



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

ELISÂNGELA NOGUEIRA TEIXEIRA

PREFERÊNCIAS SINTÁTICAS E SEMÂNTICAS
NO PROCESSAMENTO DA CORREFERÊNCIA ANAFÓRICA:
EVIDÊNCIAS DE MOVIMENTAÇÃO OCULAR

FORTALEZA

2013

ELISÂNGELA NOGUEIRA TEIXEIRA

PREFERÊNCIAS SINTÁTICAS E SEMÂNTICAS
NO PROCESSAMENTO DA CORREFERÊNCIA ANAFÓRICA:
EVIDÊNCIAS DE MOVIMENTAÇÃO OCULAR

Tese apresentada como requisito parcial para
obtenção do título de Doutor em Linguística
pelo Programa de Pós-Graduação em
Linguística do Departamento de Letras
Vernáculas da Universidade Federal do Ceará.

Área de Concentração: Psicolinguística

Orientadora: Prof. Dra. Maria Elias Soares

FORTALEZA

2013

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca de Ciências Humanas

T265p Teixeira, Elisângela Nogueira.
 Preferências sintáticas e semânticas no processamento da correferência anafórica : evidências de
 movimentação ocular / Elisângela Nogueira Teixeira. – 2013.
 184 f. : il. color., enc. ; 30 cm.

 Tese(doutorado) – Universidade Federal do Ceará, Centro de Humanidades, Departamento de
Letras Vernáculas, Programa de Pós-Graduação em Linguística, Fortaleza, 2013.
 Área de Concentração: Psicolinguística.
 Orientação: Profª. Dra. Maria Elias Soares.

 1.Língua portuguesa – Anáfora. 2.Língua portuguesa – Referência. 3.Olhos – Movimentos.
4.Leitura. I.Título.

CDD 469.5

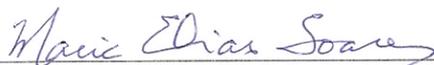
ELISÂNGELA NOGUEIRA TEIXEIRA

PREFERÊNCIAS SINTÁTICAS E SEMÂNTICAS
NO PROCESSAMENTO DA CORREFERÊNCIA ANAFÓRICA:
EVIDÊNCIAS DE MOVIMENTAÇÃO OCULAR

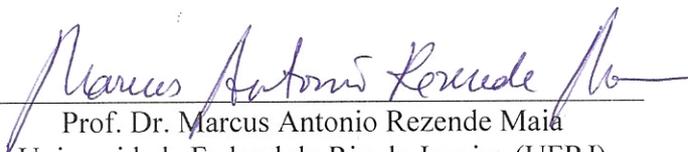
Tese apresentada como requisito parcial para
obtenção do grau de Doutor em Linguística
pelo Programa de Pós-Graduação em
Linguística do Departamento de Letras
Vernáculas da Universidade Federal do Ceará.
Área de Concentração: Psicolinguística

Aprovada em: 17/05/2013

BANCA EXAMINADORA



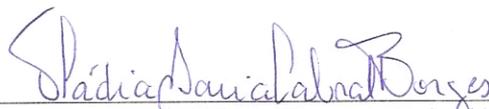
Prof. Dra. Maria Elias Soares (Orientadora)
Universidade Federal do Ceará (UFC)



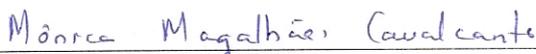
Prof. Dr. Marcus Antonio Rezende Maia
Universidade Federal do Rio de Janeiro (UFRJ)



Prof. Dr. Márcio Martins Leitão
Universidade Federal da Paraíba (UFPB)



Prof. Dra. Vlândia Maria de Cabral Borges
Universidade Federal do Ceará (UFC)



Prof. Dra. Mônica Magalhães Cavalcante
Universidade Federal do Ceará (UFC)

Para os meus dois amores, Raul e Soares.

AGRADECIMENTOS

À Prof. Maria Elias Soares, minha orientadora e amiga, que tem me permitido construir um caminho de pesquisa experimental sobre o processamento da linguagem, aceitando e viabilizando, com empolgação e ousadia, muitos dos meus sonhos de linguista iniciante, meus mais sinceros agradecimentos.

À CAPES, agência financeira que me concedeu bolsa de doutorado nos últimos quatro anos e financiou parte dos projetos de pesquisa e dos eventos científicos em torno do tema desta tese.

Ao CNPq, agência financiadora de dois projetos de pesquisa que ajudaram a estruturar e manter os laboratórios onde os experimentos desta tese foram rodados.

