) e 33,6% por outras formas, igualmente sem diferença significativa ( $p$ -valor = 0,59), o que, de maneira geral, também segue a tendência de os antecedentes serem introduzidos por artigos definidos na maioria das pesquisas psicolinguísticas em PB sobre o processamento correferencial<sup>219</sup>.

No que diz respeito às teorias e hipóteses semânticas sobre o processamento anafórico, gostaríamos de reafirmar que compartilhamos da crença de que o processamento correferencial passa por dois estágios, conforme proposto por Van Gompel, Liversedge e Pearson (2004).

Acreditamos que, em tarefas de resolução correferencial, o usuário da língua faz análise de, no mínimo, dois estágios, sendo o primeiro estágio voltado para a identificação do antecedente, envolvendo questões de cunho sintático, e, um estágio subsequente, no qual a forma semântica da anáfora é verificada,

---

<sup>219</sup> Questões mais aprofundadas sobre questões semânticas ligadas ao processamento correferencial não são objetivo de nosso trabalho e, por isso, critérios como animacidade/inanimacidade não foram levados em conta em nosso estudo.

estando envolvidas, aqui, questões de número, de gênero e de animacidade/inanimacidade, por exemplo.

Reiteramos, também, nossa crença de que, dentre as opções que o usuário da língua dispõe para efetuar a correferência, há formas que são mais frequentes do que outras, o que desemboca nos achados de Almor (1999). Assim como este autor, acreditamos que existem formas que demandam maior esforço para ser ligado ao seu antecedente e, por isso, podem gerar prejuízos para a compreensão global do texto.

Por fim, acreditamos ter indícios para corroborar a Teoria da Acessibilidade (ARIEL, 1991), pois, para a autora, há um padrão universal, que pode variar entre as línguas, para como as formas linguísticas são acessadas na mente do usuário, variando de formas mais acessíveis a formas menos acessíveis.

Na escala proposta por Ariel (1991), os pronomes nulos, que ela chama de *zeros*, são altamente acessíveis enquanto os nomes, a depender de como usados no texto, ficam cada vez menos acessível.

Em nosso trabalho, associamos esta acessibilidade ao fator de distanciamento entre os antecedentes e as retomadas em quantidade de palavras e não analisamos todas as formas propostas pela autora, mas, dentre as formas pesquisadas neste trabalho, temos fortes indícios de que Ariel (1991) estava certa.

Contudo, não podemos deixar de fazer a ressalva de que, no Português Brasileiro falado, os pronomes plenos, parecem estar sendo usados, senão com a mesma, mas com uma função bem próxima da função dos pronomes nulos, o que, de certa forma, na escala de acessibilidade de Ariel (1991), poderia mostrar uma reconfiguração para o pronome nulo.

De qualquer maneira, encontramos indícios do que Ariel (1991) chama de *padrão universal de acessibilidade*, pois, analisando dados de *corpus* fala, vimos que os pronomes nulos retomam antecedentes mais acessíveis, os pronomes plenos e o nomes recategorizados estão em nível intermediário e os nomes repetidos são utilizados para trazer de volta à mente do interlocutor os referentes que estão em baixa acessibilidade, ou seja, mais distante do antecedente.

E, após a descrição minuciosa de nossos resultados e de buscar fazer uma ampla e profunda discussão com as pesquisas que nortearam este trabalho,

esperamos lançar luz em pesquisas futuras que tratem sobre a correferência anafórica.

No intuito de fechar nossas interpretações, apresentaremos no capítulo seguinte uma discussão geral para analisar, de forma conjunta, os resultados do *corpus* escrito e do *corpus* de fala na busca pelas explicações de como estão distribuídos os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos com função anafórica na posição de sujeito em Português Brasileiro.

## 6 DISCUSSÃO GERAL

Neste capítulo, empreendemos uma discussão sobre como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos com função anafórica em posição de sujeito estão distribuídos em Português Brasileiro, levando em conta os dados coletados em *corpus* escrito e em *corpus* de fala.

Nossa intenção é fazer um apanhado geral do que foi exposto nos dois capítulos anteriores desta tese e mostrar as similaridades e as diferenças entre a escrita e a fala no que tange ao uso de expressões correferenciais.

Para isto, seguiremos o percurso já exposto anteriormente, revendo os resultados de cada categoria de análise por nós estudada, com os resultados gerais e os comentários necessários.

A partir desta retomada de resultados e com base em nossas interpretações, abriremos espaço para um debate aprofundado sobre o que as pesquisas psicolinguísticas que tratam do processamento correferencial em PB têm mostrado e, com base nos resultados encontrados, termos a oportunidade de corroborar ou não estas pesquisas.

Além disso, seguiremos a ordem das teorias e hipóteses analisadas no segundo capítulo desta tese, pois buscamos ao máximo usar todas as teorias que fundamentaram nossa pesquisa como ponto de partida e de chegada de nossas interpretações.

Visando encontrar possíveis padrões de uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função correferente na posição de sujeito em PB, buscamos evidências que nos ajudassem, não somente explicar a distribuição destas formas na língua, mas, se possível, também auxiliar na construção de possíveis futuros experimentos psicolinguísticos que tratem deste tema.

A fim de começar nossa discussão, retomamos, agora, as categorias de análise que serão discutidas neste capítulo: (i) a ocorrência de nomes e pronomes com função anafórica na posição de sujeito; (ii) a ocorrência de nomes repetidos, pronomes plenos e pronomes nulos com função anafórica na posição de sujeito; (iii) a ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos na posição de sujeito; a quantidade de palavras entre o antecedente e a retomada em intervalos de (iv) 0 (zero) a (dez)

palavras, (v) 11 (onze) a 50 (cinquenta) palavras, (vi) 51 (cinquenta e uma) a 100 (cem) palavras; (vii) mais de 100 palavras; (viii) o efeito de Paralelismo Sintático; (ix) a ocorrência de antecedentes em posição sintáticas não proeminentes; o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em (x) orações subordinadas, em (xi) orações coordenadas justapostas, em (xii) orações coordenadas em períodos distintos, e em (xiii) parágrafos subsequentes; (xiv) o tipo de verbo que ocorre na sentença anterior à sentença da retomada; e, (xv) o ocorrência de artigo precedendo o antecedente.

Dentre as categorias de análise com as quais trabalhamos durante a coleta de dados, tivemos de excluir os resultados da quantidade de caracteres sem e com espaço entre o antecedente e a retomada por conta dos dados do *corpus* de fala.

Tomamos esta decisão pelo fato de acreditarmos que, quando o *corpus* é transcrito e siglas são introduzidas no texto para demarcar questões da fala no texto, o mapeamento da distância em quantidade de caracteres não é tão fiel quanto ocorre com dados do *corpus* escrito.

Por meio dos resultados do *corpus* escrito descritos no terceiro capítulo desta tese, é possível analisar como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos com função anafórica em posição de sujeito são usados levando em consideração a quantidade de caracteres sem e com espaço. Porém, estes mesmos resultados não são reportados no quarto capítulo que trata resultados dos dados do *corpus* de fala e, por conta desta distinção, esta categoria também não será retomada neste capítulo.

Antes de começar a falar especificamente de cada categoria de análise, gostaríamos de reiterar que, durante a coleta e análise dos dados, tivemos a oportunidade de observar o que as teorias e hipóteses abordadas no segundo capítulo preveem e, assim, tomamos posições que nos ajudaram a construir as categorias de análise para o nosso trabalho.

Tivemos a oportunidade de analisar o comportamento de *pro* – o pronome nulo, que é a contraparte foneticamente realizada do pronome pleno, contendo características -anafórico/+pronominal. Tal fato foi possível por termos dados que nos permitiram tecer comentários acerca do parâmetro *pro-drop* em Português Brasileiro.

Segundo Homlberg, Nayudu e Sheehan (2009), o PB é uma língua parcialmente *pro-drop* pelo fato de que a língua tem propiciado, por diversos motivos como sexo, idade, escolaridade e distribuição geográfica, o aparecimento do pronome pleno e o uso cada vez menos frequente do pronome nulo.

Apesar de não tratarmos de todas as formas pronominais nem de questões ligadas à sexo, idade, escolaridade e distribuição geográfica, podemos traçar comentários acerca do uso dos pronomes de terceira pessoa, pois observamos, dentro outras questões, a ocorrência de pronomes plenos e de pronomes nulos com função anafórica em PB.

Por termos empreendido coleta e análise de dados de *corpus* escrito e de *corpus* falado, nossos dados mostraram que, na escrita, de todas as retomadas feitas por meio de pronomes, 69,5% foram feitas pelo uso do pronome nulo e 30,5% pelo uso do pronome pleno, enquanto, na fala, foram 55,5% por pronomes plenos e 44,5% por pronomes nulos.

Estes resultados nos levam a observar que há dois padrões de uso dos pronomes em PB, sendo que, na escrita, as pessoas tendem a usar mais pronomes nulos e, na fala, usar mais pronomes plenos.

O uso de pronomes plenos e de pronomes nulos é tão próximo, em termos de quantidade de ocorrência, que nos leva a afirmar que ambos estão sendo usados quase com a mesma frequência em PB, mas não com a mesma frequência na fala e na escrita, o que nos faz corroborar Holmberg, Nayudu e Sheehan (2009) e crer que, no estágio atual, o Português Brasileiro é uma língua parcialmente *pro-drop*.

A primeira categoria de análise que estudamos tratava do uso dos nomes e dos pronomes com função anafórica na posição de sujeito em Português Brasileiro.

Tratamos, portanto, de duas macrocategorias de formas potencialmente anafóricas, pois, segundo Grosz e Sidner (1986), os pronomes são elos naturais de manutenção do mesmo referente na mente do leitor e, em pesquisas desenvolvidas por eles, foram encontrados resultados que indicam que, quando em posição de sujeito, retomando um antecedente também em posição de sujeito, o nome repetido é mais difícil de ser compreendido do que o pronome.

A fim de averiguar se na língua escrita e na língua falada a ocorrência de nomes e pronomes com função anafórica na posição de sujeito, retomando um

anterior também em posição de sujeito, poderia indicar a maior quantidade de uso de pronomes do que de nomes, começamos nossa análise pelas macrocategorias dos nomes e dos pronomes.

Nossos resultados mostraram que, nos dados do *corpus* escrito, de todas as retomadas analisadas, 51% foram feitas pelo uso de nomes como forma correferente e 49% pelo uso de pronomes, enquanto que, nos dados do *corpus* de fala, de todas as retomadas analisadas, 43,1% foram feitas pelo uso de nomes como forma correferente e 56,9% pelo uso de pronomes.

Encontramos, portanto, resultados que nos levam a discordar da afirmação de Grosz e Sidner (1986), pois podemos afirmar que, na língua como um todo, considerando a forma escrita e a forma falada, tanto o nome como o pronome podem ser considerados como elo de manutenção de referentes na mente do interlocutor.