Aos meus recentes amigos e colaboradores, que me permitiram aprender com eles, à distância ou em sua presença, Márcio Leitão, Eduardo Kenedy, Marcus Maia, Maria Luiza Cunha Lima, Maria Cristina Fonseca, Julien Schrenk, Heitor Credidio, Thiago Santiago, Maria Armanda Costa e Paula Luegi.

Aos professores do Departamento de Letras Vernáculas e do Programa de Pós-Graduação em Linguística da UFC, que sempre foram generosos em suas contribuições para o meu trabalho, em especial a Prof. Mônica Magalhães Cavalcante e a Prof. Vlândia Maria de Cabral Borges.

A Eduardo Xavier, Antonia dos Santos e Vanessa Marques, secretários do Programa de Pós-Graduação em Linguística da UFC, que estiveram sempre presentes para dar apoio no que fosse preciso.

Aos participantes da pesquisa desta tese, sem os quais a pesquisa experimental do comportamento humano não poderia acontecer.

Finalmente, agradeço à minha família, sempre presente, especialmente ao meu marido, José Soares de Andrade Júnior, que acreditou em mim e me apoiou, sem restrições, não apenas no plano familiar, mas como cientista experiente que é, incentivando-me incondicionalmente nestes duros anos, tanto no plano acadêmico quanto no pessoal.

Agradeço a imensa paciência e o carinho sem fim de Raúl B. Gómez del Estal Teixeira, meu filho, que soube ceder sua mãe ao onipresente computador e que, mesmo assim, torceu e vibrou comigo em todas as minhas conquistas, esperando ansiosamente pelo momento, como ele diz, de dar o ponto final na tese.

E desde e para sempre agradeço aos meus pais, à minha mãe querida e visionária, Francisca Nogueira Teixeira, e a meu pai, generoso e desprendido, Joaquim Martins Teixeira, pela criação e companhia que me ofereceram e ainda oferecem, fazendo-me sentir tão tranquila e apoiada em qualquer momento da minha vida.

Aos meus queridos irmãos artistas, Francimara Nogueira Teixeira e Frederico Nogueira Teixeira, sempre dispostos a me auxiliar e apoiar, com quem sei que posso contar a qualquer hora e em qualquer lugar, meu muito obrigada.

No estamos recorriendo el trecho que ya hemos recorrido.

No estamos recorriendo el trecho que aún falta recorrer.

Un trecho no recorrido ni por recorrer es incomprensible.

Nagarjuna séc. II [Radhakrishnan]

RESUMO

Esta tese tem como objetivo principal contribuir com o desenvolvimento dos estudos psicolinguísticos que procuram demonstrar experimentalmente conjecturas teóricas a respeito do processamento anafórico. Tomando por base a Teoria da Acessibilidade (ARIEL, 1991, 2001), a Teoria da Centralização (GROSZ; JOSHI; WEINSTEIN, 1995), os trabalhos em torno da tipicidade do termo antecedente (GARROD; SANFORD, 1977; VAN GOMPEL; LIVERSEDGE; PEARSON, 2004), a Hipótese da Carga Informacional (ALMOR, 1999) e a Hipótese da Posição do Antecedente (CARMINATI, 2002), trabalhamos com a hipótese de que, em períodos complexos por coordenação e subordinação, formados por no máximo duas orações, a saliência da posição sintática de sujeito é o principal fator para a resolução anafórica em língua portuguesa. Fazendo uso de metodologia experimental *on-line* e *off-line*, procuramos evidências para nossa hipótese em um conjunto formado por quatro estudos, composto por (i) um experimento de compreensão de períodos complexos por coordenação, em que foram manipulados a posição do antecedente e o tipo de relação semântica entre antecedente e anáfora; (ii) um experimento de compreensão de períodos complexos por subordinação, em que foram manipulados o tipo da correferência anafórica, sob a forma de pronome pleno ou nulo, e a posição da correferência, anafórica ou catafórica; (iii) uma sondagem de produção de períodos complexos com uso de pronomes plenos ou nulos como correferentes; e (iv) uma análise dos movimentos oculares durante a leitura de textos autênticos em língua portuguesa com o objetivo de encontrar padrões de fixação oculares. Os estudos foram realizados em um rastreador ocular de 120 Hz que registrou a cada 8 ms a movimentação ocular dos participantes durante a leitura dos estímulos. As variáveis dependentes de movimentação ocular analisadas foram: (i) o número de fixações; (ii) o tempo da primeira fixação; (iii) a duração média da fixação ocular; e (iv) o tempo total de fixação. A análise conjunta dos resultados dos experimentos sugere que a resolução da anáfora correferencial nos períodos complexos estudados é uma função da proeminência sintática da posição de sujeito e que a carga de informação das expressões anafóricas com conteúdo semântico parece levar a um aumento de custo durante o processamento anafórico de um antecedente altamente acessível.