Dizemos isto pelo fato de termos encontrado indícios de que, na escrita, há predominância do uso dos nomes e, na fala, há predominância do uso dos pronomes, mas são variações sem expressividade estatística, o que aponta o uso de nomes e pronomes com função anafórica como algo quase aleatório, se analisado apenas de forma geral.

A partir destes resultados, tivemos de criar subcategorias para as macrocategorias analisadas acima para, assim, tratarmos diretamente do que as pesquisas psicolinguísticas levam em conta.

Conforme explicamos nos capítulos anteriores, em um primeiro momento, tratamos da divisão da macrocategoria dos pronomes, analisando, portanto, as ocorrências dos pronomes nulos e dos pronomes plenos e, durante esta fase, renomeando a macrocategoria dos nomes para nomes repetidos.

Foi necessário fazer esta divisão da macrocategoria dos pronomes, a fim de respeitar o que fora proposto por Grosz e Sidner (1986), Gordon, Grosz e Guilliom (1993). Os autores analisaram o que ocorria quando uma entidade, na posição de sujeito, era introduzida por meio de um nome e depois retomada, também em posição de sujeito, pela repetição do mesmo nome ou do pronome e, assim, encontraram indícios de que os centros anafóricos eram preferencialmente preenchidos por pronomes e que o uso dos nomes repetidos tornava a leitura mais lenta, o que seria um indicativo de que, quando os nomes repetidos eram usados como forma anafórica, era mais difícil de compreender o segmento discursivo.

Foi a partir desta constatação que Grosz, Joshi e Weinstein (1995) postularam o conceito do efeito da Penalidade do Nome Repetido, afirmando que quando há um centro anafórico a ser preenchido na posição de sujeito, retomando um antecedente também em posição de sujeito, a forma anafórica que é mais fácil de ser compreendida é o pronome, penalizando, portanto, o uso dos nomes repetidos.

Utilizando este conceito e aplicando em Espanhol Argentino, uma língua *pro-drop* e que, portanto, admite o parâmetro do sujeito nulo, Gerlomini-Lezama (2008) comparou o tempo de processamento dos nomes repetidos, dos pronomes plenos e dos pronomes nulos com função anafórica em posição de sujeito, retomando um antecedente também em posição de sujeito, e encontrou que, nesta língua, a forma mais fácil de ser processada é o pronome nulo e não o pronome pleno e, assim, postulou o conceito do efeito da Penalidade do Pronome Pleno.

Em Português Brasileiro, pesquisadores da área da psicolinguística têm tratado tanto sobre o efeito da Penalidade do Nome Repetido quanto da Penalidade do Pronome Pleno, encontrando sempre resultados que vão de encontro um ao outro, pois as pesquisas de Leitão (2005) e colaboradores (QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; GONDIM, 2017) confirmam que há, em PB, o efeito da RNP, enquanto as pesquisas de Maia e Cunha-Lima (2011, 2012, 2014), Maia (2013) e Lima (2015) afirmam que há, em PB, o efeito da OPP.

Levando em consideração esta realidade e renomeada/dividida a macrocategoria nas mesmas formas anafóricas estudadas nas pesquisas psicolinguísticas que motivaram nosso trabalho, analisamos o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função anafórica na posição de sujeito em PB na busca da distribuição das formas correferentes na língua usada em contexto autêntico.

Porém, tratar do conceito de penalidade em dados de *corpus* e não em pesquisa experimental, não é algo usual, pois as pesquisas que tratam do *corpus* mostram evidências no produto, enquanto as pesquisas experimentais trazem evidências do processamento linguístico. Dada a natureza distinta destas metodologias, não se pode associar frequência de uso com compreensão, pois não é necessário que algo seja frequente para ser compreendido e, por isso,

preferimos não tocar sobre penalidades, salvo casos em que podemos hipotetizar uma possibilidade a ser averiguada em pesquisas futuras.

Também encontramos algo a ser tratado com cuidado durante a leitura e coleta dos dados no *corpus* escrito e no *corpus* falado, pois os nomes repetidos não eram simplesmente repetidos sempre tal qual as pesquisas psicolinguísticas estudavam, mas poderia ocorrer de ser repetido de forma idêntica ou quase idêntica ou, ainda, ser usado um nome recategorizado, ou seja, um sinônimo ou um hiperônimo, por exemplo.

Por conta disto, tivemos de também separar a macrocategoria dos nomes em duas: os nomes repetidos e os nomes recategorizados. E, a fim de averiguar a quantidade de ocorrência de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função anafórica na posição de sujeito, chegamos a duas interpretações possíveis: (i) a ocorrência de nomes – incluindo os nomes recategorizados, pronomes plenos e pronomes nulos; e, (ii) a ocorrência de nomes repetidos – excluindo os nomes recategorizados, pronomes plenos e pronomes nulos.

Nos dados do *corpus* escrito, para (i), de todas as retomadas analisadas, 51% foram feitas pelo uso de nomes como forma correferente, 15% pelo uso de pronomes plenos e 34% pelo uso de pronomes nulos; e, para (ii), de todas as retomadas analisadas, 35,1% foram feitas pelo uso de nomes repetidos, 19,7% pelo uso de pronomes plenos e 45,2% pelo uso de pronomes nulos.

Nos dados do *corpus* de fala, para (i), de todas as retomadas analisadas, 43,1% foram feitas pelo uso de nomes como forma correferente, 31,6% pelo uso de pronomes plenos e 25,3% pelo uso de pronomes nulos; e, para (ii), de todas as retomadas analisadas, 30,1% foram feitas pelo uso de nomes repetidos, 38,9% pelo uso de pronomes plenos e 31% pelo uso de pronomes nulos.

Na busca de melhor compreender como as formas correferentes estão distribuídas, optamos por analisar os dados com os valores de retomada dos nomes repetidos, sem contabilizar as retomadas com nomes recategorizados, e comparar com a quantidade de ocorrências de pronomes plenos e de pronomes nulos.

Encontramos que, na escrita, 64% das retomadas analisadas foram feitas pelo uso de nomes repetidos e 36% pelo uso de pronomes plenos. Encontramos, também, na escrita que, 43,9% das retomadas analisadas foram feitas pelo uso

de nomes repetidos e 56,1% pelo uso de pronomes nulos, o que aponta para o uso mais frequente de pronomes nulos do que de pronomes plenos na escrita.

Já na fala, encontramos 43,6% das retomadas analisadas feitas pelo uso de nomes repetidos e 56,4% pelo uso de pronomes plenos, o que indica o uso mais frequente de pronomes plenos na fala. Encontramos, também, na fala que, 49,2% das retomadas analisadas feitas pelo uso de nomes repetidos e, 50,8% pelo uso de pronomes nulos, o que mostra que os pronomes nulos são menos frequentes na fala.

O que podemos inferir, de fato, é que há dois comportamentos muito distintos dos pronomes em PB: (i) a predominância do pronome nulo na escrita e, (ii) a predominância do pronome pleno na fala.

No que trata especificamente sobre os nomes recategorizados, observamos que, de modo geral, o uso deste tipo de nome tende a concorrer com o uso dos pronomes plenos na escrita, o que nos levou a acreditar que, na mente do escritor/falante da língua, há duas formas básicas para retomar os antecedentes: os nomes e os pronomes nulos.

O que parece acontecer é que, os nomes podem ser retomados pelos mesmos nomes ou por outras formas correferentes ao nome, o que implica nos nomes recategorizados, enquanto os pronomes nulos podem ser retomados por pronomes plenos.

De fato, o que acontece na escrita é uma concorrência entre o nome recategorizado e o pronome pleno para ocupar a posição sintática de sujeito, enquanto, na fala, mesmo que esta concorrência ocorra, o uso do pronome pleno é aceitável em qualquer contexto.

Como acreditamos que o uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função anafórica em posição de sujeito depende, na verdade, do contexto no qual a retomada ocorre, trabalhamos com a possibilidade de a distância entre o antecedente e a retomada ter influência direta sobre a forma correferente que o usuário da língua usará para preencher o centro anafórico<sup>220</sup>.

---

<sup>220</sup> É importante levar em consideração que os trabalhos de Vasconcelos e Leitão (2012) e de Alves (2012) em indivíduos com patologia e com idosos já mostravam um direcionamento para o uso de nomes, o que sugere contextos nos quais o nome é a melhor forma correferente a ser processada. Não utilizamos diretamente estes resultados por não termos textos produzidos por sujeitos com as mesmas especificidades, o que pode mostrar indícios distintos se comparados nossos resultados aos destes pesquisadores.

Esta possibilidade nos veio à mente pelo fato de estranharmos que, em nenhuma das pesquisas citadas, o nome foi a forma mais fácil de ser processada, pois tanto na RPN quanto na OPP, os pronomes são as formas mais fáceis de serem processadas e o nome é sempre penalizado.

Acreditando que a distância influencia no tipo de retomada, optamos por quantificar as palavras entre o antecedente e a retomada, já que, segundo Leitão e Simões (2008), se houver até 38 sílabas entre o antecedente e a retomada, o uso do pronome pleno gera efeito de penalidade sobre o nome e, segundo Lima (2015), a distância sintática entre o antecedente e o referente não tem influência no processamento correferencial.

Para termos resultados que nos possibilitassem inferências mais precisas, dividimos a quantidade de palavras em intervalos e analisamos quando o antecedente e a retomada estão distantes entre si por (i) até 10 (dez) palavras, (ii) de 11 (onze) a 50 (cinquenta) palavras, (iii) de 51 (cinquenta e uma) a 100 (cem) palavras e (iv) mais de 100 (cem) palavras.

Encontramos evidências muito distintas entre como as formas correferentes são usadas quando levadas em consideração a quantidade de palavras entre o antecedente e a retomada, sendo preciso compreender como ocorre a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em posição de sujeito em PB na escrita e na fala.

Na escrita, quando há até 10 (dez) palavras entre o antecedente e a retomada, o pronome nulo é a forma mais utilizada. Já na distância de 11 (onze) a 50 (cinquenta) palavras, o pronome pleno se torna bastante frequente, mas concorre com o pronome nulo na escrita, quando levado em consideração a quantidade de palavras, e o nome recategorizado surge como um concorrente para o pronome pleno em quantidade de ocorrências. Na distância de 51 (cinquenta e uma) a 100 (cem) palavras entre o antecedente e a retomada, o uso do pronome nulo diminui exponencialmente, enquanto o pronome pleno e o nome recategorizado são as formas preferencialmente usadas, despontando, também, o uso mais regular do nome repetido. E, quando a distância é acima de 100 (cem) palavras entre o antecedente e a retomada, o nome repetido é a forma mais frequente, na escrita.

Na fala, o padrão de distribuição é bastante peculiar, pois, ao mesmo tempo que parece seguir o padrão encontrado na escrita, vemos o uso do

pronome pleno surgir como frequente e ser viável em qualquer distância, enquanto os nomes recategorizados são menos frequentes do que na escrita e os nomes repetidos recorrentes. Além disso, notamos que, na fala, quando há até 10 (dez) palavras entre o antecedente e a retomada, apesar de o uso do pronome nulo ser mais frequente, o nome repetido também é regularmente usado. Já de 11 (onze) a 50 (cinquenta) palavras, de acordo com os textos analisados, o pronome pleno e nome recategorizado são usados, sendo preferencialmente usado o pronome pleno, mas, os pronomes nulos e os nomes recategorizados também aparecem nesta distância. Já quando há, entre o antecedente e a retomada, de 51 (cinquenta e uma) a 100 (cem) palavras, o pronome nulo é pouquíssimo usado.