Palavras-chave: Processamento anafórico. Movimentação ocular. Proeminência sintática. Preferência semântica.

ABSTRACT

In this dissertation, our main objective is to contribute for the development and understanding of psycholinguistics studies that attempt to experimentally demonstrate relevant theoretical conjectures about anaphoric processing. Under the conceptual frameworks of the Theory of Accessibility (ARIEL, 1991, 2001), the Theory of Centering (GROSZ; JOSHI; WEINSTEIN, 1995), the studies on the typicality of the antecedent term (GARROD; SANFORD, 1977; VAN GOMPEL; LIVERSEDGE; PEARSON, 2004), the Informational Load Hypothesis (ALMOR, 1999), and the Position of Antecedent Hypothesis (CARMINATI, 2002), we propose that the prominence of the syntactic position in complex sentences plays a major role on the anaphoric resolution in the Portuguese language. Adopting a psycholinguistic methodology based on on-line (tracking of eye movements) as well as off-line observations, we searched for evidence to support our hypothesis from the results of the following set of studies: (i) an experiment to evaluate the comprehension of complex sentences due to coordination, in which both the position of the antecedent and the type of semantic relationship between antecedent and anaphora are manipulated; (ii) an experiment to evaluate the comprehension of complex sentences due to subordination, in which both the type of anaphoric coreference, in the form of a plain or null pronoun, and the position of the coreference, anaphoric or cataphoric, are manipulated; (iii) an experiment for generation of complex sentences, using plain or null pronouns as coreferentials; and (iv) a reading experiment of non-manipulated texts to establish a comparative standard for reading flux in Brazilian Portuguese. Our on-line experiments were performed with an eye-tracker of 120 Hz, which allowed eye movements to be recorded at each 8 milliseconds. The following dependent variables related with the eye movement have been analyzed: (i) the number of fixations; (ii) the duration time of the first fixation; (iii) the average duration of the fixations; and (iv) the total time of fixation. The overall analysis of our results, based on the investigation of complex sentences, suggests that the resolution of the coreferential anaphora is a function of the prominence of the subject position. Moreover, the information load of anaphoric expressions with semantic content seems to increase the cost of the anaphoric processing of a highly accessible antecedent.

Keywords: Anaphoric processing. Eye movement. Syntactic prominence. Semantic preferences.

LISTA DE FIGURAS

Figura 1 – Exemplo de tradução do Português para Inglês (Feita em março de 2013).....	20
Figura 2 - Ilustração das noções de c-comando, domínio, ligação.....	38
Figura 3 - Hierarquia do marcadores de acessibilidade (ARIEL, 1991)	43
Figura 4 – Resultados do experimento 1 de Gordon, Grosz e Gilliom (1993).....	52
Figura 5 – Exemplo de slide usado para exibir os textos não manipulados	75
Figura 6 - Exemplo de uma sequência de estímulos	87
Figura 7 – Exemplos de variação na posição de fixações oculares	107

LISTA DE TABELAS

Tabela 1 - Reprodução da tabela de Ariel (1991).....	45
Tabela 2 – Distribuição dos tempos médios de duração da fixação de Rayner (1998).....	71
Tabela 3 – Dados dos participantes do estudo de textos não manipulados.	74
Tabela 4 – Tempo médio das fixações encontradas na leitura dos textos não manipulados.	77
Tabela 5 – Características das palavras nos textos estudados	78
Tabela 6 – Padrão de duração de fixação e tamanho médio da palavra em português.	79
Tabela 7 - Sumário dos dados demográficos do experimento do Capítulo 5.....	84
Tabela 8 - Média da performance comportamental para as diferentes condições.....	90
Tabela 9 - Tempo médio total de leitura da frase em segundos.	91
Tabela 10 - Sumário dos tempos de duração média e duração da primeira fixação.....	94
Tabela 11 – Dados do experimento de Sorace e Filiaci (2006).....	103
Tabela 12 - Dados do experimento de Fonseca e Guerreiro (2012).....	104
Tabela 13 – Dados demográficos do estudo sobre pronomes nulos e plenos	105
Tabela 14 – Controle de frequência dos nomes usados como competidores.	108
Tabela 15 - Preferências no estabelecimento da correferência em tarefa de produção.....	118