Estes resultados levam a indícios já apontados por Leitão e Simões (2011), mas não encontrados por Lima (2015), pois percebemos que a distância, em quantidade de palavras, entre o antecedente e a retomada tem influência direta sobre como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos são usados enquanto formas correferentes, pois, quanto menor a distância, mais frequente é o uso do pronome nulo e, quanto maior a distância, mais frequente é o uso do nome repetido, o que pode mostrar uma possível influência das informações na memória de trabalho (DANEMAN; CARPENTER, 1980; CIASCA, 2003; ALLOWAY, 2006; DEJONG, 2006; ALMOR, 1999; SOUSA; SISTO, 2011; KEREMER, 2014; BADDELEY, 2017) sobre a forma como usamos as expressões correferentes.

Seguimos nossa pesquisa analisando como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos em função anafórica na posição de sujeito são usados quando levado em consideração a disposição textual em orações e em parágrafos.

Por conta disto, analisamos quatro contextos de retomada: a retomada em (i) orações subordinadas, em (ii) orações coordenadas justapostas, em (iii) orações coordenadas em períodos distintos e em (iv) parágrafo subsequente.

A retomada de um antecedente por meio de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função anafórica na posição de sujeitos em contextos específicos ao longo do texto nos deu evidências bastante interessantes.

Vimos que, na escrita, as retomadas em orações subordinadas tendem a ser preferencialmente feitas por pronomes nulos, neste contexto e que, o uso

dos pronomes nulos dá espaço para o aparecimento dos pronomes plenos, para o preenchimento do centro anafórico na posição de sujeito de orações coordenadas justapostas.

Já nas orações coordenadas em períodos distintos, vimos que há tendência de a retomada ser feita por pronomes plenos ou por nomes recategorizados, enquanto a primeira retomada em parágrafo subsequente é, preferencialmente, feita por nomes repetidos.

Encontramos, ao longo deste trabalho, que a distância entre o antecedente e a retomada, em quantidade de palavras, mostra que, quanto menor a distância, mais frequente é o uso do pronome nulo e que, inversamente, quanto maior a distância, mais frequente é o uso do nome repetido.

Se levarmos em consideração que, geralmente, o sujeito da oração subordinada está próximo do antecedente, entendemos a grande quantidade de ocorrências de pronomes nulos neste contexto. Por outro lado, se levarmos em consideração os parágrafos do texto e a última vez que o referente foi usado em dado parágrafo, ao mudar de parágrafo e analisar a primeira retomada na posição de sujeito, geralmente, há uma grande distância entre o antecedente e a retomada em quantidade de palavras, o que explica a grande quantidade de ocorrências de nomes repetidos neste contexto.

Como em orações coordenadas justapostas, ligadas por vírgulas e/ou por conjunções, pode acontecer de o antecedente estar muito ou medianamente próximo da retomada, temos a explicação da ocorrência regular de pronomes nulos e do aumento na quantidade de ocorrências de pronomes plenos e, levando em consideração que os pronomes plenos e os nomes recategorizados tendem a estar medianamente afastados de seus antecedentes em quantidade de palavras, fica explicado o aumento de ocorrências destas duas formas em orações coordenadas em períodos distintos.

Já na fala, o padrão parece ser menos rígido do que na escrita, pois, levando em consideração os mesmo quatro contextos, notamos que as formas correferentes são mais soltas, no sentido de que podem ocorrer de forma regular em quase todos os contextos.

Vimos que, em orações subordinadas, há tendência ao preenchimento do sujeito por pronomes nulos, mas que os pronomes plenos também aparecem com frequência, o que pode indicar que não há prejuízo para o usuário da língua

em retomar o sujeito de uma oração subordinada por meio de pronome nulo ou de pronome pleno na fala.

Nas orações coordenadas justapostas, os pronomes nulos passam a ser menos usados e a quantidade de ocorrências de pronomes plenos se destaca, sendo seguida pelo uso de nomes recategorizados retomando sujeitos neste contexto.

Já quando analisamos as orações coordenadas em períodos distintos, vimos que os pronomes plenos e os nomes repetidos começam a ocorrer regularmente. Neste contexto, o uso de nomes recategorizados também é frequente, enquanto o uso do pronome nulo cai drasticamente. Já na primeira retomada em parágrafo subsequente, o uso do nome repetido é frequente.

O que percebemos, portanto, é que, assim como na escrita, há influência da distância em quantidade de palavras e a forma correferente usada em posição de sujeito, pois, quanto menor for a distância em quantidade de palavras entre o antecedente e a retomada, maior é a frequência de uso dos pronomes nulos e, inversamente, quanto maior for a distância em quantidade de palavras entre o antecedente e a retomada, maior é a frequência de usos de nomes repetidos.

Sobre a fala, é interessante destacar alguns pontos: (i) o uso dos pronomes nulos é baixo; (ii) o uso dos pronomes plenos é frequente em qualquer distância entre o antecedente e a retomada; (iii) os nomes recategorizados são menos frequentes na fala do que na escrita.

Com relação à frequência menor do uso de nomes recategorizados na fala do que na escrita, não tivemos nenhuma surpresa, pois é do conhecimento comum que, durante a vida escolar, sobretudo no que diz respeito à escrita, as pessoas são treinadas a não repetir palavras em textos e, isto, acaba sendo refletido na frequência de palavras distintas para retomar o mesmo termo.

É comum um aluno de nível fundamental escrever um texto falando sobre uma entidade no mundo, como *Brasil*, por exemplo, e escrever várias vezes o mesmo nome. Os professores, por sua vez, começam uma tarefa de ensinar o discente a substituir a palavra *Brasil* por *nação*, *estado*, *país*, entre outras e, assim, fazendo-o evitar a repetição do mesmo termo.

Já na fala, mesmo que as pessoas conheçam nomes recategorizados para retomar o seu referente, elas não têm as regras rígidas que a escrita

apresenta e não se sentem monitoradas o tempo todo a estarem corretas e não cometer desvios com relação à norma padrão.

Este mesmo fator poderia explicar, até certo ponto, o uso frequente de pronomes plenos em qualquer distância em quantidade de palavras entre o antecedente e a retomada, pois, enquanto na escrita há tempo para pensar nas melhores formas para indicar aquilo que se pretende dizer, na fala, o monitoramento é mais flexível e, por diversos fatores que não são do escopo de nosso trabalho, é sabido que as pessoas têm tido preferência ao preenchimento do sujeito foneticamente realizado na posição de sujeito.

De qualquer forma, podemos perceber que há um padrão que as pessoas seguem, mesmo que mentalmente, ao usar os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos em Português Brasileiro e que, este padrão, depende inicialmente de dois fatores que estão diretamente ligados ao primeiro estágio que Van Gompel, Liversedge e Pearson (2004) propôs: o estágio sintático.

Queremos dizer com isso que acreditamos no fato de que, quando as pessoas fazem as escolhas das formas correferentes que serão usadas, elas levam em consideração dois fatores: a distância em quantidade de palavras entre o antecedente e a retomada e a disposição textual com relação ao fato de a retomada ocorrer em oração subordinada, em oração coordenada justaposta, em oração coordenada em períodos distintos ou se é a primeira retomada em parágrafo subsequente.

Em nossa busca por explicações sobre a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função anafórica em PB, acreditamos ter encontrado indícios de dois fatores que podem ser considerados em pesquisas que tratam dos efeitos de penalidade: (i) a quantidade de palavras entre o antecedente e a retomada e (ii) a disposição textual.

Ao longo de nossa pesquisa, analisamos também outros fatores que poderiam influenciar na resolução correferencial, como o tipo de verbo que ocorre na sentença anterior à sentença da retomada e a definitude do referente, refletida por meio do artigo que precede o antecedente dentro do sintagma nominal.

Sobre o tipo de verbo que ocorre na sentença anterior à sentença da retomada não encontramos nada que pudesse indicar uma influência direta deste fator na escolha da forma correferente.

A única constatação que tivemos foi de que, aparentemente, os verbos transitivos são os verbos mais frequentes em PB tanto na escrita quanto na fala e que, na fala, há tendência para o uso de verbos de ligação com mais frequência do que na escrita.

Sobre a definitude do antecedente, fator que pode ter algum tipo de influência na resolução correferencial, não no primeiro estágio, mas no segundo, o estágio semântico-pragmático (VAN GOMPEL; LIVERSEDGE; PEARSON (2004), percebemos que é muito natural retomar um antecedente definido.

Isto é refletido na grande quantidade de ocorrências do artigo definido precedendo o antecedente, enquanto o artigo indefinido é menos frequentemente utilizado para introduzir uma entidade no texto.

Como analisamos somente os artigos definidos e os artigos indefinidos, sendo as demais possíveis palavras que precedem o antecedente classificadas como outras formas, não podemos afirmar que o fator definitude tem influência direta sobre o processamento correferencial, mas indicamos que, provavelmente, de algum modo, o fato de o antecedente ser definido no mundo pode ter algum papel na escolha das formas correferentes que as pessoas usam ao longo do texto.

Desta forma, chegamos ao final da discussão sobre os resultados que nossas categorias de análise nos deram, levando-nos a, em discussão com pesquisas feitas anteriormente sobre o processamento correferencial, chegar a novas descobertas, porém, antes de dar prosseguimento às demais teorias e hipóteses que nortearam nosso trabalho, precisamos falar sobre a metodologia das pesquisas psicolinguísticas que levam aos resultados encontrados.

A questão de como os resultados são analisados vêm ganhando forma, mas foi Gondim (2017) que buscou, por meio de uma análise multifatorial, observar se a maneira como os dados eram analisados poderia influenciar, de algum modo, no resultado final dos dados coletados em experimentos psicolinguísticos.

Segundo Gondim (2017), dentre alguns fatores estudados, o tempo de processamento da região ou segmento crítico – local onde ocorre a retomada –

ou o tempo de processamento de toda a sentença, não interfere diretamente no resultado.

Ao analisarmos as pesquisas que nortearam nosso estudo, vimos que, na maioria, os *designs* experimentais tratam de orações coordenadas em períodos distintos, mas muito próximas em quantidade de palavras, não chegando a ultrapassar 20 (vinte) palavras entre o antecedente e a retomada.

Levando em consideração o que propomos, encontramos, nestas pesquisas, o seguinte contexto: distância curta em quantidade de palavras entre o antecedente e a retomada e orações coordenadas em períodos distintos, o que poderia propiciar o surgimento do pronome nulo, no primeiro caso, e o aparecimento do pronome nulo, no segundo.

Fazendo uma análise das sentenças estudadas nestas pesquisas, é notório que, apesar de as sentenças estarem separadas por ponto final, o que faz com que sejam classificadas como orações coordenadas em períodos distintos em nosso trabalho, elas poderiam estar perfeitamente ligada por conjunção, o que, na mente do leitor, pode ter gerado a configuração contextual de pequena quantidade de palavras entre o antecedente e a retomada e oração coordenada justaposta, o que, na escrita, de acordo com nossos resultados, propicia o aparecimento do pronome nulo.

Entretanto, como também propomos em nosso trabalho, se o usuário da língua cogita uma forma correferente a ser usada e esta forma não está disponível por algum motivo, a forma subsequente será usada até que se esgotem as possibilidades, se é que isso seria possível.

Desta forma, se o contexto propicia a retomada por pronome nulo e este não está compondo as sentenças lidas, a próxima forma mais frequente de uso nesse contexto é o pronome pleno e, a última possibilidade, seria o nome repetido, para nós, explica os resultados encontrados nas pesquisas psicolinguísticas sobre a correferência anafórica em PB.

Outro ponto importante e que não poderia deixar de estar em nossa discussão é o fato de que, por ser comum o uso de sentenças em experimentos psicolinguísticos, o termo *texto* e o uso de blocos textuais ou de textos inteiros é, silenciosamente, recriminado.

Entretanto, levado por conceitos mais atuais, entendemos que texto é qualquer produção que passe uma informação (KOCH, 2015). Por esta concepção, uma placa de trânsito é um texto; um espetáculo de dança ou um

filme de cinema mudo é um texto; um gesto em dado contexto é um texto; uma sentença é um texto.

Advogamos que futuras pesquisas experimentais possam fazer uso de textos longos, observando uma cadeia correferencial para, a partir de então, temos mais indícios não somente da distribuição das formas correferentes, mas também, de como nossa mente é capaz de processá-las, uma vez que nosso estudo nos mostra padrões a partir de frequência, dando-nos, quando muito, hipóteses do que poderia ocorrer na mente do falante.

Feitas as considerações que achamos pertinentes, aproveitamos para comentar que compartilhamos da crença de que o processamento correferencial passa por dois estágios, conforme proposto por Van Gompel, Liversedge e Pearson (2004).

Estes autores, mostraram que, durante o processamento correferencial, feitas todas as ligações sintáticas necessárias entre o antecedente e o referente, entra em questão o estágio no qual o antecedente é verificado semanticamente, levando em consideração o número, o gênero e a animacidade/inanimacidade, por exemplo.

Acreditamos, também, que o uso das formas correferenciais, sobretudo a possibilidade de retomada por nomes recategorizados, gera diferentes nuances no processamento correferencial por conta da carga informacional que cada forma usada tem e, assim, compartilhamos da crença de que Almor (1999) estava certo ao prever que existem formas que demandam mais esforço cognitivo para ser reconhecida como correferente, podendo, em alguns casos, gerar prejuízos para a compreensão global do texto.

Por fim, acreditamos ter encontrado resultados que nos levam a corroborar a Teoria da Acessibilidade (ARIEL, 1991), pois, de acordo com a autora, há um padrão universal, que pode variar entre as línguas, para como as formas correferenciais são acessadas na mente do usuário, variando de formas mais acessíveis a formas menos acessíveis.

De acordo com Ariel (1991), os pronomes nulos são ligados a antecedentes altamente acessíveis, enquanto os nomes tendem a retomar antecedentes que estão menos acessíveis no texto.

Em nosso trabalho, encontramos o que consideramos dois indícios de que Ariel (1991) estava certa em sua escala de acessibilidade, pois vimos que os pronomes nulos, na escrita, retomam antecedentes altamente acessíveis por

estarem próximos em quantidade de palavras, enquanto os nomes repetidos são usados por conta da longa distância de seus antecedentes que, ao longo do texto, têm ficado menos acessíveis. Já na fala, vimos que o mesmo padrão se mantém, feita a ressalva de que os pronomes nulos e os pronomes plenos têm, quando estão próximos de seus antecedentes, o mesmo comportamento, ou seja, servem para retomar um antecedente altamente acessível.

Acreditamos que, quando Ariel (1991) pressupôs que em algumas línguas, a escala de acessibilidade pudesse variar, indiretamente era previsto casos como o do PB que, por ser uma língua em que o parâmetro *pro-drop* está em oscilação, tem formas que estão tendo comportamentos inesperados, como é o caso do pronome pleno na fala que tem o mesmo papel que um pronome nulo, um nome recategorizado ou um nome repetido.

E, assim, chegamos ao fim de nossa discussão geral, na qual buscamos correlacionar o máximo possível as teorias citadas ao longo de nosso segundo capítulo, destacando as pesquisas psicolinguísticas que tratam sobre o processamento correferencial em PB.

Na busca por explicação de como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos em função anafórica na posição de sujeito são distribuídos em PB, fizemos a análise de diversos fatores sintáticos que pudessem influenciar nas formas correferentes que os usuários da língua preferem usar e, assim, encontrar possíveis padrões de uso destas formas e lançar luz sobre possíveis parâmetros que possam ser respeitados em pesquisas futuras que tratem deste assunto.

Após empreendermos pesquisas na busca de explicar como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos estão distribuídos em PB em dados de *corpus* escrito e em dados de *corpus* de fala, esperamos ter cumprido nossos objetivos e ter feito discussões esclarecedoras sobre nosso objeto de estudo e que, com este trabalho, tenhamos contribuído com novos achados sobre a correferência anafórica em PB.

## 7 CONCLUSÕES

O estudo que desenvolvemos, nesta tese, teve como objetivo explicar a distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos com função anafórica em posição de sujeito em Português Brasileiro, a fim de averiguar se efeitos encontrados em pesquisas psicolinguísticas são notados em dados de *corpus* escrito e em dados de *corpus* de fala.

Conforme descrito ao longo deste trabalho, o estudo da correferenciação tem sido feito em Psicolinguística por meio de pesquisas que visam explicar como o usuário da língua faz, desde ligações entre o antecedente e a forma correferente (CHOMSKY, 1981), até possíveis penalidades entre nomes repetidos, pronomes plenos e pronomes nulos (GORDON; GROSZ; GUILLIOM, 1993; GERLOMINI-LEZAMA, 2008).

Segundo Gordon, Grosz e Guilliom (1993), quando uma entidade é introduzida na posição de sujeito, ela é naturalmente retomada por um pronome pleno também na posição de sujeito, sendo o nome repetido a forma mais difícil de ser compreendida, dando a este efeito o nome de Penalidade do Nome Repetido.

Gerlomini-Lezama (2008), aplicando o mesmo parâmetro – entidade introduzida em posição de sujeito sendo retomada por forma correferente também em posição de sujeito – observou que, em uma língua *pro-drop* como o Espanhol Argentino, o pronome nulo é mais fácil de ser processado do que o pronome pleno e do que o nome repetido, dando a este efeito o nome de Penalidade do Pronome Pleno.

Em ambas as pesquisas, percebemos o destaque para a compreensão da retomada feita na posição de sujeito da sentença, o que decorre dos estudos de Gordon e Hendrick (1997), no qual os autores afirmam que o sujeito é a forma mais proeminente da sentença e que tende a ser retomada também em posição de sujeito, dando às pesquisas psicolinguísticas o conceito de Proeminência Sintática.

Com base neste conceito, Chambers e Smith (1998) desenvolveram estudos, a fim de comprovar que os antecedentes proeminentes na sentença são mais facilmente processados na mente do leitor quando são retomados

também na posição de sujeito, ou seja, em posições paralelas, dando a este efeito o nome de Paralelismo Sintático.

Porém, nestas pesquisas, não encontramos comentários sobre como os autores entendem as ligações entre os antecedentes e os correferentes ou quais são os estágios tanto sintáticos quanto semânticos ou até pragmáticos que poderiam nos fazer compreender melhor a correferência anafórica.

Buscamos compreender melhor o comportamento de *pro*, a categoria vazia que é a contraparte do pronome foneticamente realizado, ou seja, o pronome nulo, uma das formas correferentes por nós analisadas, além de entender melhor sobre como o parâmetro *pro-drop* é configurado e o que leva uma língua a ser classificada como *pro-drop* ou não *pro-drop*.

Van Gompel, Liversedge e Pearson (2004), baseados nos estudos de Garrod e Sanford (1977) propõem que o processamento correferencial tem dois estágios, um estágio sintático e outro semântico.

Além disso, ainda em relação à semântica, vimos que Almor (1999) estudou como as formas correferentes podem ter nuances distintas de informação e, assim, influenciar de alguma maneira no processamento anafórico, pois é sabido que o mesmo nome sendo repetido não tem a mesma carga informacional de um hiperônimo ou de um hipônimo, por exemplo.

E, por fim, vimos que, segundo Ariel (1991), todas as línguas têm, em graus variados, um *padrão universal de acessibilidade* entre o antecedente e a forma correferente, dentro das quais, os pronomes nulos retomam entidades altamente acessíveis e, na direção oposta, os nomes retomam entidades menos acessíveis no texto.

As pesquisas que tratam sobre o processamento correferencial, em PB, tem se debruçado em analisar se os efeitos da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008) são produtivos nesta língua e, por meio de técnicas experimentais de monitoramento de tarefas de leitura, têm mostrado que tanto este efeito (MAIA; CUNHA-LIMA, 2011, 2012, 2014; MAIA, 2014; LIMA, 2015) quanto aquele (LEITÃO, 2005; QUEIROZ; LEITÃO, 2008; LEITÃO; PEIXOTO; SANTOS, 2008; LEITÃO; SIMÕES, 2011; LEITÃO; RIBEIRO; MAIA, 2012; VASCONCELOS; LEITÃO, 2012; SIMÕES, 2014; LIMA, 2014; GONDIM, 2017) existem em PB, mas não foram encontrados contextos em que os dois efeitos co-ocorressem.

Os demais pressupostos encontrados em pesquisas psicolinguísticas como o efeito da Proeminência Sintática (GORDON; HENDRICK, 1997, LEITÃO, 2005a) e do Paralelismo Sintático (CHAMBERS; SMITH, 1998) são utilizados na maioria destas pesquisas, pois já são aceitos como efeitos reais das línguas e que influenciam diretamente em como o usuário da língua escolhe a forma correferente.

Entretanto, em meio a tantos resultados quase alinhados, encontramos duas questões fortemente marcantes: (i) Como os nomes nunca foram reportados sendo formas correferentes boas, ou seja, mais fáceis de serem processadas? Isto significa dizer que os nomes não devem funcionar como correferentes porque geram prejuízo à compreensão do texto?

Estas questões nos moveram para, então, fazer a análise de como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos em função anafórica em posição de sujeito estão distribuídos em PB para, a partir desta análise, averiguar, pautados em frequência de uso, como estas formas estão distribuídas em PB.

Esperávamos encontrar, em dados de *corpus* escrito e em dados de *corpus* de fala, contextos nos quais os nomes seriam tão frequentes quanto as demais formas, apontando para uma gradação de uso a depender do contexto sintático.

Como pesquisas anteriores (LIMA, 2015; ALMOR *et al.*; 2017, GONDIM, 2017) deram indicações de que o contexto pode influenciar no surgimento de uma penalidade na língua, decidimos elencar fatores constantemente usados em pesquisas psicolinguísticas em PB, como o tipo de oração em que a retomada é feita, o tipo de verbo que ocorre na sentença anterior à retomada e o tipo de artigo que precede o antecedente.

Levando em consideração os fatores acima, partimos para uma pesquisa em *corpora* linguísticos contabilizando, também, a quantidade de sujeitos que foram retomados em posições de sujeito (proeminentes) ou em posições de objeto, de adjunto adnominal ou de adjunto adverbial (não proeminentes).

Desta forma, esperávamos encontrar padrões de uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em posição de sujeito em PB e, assim, esclarecer alguns pontos que poderiam vir a ser respeitados em futuros experimentos psicolinguísticos sobre o processamento correferencial.

A fim de chegar a resultados que pudessem responder às perguntas que guiaram o nosso trabalho, desenvolvemos a coleta e análise de dados em dois *corpora* distintos: um *corpus* escrito e um *corpus* de fala.

Em nossa primeira pesquisa, analisamos 10 (dez) texto do *corpus* escrito Lácio-Ref e, através de preenchimento de formulários para computar a quantidade de ocorrências, posteriormente transcritos em planilhas e apresentados em formas de gráficos com os resultados, analisamos diversas categorias que nos levaram às inferências discutidas nos capítulos anteriores.

Os dados do *corpus* escrito nos mostraram que há dois fatores que propiciam o surgimento de uma forma correferente em vez de outra: a distância em quantidade de palavras entre o antecedente e a retomada e a disposição textual.

A distância em quantidade de palavras pode mostrar indícios da influência que a memória de trabalho (DANEMAN; CARPENTER, 1980; CIASCA, 2003; ALLOWAY, 2006; DEJONG, 2006; ALMOR, 1999; SOUSA; SISTO, 2011; KEREMER, 2014; BADDELEY, 2017) tem sobre a forma anafórica que usamos em textos escritos e falados.

Encontramos evidências de que os pronomes nulos, na escrita, retomam os antecedentes que estão próximos em até 10 (dez) palavras. Na direção oposta, encontramos evidências de que os nomes repetidos, na escrita, retomam os antecedentes que estão distantes em mais de 100 (cem) palavras da forma correferente.

Ainda sobre a distância em quantidade de palavras entre o antecedente e a retomada, encontramos evidências de que, na escrita, os pronomes plenos e os nomes recategorizados funcionam bem em distâncias medianas, sendo os pronomes plenos mais frequentes quando há, entre o antecedente e a retomada, de 11 (onze) a 50 (cinquenta) palavras, enquanto os nomes recategorizados são mais frequentes quando há entre 51 (cinquenta e uma) e 100 (cem) palavras entre o antecedente e a retomada, mas, nestas distâncias, apontando para um padrão de uso das formas mais frequentes para as formas menos frequentes com relação à quantidade de palavras.

Percebemos que, em quantidade de palavras, há um padrão de distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB, já que quanto menor a distância em quantidade de palavras, mais frequente é o uso do pronome nulo e, quanto

maior a distância em quantidade de palavras, mais frequente é o uso do nome repetido, enquanto pronomes plenos e nomes recategorizados são mais frequentes em distâncias medianas.

Em seguida, vimos que, em termos de disposição textual, na escrita, o uso dos pronomes nulos é bastante frequente na posição de sujeito de orações subordinadas, enquanto os pronomes nulos e os pronomes plenos surgem quase na mesma frequência como sujeitos de orações coordenadas justapostas. Já o uso do pronome nulo cai drasticamente quando analisadas as retomadas em posição de sujeito de orações coordenadas em períodos distintos, sendo tanto os pronomes plenos quanto os nomes recategorizados as formas mais frequentes, nestes contextos.

Quando analisadas as primeiras retomadas em parágrafo subsequente, os resultados mostraram que o uso dos nomes repetidos foi bastante frequente.

Logo, se levarmos em consideração a distribuição das retomadas em posição de sujeito com relação à disposição textual, não fica difícil de interligar ao fator de quantidade de palavras entre o antecedente e a retomada a alta frequência de nomes como forma correferente.

Esta interação nos levou hipotetizar que, durante a escolha da forma correferente a ser usada, no estágio sintático, os usuários da língua fazem cálculos mentais e averiguam as possibilidades de retomadas levando em consideração a distância em quantidade de palavras entre o antecedente e a retomada e o tipo de oração/parágrafo em que ocorre a retomada.

Encontramos, também, indícios, por meio da frequência de ocorrências, que um antecedente em posição de sujeito tende a ser preferencialmente retomado em posição de sujeito em vez de ser retomado em posições não proeminentes, o que corrobora que, na língua escrita, tanto o efeito da Proeminência Sintática (GORDON; HENDRICK, 1997; LEITÃO, 2005a) quanto o efeito do Paralelismo Sintático (CHAMBERS; SMITH, 1998) podem ser percebidos.

Entretanto, como encontramos antecedentes introduzidos em posições não proeminentes, optamos por observar se havia um padrão de distribuição do uso de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos entre estas posições. Percebemos que um antecedente tende a ser retomado em posição de objeto mais frequentemente do que em posição de adjunto adnominal e esta, por sua vez, é mais frequente do que a retomada em

posição de adjunto adverbial, o que indica uma escala de posições em que a introdução de referentes ocorre com mais frequência.

Isto nos levou a aceitar ainda mais que a escolha da forma correferente depende de contextos no qual a forma correferente é usada, pois acreditamos que um antecedente pode ser retomado, sem prejuízos para a produção textual, tanto em posição proeminente quanto em posição não proeminente, o que nos leva a corroborar Leitão (2005) e Queiroz (2009), mostrando, assim, que são necessárias mais evidências sobre efeitos de penalidade quando não há paralelismo sintático para, futuramente, serem propostas reformulações no conceito da Penalidade do Nome Repetido (GORDON; GROSZ; GUILLIOM, 1993) e da Penalidade do Pronome Pleno (GERLOMINI-LEZAMA, 2008).

Além destes fatores, analisamos a ocorrência de verbos intransitivos, verbos transitivos e verbos de ligação na sentença anterior à sentença em que ocorre a retomada e percebemos que há um padrão de uso destes verbos, sendo os verbos transitivos mais frequentes do que os verbos intransitivos e estes, por sua vez, mais frequentes do que os verbos de ligação, mas não encontramos nenhuma evidência que nos levasse a acreditar que o tipo de verbo possa ter influência direta durante o processamento correferencial.

Por fim, observamos a frequência de uso de artigo definido e de artigo indefinido precedendo o antecedente, o que é um fator semântico que indica características como a definitude ou a indefinitude e animacidade ou inanimacidade do referente. Como este não é um fator central de nossa pesquisa, não nos atemos a isto, restando-nos somente a dizer que os antecedentes definidos tendem, substancialmente, a ser mais retomados do que os antecedentes indefinidos. Entretanto, alertamos que mais pesquisas que tratem da animacidade/inanimacidade do referente são necessárias, pois acreditamos que este fator semântico tem influência no processamento correferencial.

Nossa segunda coleta de dados, em *corpus* de fala, mostrou-nos que, até certo ponto, nossas interpretações sobre o padrão de distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB estavam corretas, pois confirmamos resultados quase idênticos sobre o efeito do Paralelismo Sintático (CHAMBERS; SMITH, 1998) e sobre o efeito da Proeminência Sintática (GORDON; HENDRICK, 1997), bem como os resultados sobre a ocorrência dos tipos de verbos da sentença anterior

à sentença em que ocorre a retomada, pois é percebido o mesmo padrão além de, na fala, os artigos definidos ocorrerem, substancialmente, mais vezes do que os artigos indefinidos, o que nos levou a reforçar nossas previsões sobre a possível influência do fator animacidade/inanimacidade.

Assim como nos dados do *corpus* escrito, vimos que, na fala, a quantidade de palavras entre o antecedente e a retomada tem influência sobre a forma correferente usada e, além disso, encontramos indícios de que o mesmo padrão da escrita ocorre, pois os pronomes nulos tendem a retomar antecedentes próximos em quantidade de palavras e os nomes repetidos tendem a retomar antecedentes mais distantes, enquanto os nomes recategorizados, ocorrendo substancialmente em menor quantidade do que na escrita, surgem quando há quantidade intermediária de palavras entre o antecedente e a retomada.

O que encontramos de diferente entre o padrão do *corpus* escrito e o padrão do *corpus* de fala foi o fato de que, independentemente da distância em quantidade de palavras, o pronome pleno surge frequentemente.

É interessante notar que, em termos de frequência, em distâncias curtas, na fala, o pronome nulo é mais frequente do que o pronome pleno, o que vai na direção oposto dos resultados encontrados nos dados do *corpus* escrito.

Por outro lado, vimos que, na fala, assim como ocorre na escrita, quando há mais de 100 (cem) palavras entre o antecedente e a retomada, o uso dos nomes repetidos é mais frequente.

Uma vez que analisamos a quantidade de palavras entre o antecedente e a retomada, analisamos o padrão de distribuição de nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos em função anafórica em PB em dados de fala, buscamos mais evidência desta distribuição de acordo com a disposição textual.

Encontramos, mais uma vez, um padrão muito parecido com o padrão da escrita, salvo para o comportamento dos pronomes plenos, pois, em sujeitos de orações subordinadas, há tendência de a retomada ser feita por meio de pronomes nulos com alternância de uso com os pronomes plenos que são frequentes nesta posição.

Em orações coordenadas, tanto justapostas quanto em períodos distintos, os sujeitos são retomados, na fala, majoritariamente, por pronomes plenos, mas há ocorrência tanto de nomes repetidos e nomes recategorizados quanto de pronomes nulos. Os pronomes nulos tendem a aparecer em orações

coordenadas justapostas, enquanto os nomes repetidos e os nomes recategorizados tendem a aparecer em orações coordenadas em períodos distintos. Já quando analisada a primeira retomada em parágrafo subsequente, percebemos que, na fala, há grande tendência de a retomada ser feita por nomes repetidos.

Além de nossas descobertas acerca dos fatores linguísticos envolvidos no processamento correferencial, ao longo de nossa tese, optamos por nos posicionar sobre as teorias que compõem o nosso segundo capítulo e, por conta disto, achamos necessário ressaltá-las pelo fato de esperarmos que, estudos futuros possam vir a corroborar ou não nossos achados, desde que obedeçam, durante a construção de possíveis *designs* experimentais, aos fatores que levaram aos efeitos que postulamos.

Por termos também tocado no conceito de língua *pro-drop*, temos indícios da quantidade de ocorrências de pronomes plenos e de pronomes nulo em posição de sujeito tanto na escrita quanto na fala, sendo que nesta forma da língua, há predominância dos pronomes plenos e, naquela, dos pronomes nulos, o que nos levou a aceitar que o PB é uma língua parcialmente *pro-drop*, conforme previram Holmberg, Nayudu e Sheehan (2009).

Ao tocarmos em questões de semântica, mesmo não sendo o objeto central de nosso trabalho, defendemos também a Teoria do Processamento em Dois Estágios (VAN GOMPEL; LIVERSEDGE; PEARSON, 2004) defendem que, quando há tarefas de processamento anafórico, são levadas em consideração questões de cunho sintático e, posteriormente, semânticas.

Somos também defensores de que, entre as formas correferentes, sobretudo na escolha dos nomes recategorizados quando funcionam como correferentes de nomes, há nuances distintas de cargas informacionais e, por conta disso, acreditamos que haja formas mais fáceis de serem processadas por conta da carga informacional que contêm, conforme postulado por Almor (1999).

E, por fim, encontramos indícios que também nos levam a corroborar a Teoria da Acessibilidade (ARIEL, 1991), pois, tal como dito pela autora, os pronomes nulos retomaram antecedentes altamente acessíveis e os nomes, antecedentes mais difíceis de serem acessados.

Além de termos encontrado estes indícios, percebemos ainda que, na fala, como os pronomes plenos retomam antecedentes em qualquer distância em PB, temos uma variação na escala de acessibilidade também prevista por

Ariel (1991) ao dizer que, mesmo havendo um *padrão universal de acessibilidade*, algumas línguas podem ter especificidades e acreditamos que o comportamento do pronome pleno em PB indique uma especificidade na Teoria da Acessibilidade, o que pode ser característico de uma língua em que o parâmetro *pro-drop* parece oscilar.

Os resultados encontrados nesta tese exigirão, naturalmente, a elaboração de novas pesquisas, bem como o constante debate com estudos psico e sociolinguísticos acerca do tema investigado. Apesar disso, este estudo oferece novas informações para o debate de como os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos em função anafórica na posição de sujeito são utilizados em Português Brasileiro.

Lançando diversas questões para investigações futuras temos, por exemplo, a necessidade de averiguar, por meio de pesquisas experimentais, se as distâncias delimitadas por nós têm influência no processamento anafórico e como os nomes repetidos, nomes recategorizados, pronomes plenos e pronomes nulos são processados em cada uma das disposições textuais por nós elencadas, fazendo o cruzamento entre disposição textual e quantidade de palavras entre o antecedente e a retomada.

Além disso, por acreditar que o efeito do Paralelismo Sintático (CHAMBERS; SMITH, 1998) não tem influência direta no processamento anafórico, mas ajuda as retomadas a serem mais facilmente processadas, entretanto, com encontramos indícios de retomadas não paralelas e não proeminentes, novas comprovações poderão ser encontradas, tal como encontradas por Leitão (2005a) e por Queiroz (2009), o que poderia apontar até para reformulações futuras na Teoria da Centralização (GROSZ; SIDNER, 1986; GROSZ; JOSHI; WEINSTEIN, 1995).

E, por fim, acreditamos que o uso de textos mais longos e do mapeamento de cadeias referenciais em pesquisas experimentais pode dar mais evidências sobre como a mente processa as retomadas anafóricas, nos mais diversos contextos, pois, tal como aponta Gondim (2017), acreditamos que diversos fatores estão envolvidos não somente no processamento correferencial, mas também na produção, tal como mostram as pesquisas em *corpora* linguísticos sobre o objeto nulo em PB.

Levando em consideração o que foi proposto neste trabalho e as pesquisas empreendidas, os resultados analisados e amplamente discutidos

com as pesquisas que nos levaram a este estudo, esperamos ter explicado como estão distribuídos os nomes repetidos, os nomes recategorizados, os pronomes plenos e os pronomes nulos em função anafórica na posição de sujeito em Português Brasileiro, mostrando padrões de contextos que podem vir a ser analisados futuramente e, assim, ter contribuído para o avanço de questões teórico-metodológicas importantes sobre o processamento correferencial.

## REFERÊNCIAS

ALBUQUERQUE, Guiomar Silva de. **Processamento da linguagem no déficit de atenção e hiperatividade**. Orientador: Marcus Antônio Rezende Maia. 2008. 148f. Tese (Doutorado em Linguística) – Pós-Graduação em Linguística, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.

ALLOWAY, T. P. How does working memory work in the classroom?. **Educational Research and reviews**, v.1, n.4, p.134-149, 2006.

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

ALMOR, Amit et al. Language processing, acceptability, and statistical distribution: A study of null and overt subjects in Brazilian Portuguese. **Journal of Memory and Language**, v. 92, p. 98-113, 2017.

ALVES, Giorvan Ânderson dos Santos. **Processamento correferencial em idosos com e sem doença de Alzheimer**. Orientador: Márcio Martins Leitão. Tese (Doutorado em Linguística) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba, João Pessoa, 2012.

ARAÚJO, Aluiza Alves de; VIANA, Rakel Beserra de Macedo; PEREIRA, Maria Lidiane de Sousa. O projeto Descrição do Português Oral Culto de Fortaleza – PORCUFORT: das origens aos dias atuais. **Web Revista SOCIODIALETO**, v. 8, p. 174-198, 2018.

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

ATKINSON, Richard. C.; SHIFFRIN, Richard. M. Human memory: A proposed system and its control processes. In: **Psychology of learning and motivation**. Academic Press, p. 89-195, 1968.

AXMACHER, Nikolai; RASCH, Björn (Ed.). **Cognitive neuroscience of memory consolidation**. Springer International Publishing, 2017.

BADDELEY, Alan. The episodic buffer: a new component of working memory?. **Trends in cognitive sciences**, v. 4, n. 11, p. 417-423, 2000.

BADDELEY, Alan. **Working memory, thought, and action**. OUP Oxford, 2007.

BADDELEY, Alan. **Essentials of human memory (classic edition)**. Psychology Press, 2013.

BADDELEY, Alan. **Exploring Working Memory: Selected Works of Alan Baddeley**. Routledge, 2017.

BADDELEY, A.; EYSENCK, M. W.; ANDERSON, M. C. **Memory (2nd edn)**. 2014.

- BADDELEY, Alan. **Working memory, thought, and action**. OUP Oxford, 2007.
- BADDELEY, Alan D.; HITCH, Graham. Working memory. In: **Psychology of learning and motivation**. Academic press, 1974.
- BADDELEY, Alan D. EYSENCK, Michael W. ANDERSON, Mike. **Memory** (2nd edn. Hoboken: Taylor and Francis, 2014.
- BADDELEY, Alan D.; KOPELMAN, Michael D.; WILSON, Barbara A. (Ed.). **The handbook of memory disorders**. John Wiley & Sons, 2003.
- BADDELEY, Alan D.; KOPELMAN, Michael; WILSON, Barbara A. (Ed.). **The essential handbook of memory disorders for clinicians**. John Wiley & Sons, 2004.
- BADECKER, William; STRAUB, Kathleen. The processing role of structural constraints on interpretation of pronouns and anaphors. **Journal of Experimental Psychology: Learning, Memory, and Cognition**, v. 28, n. 4, p. 748-769, 2002.
- BARBOSA, Pilar; DUARTE, Maria Eugênia Lamoglia; KATO, Mary Aizawa Null subjects in European and Brazilian. NEGRÃO, E. V. (Eds.). **Brazilian Portuguese and the Null Subject Parameter**. Frankfurt: Vervuert-Iberoamericana, 2000
- BASSETTO, Livia Maria Turra. Nomear e (re) categorizar: o papel do nome próprio na referenciação. **Revista de Letras**, n. 34, v. 02, p. 07-19, 2015.
- BEAR, Mark F.; CONNORS, Barry W.; PARADISO, Michael A. Structure of the Nervous System. IN: BEAR, Mark F.; CONNORS, Barry W.; PARADISO, Michael A. (Eds) **Neuroscience: Exploring the Brain**. Williams & Wilkins, p. 152-185, 1996.
- BENVENISTE, Émile. **Problemas de Linguística Geral I**: tradução de Maria da Glória Novak e Maria Luisa Neri: revisão do prof. Isaac Nicolau Salum, 5 ed., 2005.
- BHAT, D. N. S. **Pronouns**. Oxford: Oxford University Press, 2004.
- CABANA, Nasle Maria. **Da realização do sujeito no Português do Brasil**: um estudo em tempo real do uso do sujeito nulo na fala de Belo Horizonte/MG. Dissertação (Mestrado em Linguística) – Universidade Federal de Minas Gerais, Belo Horizonte, 2004.
- CAVALCANTE, Mônica Magalhães; BRITO, Mariza Angélica Paiva. O caráter naturalmente recategorizador das anáforas. **Estudos do discurso: caminhos e tendências**. São Paulo: Paulistana, p. 119-133, 2016.
- CAVALCANTE, Silvia Regina de O.; DUARTE, Maria Eugênia Lamoglia. The subject position in Brazilian Portuguese: the embedding of a syntactic change. **University of Pennsylvania Working Papers in Linguistics**, v. 14, n. 2, p. 8, 2008.

CHAMBERS, Craig G.; SMITH, Ron. Structural parallelism and discourse coherence: A test of centering theory. **Journal of Memory and Language**, v. 39, n. 4, p. 593-608, 1998.

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

CIASCA, Sylvia Maria. Distúrbios e dificuldades de aprendizagem: questão de nomenclatura. **Distúrbios de aprendizagem: proposta de avaliação interdisciplinar**. São Paulo: Casa do Psicólogo, p. 19-31, 2003.

CLIFTON, Charles; KENNISON, Shelia M.; ALBRECHT, Jason E. Reading the wordsher, his, him: implications for parsing principles based on frequency and on structure. **Journal of Memory and language**, v. 36, n. 2, p. 276-292, 1997.

CONCEIÇÃO, Sanir da. **PRO e o Português Brasileiro**. Orientador: Carlos Miotto. 1999. 96f. Dissertação (Mestrado em Linguística) – Pós-Graduação de Letras/Linguística, Universidade Federal de Santa Catarina, Florianópolis, 1999.

CÔRREA, Vilma Reche. **O objeto direto nulo no português do Brasil**. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Campinas, 1991.

CORREIA, Débora Vasconcelos; NETO, José Ferrari; LEITÃO, Márcio Martins. Processamento correferencial de nomes e pronomes do português brasileiro em pessoas portadoras de gagueira. **Letras de Hoje**, v. 48, n. 1, p. 59-67, 2013.

COSTA, Sueli. **O (não) preenchimento do objeto anafórico na língua portuguesa: análise diacrônica do PB e do PE dos séculos XIX e XX**. Tese de Doutorado, Universidade Federal de Santa Catarina, 2011.

CYRINO, Sonia Maria Lazzarini. **O objeto nulo no português brasileiro—um estudo sintático-diacrônico**. Londrina: UEL, 1997.

DANEMAN, Meredyth; CARPENTER, Patricia A. Individual differences in working memory and reading. **Journal of verbal learning and verbal behavior**, v. 19, n. 4, p. 450-466, 1980.

DE JONG, Peter F. Understanding normal and impaired reading development: A working memory perspective. In: **Working memory and education**. Academic Press, 2006. p. 33-60.

DIVIDINO, Renata Queiroz; FAIGLE, Ariadne. Distinções entre memória de curto prazo e memória de longo prazo. **Acessado em**, v. 1, p. 255-276, 2004.

DUARTE, Maria Eugenia Lamoglia. **Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no português do Brasil**. Dissertação de Mestrado. 1986.

DUARTE, Maria Eugênia Lammoglia. Tópico e sujeito: duas categorias em sintaxe? **Cadernos de Estudos Linguísticos**, Campinas, v.17, p.109-132, 1989.

DUARTE, Maria Eugênia Lammoglia. **A perda do princípio “Evite Pronome” no português brasileiro**. Tese (Doutorado em Linguística), Universidade de Campinas, Campinas, 1995.

DUARTE, Maria Eugênia Lammoglia. 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**. 2. ed. Campinas: Editora da UNICAMP, 1996.

DUARTE, Maria Eugênia Lammoglia. The loss of the Avoid Pronoun Principle in Brazilian Portuguese. In: KATO, M. A.; NEGRÃO, E. V. (Eds.). **Brazilian Portuguese and the Null Subject Parameter**. Frankfurt: Vervuert-Iberoamericana, 2000.

DUARTE, Maria Eugênia Lamoglia. The loss of the ‘Avoid Pronoun’ principle in Brazilian Portuguese. **Brazilian Portuguese and the null subject parameter**, p. 17-36, 2000.

DUARTE, Maria Eugênia Lammoglia. A evolução na representação do sujeito pronominal em dois tempos. In: PAIVA, M. C.; DUARTE, M. E. L. (Orgs.). **Mudança linguística em tempo real**. Rio de Janeiro: Contracapa, 2003.

DUARTE, Maria Eugênia Lammoglia. Sujeitos de referência definida e arbitrária: aspectos conservadores e inovadores na escrita padrão. **Linguística / Revista do Programa de Pós-graduação em Linguística**, v. 3, n. 1, 2007.

DUARTE, Maria Eugênia Lamoglia. Sujeitos de referência definida e arbitrária: aspectos conservadores e inovadores na escrita padrão. **Revista Linguística**, v. 3, n. 1, 2007.

DUARTE, Maria Eugênia Lammoglia; KATO, MARY AIZAWA A Variação entre construções finitas pessoais e impessoais no Português Brasileiro. **Web-Revista SOCIODIALETO**, v. 4, p. 153-177, 2014.

DUARTE, Maria Eugênia Lammoglia; KATO, MARY AIZAWA. Do pronome nulo ao pronome pleno: a trajetória do sujeito nulo no português do Brasil. **Português Brasileiro: uma viagem diacrônica**, 2017.

FILLMORE, Charles J. An alternative to checklist theories of meaning. In: **Annual Meeting of the Berkeley Linguistics Society**. p. 123-131, 1975

FRAZIER, Lyn *et al.* **Construal**. Mit Press, 1996.

GADELHA, Luísa de Araújo Pereira. **Processamento da correferência anafórica de pronomes e nomes repetidos em brasileiros aprendizes de francês como L2**. Orientador: Márcio Martins Leitão. 2009. 84f. Dissertação (Mestrado em Linguística) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba, João Pessoa, 2009.

GARROD, Simon; SANFORD, Anthony. Interpreting Anaphoric Relations: The Integration of Semantic. **Journal of Verbal Learning and Verbal Behavior**, v. 16, p. 77-90, 1977.

GELORMINI-LEZAMA, Carlos. **The overt pronoun penalty**: a processing delay in Spanish anaphora comprehension. Orientador: Amit Almor. 81f. Dissertation (Doctor of Philosophy) – College of Arts and Sciences, University of South Carolina, 2010.

GERNSBACHER, Morton Ann; HARGREAVES, David J. Accessing sentence participants: The advantage of first mention. **Journal of memory and language**, v. 27, n. 6, p. 699-717, 1988.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. Editora Atlas SA, 2008.

GODOY, Mahayana Cristina. **Resolvendo a anáfora conceitual: um olhar para além da relação antecedente/anafórico**. Orientador: Edson Françoso. 2010. 96f. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 2010.

GODOY, Mahayana Cristina. **Idiossincrasias do processamento de pronomes plurais**. Orientador: Edson Françoso. 2014. 136f. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 2014.

GOLDMAN-RAKIC, Patricia S. Cellular basis of working memory. **Neuron**, v. 14, n. 3, p. 477-485, 1995.

GONDIM, Eva Vilma Aires Cabral. **Investigação teórico-metodológica sobre a penalidade do nome repetido em português brasileiro**. Orientador: Márcio Martins Leitão. 2017. 96f. Dissertação (Mestrado em Linguística) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba, 2017.

GORDON, Peter C.; CHAN, Davina. Pronouns, passives, and discourse coherence. **Journal of Memory and Language**, v. 34, n. 2, p. 216-231, 1995.

GORDON, Peter C.; GROSZ, Barbara J.; GILLIOM, Laura A. Pronouns, names, and the centering of attention in discourse. **Cognitive science**, v. 17, n. 3, p. 311-347, 1993.

GORDON, Peter C.; HENDRICK, Randall. Intuitive knowledge of linguistic co-reference. **Cognition**, v. 62, n. 3, p. 325-370, 1997.

GORDON, Peter C.; HENDRICK, Randall. The representation and processing of coreference in discourse. **Cognitive science**, v. 22, n. 4, p. 389-424, 1998.

GRODZINSKY, Yosef; REINHART, Tanya. The innateness of binding and coreference. **Linguistic inquiry**, v. 24, n. 1, p. 69-101, 1993.

GROOME, David; EYSENCK, Michael W. **An introduction to applied cognitive psychology**. Psychology Press, 2016.

GROSZ, Barbara J.; SIDNER, Candace L. Attention, intentions, and the structure of discourse. **Computational linguistics**, v. 12, n. 3, p. 175-204, 1986.

GROSZ, Barbara J.; WEINSTEIN, Scott; JOSHI, Aravind K. Centering: A framework for modeling the local coherence of discourse. **Computational linguistics**, v. 21, n. 2, p. 203-225, 1995.

GUY, Gregory R.; ZILLES, Ana. **Sociolinguística quantitativa: instrumental de análise**. Ed. Parábola, 2007.

HAEGEMAN, Liliane M. V. **Introduction to government and binding theory**. Oxford: Blackwell, 1991.

HOLMBERG, Anders; NAYUDU, Aarti; SHEEHAN, Michelle. Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi. **Studia Linguistica**, v. 63, n. 1, p. 59-97, 2009.

HORA, Katharine de Freitas Pereira Neto Aragão da. **O processamento da correferência pronominal anafórica em estruturas complexas do Português Brasileiro**. Orientador: Marcus Antônio Rezende Maia. 2014. 215f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro. Dissertação de Mestrado, 2014.

HUANG, Yan. **Anaphora: A cross-linguistic approach**. Oxford University Press on Demand, 2000.

HUETTIG, Falk; ROMMERS, Joost; MEYER, Antje S. Using the visual world paradigm to study language processing: A review and critical evaluation. **Acta psychologica**, v. 137, n. 2, p. 151-171, 2011.

IZQUIERDO, Iván. Memória. 2ª edição, revista e ampliada. **Porto Alegre: Artmed**, 2011.

JAEGGLI, Osvaldo A. Passive. **Linguistic inquiry**, p. 587-622, 1986.

JUBRAN, Clelia Candida Abreu Spinardi et al. Organização tópica da conversação. **Gramática do português falado**, 4 ed.rev. Campinas – SP: Editora da Unicamp, p. 341-375, 2002.

KATO, Mary Aizawa.; TARALLO, Fernando “Anything YOU can do in Brazilian Portuguese”. In. O. Jaeggli e C. Silva-Corvalán (orgs). **Studies in Romance Linguistics**. Dordrecht: Foris: 1986.

KATO, Mary Aizawa. The partial pro-drop nature and the restricted VS order in Brazilian Portuguese. In: KATO, Mary Aizawa; NEGRÃO, Esmeralda Vailati.

**Brazilian Portuguese and the Null Subject Parameter.** Frankfurt: Vervuert-Iberoamericana, 2000.

KATO, Mary Aizawa; DUARTE, Maria Eugênia Lammoglia. Semantic and phonological constraints on the distribution of null subjects in Brazilian Portuguese. **Comunicação apresentada no NWAV32**, realizado na Universidade da Pensilvânia, em outubro de 2003.

KATO, Mary Aizawa; DUARTE, Maria Eugênia Lammoglia. Restrições na distribuição de sujeitos nulos no português brasileiro. **Revista Veredas**, v. 18, n. 1, p. 1-22, 2014.

KEMMERER, David. **Cognitive neuroscience of language.** Psychology Press, 2014.

KENNISON, Shelia M. Comprehending the pronouns her, him, and his: Implications for theories of referential processing. **Journal of Memory and Language**, v. 49, n. 3, p. 335-352, 2003.

KENNISON, Shelia M.; GORDON, Peter C. Comprehending referential expressions during reading: Evidence from eye tracking. **Discourse Processes**, v. 24, n. 2-3, p. 229-252, 1997.

KOCH. Ingedore Grufeld Villaça. **Introdução à Linguística Textual: trajetória e grandes temas.** 2ª ed. São Paulo: Contexto, 2015

LÁCIO-WEB. **Lácio-Web: Compilação de Córpus do Português do Brasil e Implementação de Ferramentas para Análises Linguísticas.** c.2014. Descrição. Disponível em: <<http://143.107.183.175:22180/lacioweb/descricao.htm>>. Acesso em: 20 de jun. de 2019.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de metodologia científica.** 4. ed. São Paulo: Atlas, 2001.

LAPOINTE, Leonard L. **Paul Broca and the Origins of Language in the Brain.** Plural Publishing, 2012.

LEITÃO, Márcio Martins. **O processamento do objeto direto anafórico no português brasileiro.** Rio de Janeiro: UFRJ/ Faculdade de Letras. Tese de Doutorado em Linguística, 2005a.

LEITÃO, Márcio Martins; MAIA, Marcus. Processamento na interface sintaxe-semântica: objeto direto anafórico e traço de animacidade. In: FINGER, I. & MAIA, M. (Eds.). **Processamento da linguagem.** Pelotas: Educat, p. 337-362, 2005b.

LEITÃO, Márcio Martins; PEIXOTO, Priscilla Cristina de Carvalho; SANTOS, Suzana Thais Pedroza. Processamento da co-referência intra-sentencial em português brasileiro. **Veredas on-line**, v. 2, p. 50-61, 2008.

LEITÃO, Márcio Martins; SIMÕES, Antônio Barros Gibson. A influência da distância no processamento correferencial de pronomes e nomes repetidos em Português Brasileiro. **Veredas on-line**, Juiz de Fora, p. 262-272, 2008.

LEITÃO, Márcio Martins; RIBEIRO, Antônio João Carvalho; MAIA, Marcus. Penalidade do nome repetido e rastreamento ocular em português brasileiro. **Revista Linguística / Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro**, v. 8, n. 2, 2012.

LEITÃO, Márcio Martins; OLIVEIRA, Rosana Costa de; TEIXEIRA, Elisângela Nogueira; FERRARI-NETO, J.; BRITO, D. B. S. de. Evidências experimentais do processamento da correferência e dos princípios de ligação em Português Brasileiro. **Revista da ABRALIN**, v. 13, n. 1, p. 289-332, jul./dez. 2014.

LEITÃO, Márcio Martins; BEZERRA, Gitanna Brito; BRITO, Dorothy Bezerra Silva de. Restrições da teoria da ligação e o processamento da correferência pronominal em dois estágios. **Cadernos de Letras da UFF**, v. 49, p. 89-108, 2014.

LEVINSON, Stephen C. **Pragmática**. São Paulo: Martins, 2007.

LIMA, Alisson Hudson Veras. **A influência da retomada e da distância sintática no processamento de pronomes plenos e nulos em português brasileiro**. Orientadora: Maria Elias Soares. 2015. 125f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará, 2015.

LIMA, Juciane Nóbrega; SOUZA, Flávia Gonçalves Calado de. Alguns fatores relevantes no processamento da correferência intersentencial. **Revista Escrita**, v. 16, 2016.

LIMA, Juciane Nóbrega. **Paralelismo e foco estrutural no processamento do correferência de pronomes e nomes repetidos**. Orientador: Márcio Martins Leitão. 2014. 77f. Dissertação (Mestrado em Linguística) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba, 2014.

LOPES, Célia. **A inserção de ‘a gente’ no quadro pronominal do português**. Vol. 18. Frankfurt/Madrid: Vervuert/Iberoamericana, 2003.

LUEGI, Paula. **Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes**. Orientadores: Maria Armada Costa e Marcus Rezende Maia. 2012. 267f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade de Lisboa, 2012.

LUEGI, Paula; COSTA, Amanda; MAIA, Marcus. Processamento e interpretação de sujeitos nulos e plenos em Português Europeu e em Português Brasileiro. **Cadernos de Letras da UFF**, v. 49, p. 67-88, 2014.

LYONS, John. **Semântica**. Lisboa: Editoria Presença/Martins Fontes, 1977

MAIA, Jefferson de Carvalho; LIMA, Maria Luíza Cunha. O processamento de expressões correferenciais em português brasileiro: nomes repetidos,

pronomes plenos e pronomes nulos. **Revista do GELNE (UFC)**, v. 13, n. 1/2, 2011.

MAIA, Jefferson de Carvalho; LIMA, Maria Luíza Cunha. Processamento correferencial de nomes e pronomes plenos em PB: evidências de rastreamento ocular. **ReVEL**, edição especial n. 6, 2012.

MAIA, Jefferson de Carvalho. **O processamento de expressões correferenciais em português**. Orientadora: Maria Luíza Cunha Lima. 2013. 120f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, 2013.

MAIA, Jefferson de Carvalho; LIMA, Maria Luíza Cunha. Referenciação e técnicas experimentais: aspectos metodológicos na investigação do processamento correferencial em português brasileiro. **Revista de Estudos Linguísticos**, Belo Horizonte, v.22, n.1, p. 67-93, jan./jul. 2014.

MARAFONI, Renata Lopes. A realização do objeto direto anafórico: um estudo em tempo real de curta duração. **Rio de Janeiro: UFRJ (Dissertação de Mestrado)**, 2004.

MARAFONI, Renata Lopes. **A distribuição do objeto nulo no português Europeu e no português brasileiro**. Tese de Doutorado, UFRJ/Faculdade de Letras, Rio de Janeiro, 2010.

MARAFONI, Renata Lopes. A realização do objeto direto anafórico: um estudo MIOTO, Carlos; SILVA, Maria Cristina Figueiredo; LOPES, Ruth Elisabeth Vasconcellos. **Novo manual de sintaxe**. Ed. Contexto, 2013.

MORGADO, Sara Matias. **Processamento da co-referência pronominal: informação sintáctica e semântica**. Orientadoras: Maria Armanda Costa e Maria Gabriela Matos. 2011. 199f. Dissertação (Mestrado em Psicolinguística) - Faculdade de Letras, Universidade de Lisboa, 2011.

NICOLAU, Eunice Maria das Dores. **As propriedades de sujeito nulo e ordem V-S no Português Brasileiro**. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Campinas, 1995.

RAPOSO, Eduardo; JAEGGLI, Osvaldo; SILVA-CORVALAN, Carmen. On the null object in European Portuguese. **Studies in Romance linguistics**, p. 373-390, 1986.

OLIVEIRA, Marian dos Santos. **Concordância verbal de terceira pessoa do plural em Vitória da Conquista: variação estável ou mudança em progresso?** 190f. Dissertação de Mestrado. Pós-Graduação em Letras e Linguística. Instituto de Letras. Universidade Federal da Bahia, 2005.

OLIVEIRA, Solange Mendes. Objeto direto nulo, pronome tônico de 3ª pessoa, SN anafórico e clítico acusativo no português brasileiro: uma análise de textos escolares. In.: **Revista virtual de Estudos de Linguagem – ReVEL**. Vol. 5 – número 9 – agosto de 2007.

OMENA, Nelize Pires de. **Pronome pessoal de terceira pessoa: suas formas variantes em função acusativa**. Dissertação de mestrado, PUC-RJ, 1978

OTHERO, Gabriel de Avila. **Teoria X-barras: descrição do português e aplicação computacional**. Contexto, 2006.

OTHERO, Gabriel de Ávila; SPINELLI, Ana Carolina. Analisando a retomada anafórica do objeto direto em português falado. **Revista Letras**. n.96, jul./dez. 2017.

PAREDES SILVA, Vera Lúcia. **Cartas cariocas: a variação do sujeito na escrita informal**. Tese (Doutorado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

PEREIRA, Ivelã. **O uso variado das formas anafóricas no acusativo**. Trabalho de Conclusão de Curso, Universidade Federal de Santa Catarina, 2011.

PINHEIRO, João Pinheiro *et al.* **Probabilidade e Estatística**. Elsevier Brasil, 2013.

PINTO, Cecília Augusto Vieira. **Variação do objeto anafórico acusativo na fala de Florianópolis**. Dissertação de Mestrado. 164f. Pós-Graduação em Linguística. Universidade Federal de Santa Catarina. Florianópolis, 2015.

POESIO, Massimo; STUCKARDT, Roland; VERSLEY, Yannick. **Anaphora Resolution**. Springer, 2016.

PROJETO, NURC. **Projeto norma linguística urbana culta do Rio de Janeiro**. c.1994. Histórico do projeto. Disponível em: <<http://www.nurcrj.lettras.ufrj.br>>. Acesso em: 21 de jun. de 2019.

QUEIROZ, Kamila Lima de; LEITÃO, Márcio Martins. Processamento do sujeito anafórico em português brasileiro. **Veredas** – Revista de Estudos Linguísticos Online, v. 2, p. 163-166, 2008.

QUEIROZ, Kamila Lima de. **Processamento da co-referência: pronomes lexicais, nomes repetidos, hiperônimos e hipônimos como formas de retomada anafórica inter-sentencial do sujeito em Português Brasileiro**. Orientador: Márcio Martins Leitão. 2009. 61f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística da Universidade Federal da Paraíba, 2009.

RUMEU, Maria Cristina de Brito. **A variação tu e você no português brasileiro oitocentista e novecentista: reflexões sobre a categoria de gênero social**. Alfa. São Paulo, n.57, 2013.

SILVA, Vera Lúcia Paredes Pereira da. **Cartas cariocas: a variação do sujeito na escrita informal**. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

SIMÕES, Antonia Barros Gibson. **A influência da coesão e da coerência no processamento correferencial de pronomes e nomes repetidos em**

**português brasileiro.** Orientador: Márcio Martins Leitão. 2014. 105f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística da Universidade Federal da Paraíba, 2014.

SOARES, Maria Elias. **A constituição do discurso coeso:** um estudo evolutivo da produção oral e escrita. Orientadora: Letícia Maria Sicuro Corrêa. 1991. 297f. Tese (Doutorado em Letras) – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, 1991.

SOUSA, Antonio Anderson Marques de. **As realizações do acusativo anafórico no Português Europeu e Brasileiro:** um estudo diacrônico. Dissertação de Mestrado. 127f. Pós-Graduação em Letras Vernáculas. Faculdade de Letras. Universidade Federal do Rio de Janeiro, 2017.

SOUZA, Adriana Regina Marques de; SISTO, Fermino Fernandes. Dificuldade de aprendizagem em escrita, memória e contradições. **Psicologia Escolar e Educacional**, v.5, 39-47, 2001.

TARALLO, Fernando. **Relativization strategies in Brazilian Portuguese.** Tese (Doutorado em Linguística) – University of Pennsylvania, Filadélfia, 1987.

TEIXEIRA, Elisângela Nogueira. **Preferências sintáticas e semânticas no processamento da correferência anafórica:** evidências da movimentação ocular. Orientadora: Maria Elias Soares. 2013. 185F. Tese (Doutorado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, 2013.

TEIXEIRA, Elisângela Nogueira; FONSECA, Maria Cristina Micelli; SOARES, Maria Elias. Resolução do pronome nulo em Português Brasileiro: Evidência da movimentação ocular. **Veredas on-line**, Juiz de Fora, v. 18/1, p. 281-301, 2014.

TEIXEIRA, Elisângela Nogueira; SOARES, Maria Elias. Efeito de hiperonímia no processamento da correferência interfrasal em Português Brasileiro. **Cadernos de Letras da UFF**, n. 49, p. 47-66, 2014.

TRAXLER, Matthew; GERNSBACHER, Morton Ann (Ed.). **Handbook of psycholinguistics.** Elsevier, 2011.

VAN GOMPEL, Roger P.G.; LIVERSEDGE, Simon P.; PEARSON, Jamie. Antecedent typicality effects in the processing of noun phrase anaphors. **The On-line Study of Sentence Comprehension: Eyetracking, ERP and Beyond.** Psychology Press, New York, p. 119-137, 2004.

VASCONCELOS, Manuela Leitão; LEITÃO, Márcio Martins. Processamento correferencial de pronomes e nomes repetidos em pacientes com afasia de Broca. **Revista Virtual de Estudos da Linguagem**, v. 10, p. 148-168, 2012.

VASCONCELOS, Maria Leitão. **Processamento da correferência em pacientes com afasia de expressão.** Orientador: Márcio Martins Leitão. 2012. 132f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística da Universidade Federal da Paraíba. Dissertação de Mestrado, 2012.

VERÍSSIMO, Victor. A evolução do conceito de parâmetro do sujeito nulo. **Entrepalavras**, v. 7, n. 1, p. 76-90, 2017.

WARD, Jamie. **The student's guide to cognitive neuroscience**. Psychology Press, 2015.

WILSON, Barbara A. **Memory rehabilitation: Integrating theory and practice**. Guilford Press, 2009.