LISTA DE GRÁFICOS

Gráfico 1 - Distribuição do tamanho característico da palavra em língua portuguesa.....	78
Gráfico 2 - Tempo total de leitura por participante.....	91
Gráfico 3 - Tempo total de leitura por item.....	92
Gráfico 4 – Tempo total de fixação nas regiões de interesse	92
Gráfico 5 - Número médio de fixações nas regiões de interesse.....	93
Gráfico 6 - Duração média da fixação nas regiões de interesse.....	94
Gráfico 7 - Duração da primeira fixação (dados corrigidos).....	95
Gráfico 8 - Respostas comportamentais do estudo sobre pronomes plenos e nulos.	111
Gráfico 11 - Preferência pela posição de sujeito em todas as condições experimentais.....	112
Gráfico 9 - Contraste entre pronome pleno e nulo	112
Gráfico 10 - Contraste entre anáfora e catáfora.....	112
Gráfico 12 - Tempo total de leitura (tempo de reação) em cada condição experimental.....	113
Gráfico 13 - Tempo total de fixação nas regiões de interesse.....	115
Gráfico 14 - Duração média das fixações nas regiões de interesse.....	116
Gráfico 15 - Duração média da fixação em cada segmento para a condição de Anáfora	116
Gráfico 16 - Duração média da fixação em cada segmento para a condição Catáfora	116

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Por que processamento?	16
1.2	Objetivo	22
1.3	Organização da tese.....	25
2	REPRESENTAÇÃO E PROCESSAMENTO CORREFERENCIAL	27
2.1	O conceito de correferência anafórica	29
2.2	Teorias e hipóteses sobre o processamento correferencial na gramática.....	34
2.2.1	Teoria da Ligação	35
2.2.1.1	As categorias vazias ou os pronomes nulos	39
2.2.2	Hipótese da Posição do Antecedente.....	40
2.3	Teorias e hipóteses sobre o processamento correferencial no discurso.....	41
2.3.1	Teoria da Acessibilidade	42
2.3.2	Teoria da Centralização.....	46
2.3.2.1	Penalidade do Nome Repetido	51
2.3.2.2	Paralelismo Estrutural	53
2.3.3	Hipótese da Carga Informacional.....	55
3	METODOLOGIA EXPERIMENTAL	58
3.1	A fisiologia da visão	59
3.2	Rastreamento ocular	61
3.3	O estudo experimental da correferência anafórica e a movimentação ocular....	63
3.4	Hipóteses.....	67
3.4.1	Hipótese básica.....	67
3.4.2	Hipóteses secundárias	67
3.5	Procedimentos experimentais para coleta dos dados	68
4	ESTUDO DA LEITURA DE TEXTOS NÃO MANIPULADOS.....	70
4.1	Introdução	70
4.2	Materiais e Método.....	74
4.2.1	Participantes	74
4.2.1	Procedimento experimental e materiais	74
4.3	Medidas examinadas	76
4.3.1	Duração média da fixação ocular.....	76
4.3.2	Característica dos textos apresentados.....	76

4.4	Resultados	76
4.5	Conclusões	79
5	ESTUDO DE PREFERÊNCIAS SEMÂNTICAS	80
5.1	Estudo da hierarquia semântica entre anáfora e antecedente: introdução	80
5.2	Materiais e Métodos	84
5.2.1	Participantes	84
5.2.2	Procedimento experimental e Estímulos	85
5.2.4	Equipamento	87
5.3	Variáveis dependentes	88
5.3.1	Performance comportamental.....	88
5.3.2	Tempo total de leitura	88
5.3.3	Tempo total de fixação	88
5.3.4	Número de fixações.....	89
5.3.5	Duração da Fixação	89
5.3.6	Duração da Primeira Fixação.....	89
5.4	Resultados	89
5.4.1	Performance comportamental.....	89
5.4.2	Movimentação ocular	90
5.5	Discussão	96
6	ESTUDO DE PREFERÊNCIAS SINTÁTICAS.....	99
6.1	Estudo do processamento de pronomes nulos e plenos: introdução	100
6.2	Materiais e Métodos	104
6.2.1	Participantes	104
6.2.2	Estímulos e Procedimento experimental	105
6.2.4	Equipamento	108
6.3	Variáveis dependentes.....	109
6.3.1	Performance comportamental.....	109
6.3.2	Tempo total de leitura ou tempo de reação.....	109
6.3.3	Tempo total de fixação nas regiões de interesse	109
6.3.4	Duração média de fixação nos segmentos do período complexo	110
6.4	Resultados	110
6.4.1	Performance comportamental.....	110
6.4.2	Movimentação ocular	113
6.5	Sondagem para produção de períodos complexos por subordinação.....	117

6.6	Discussão	120
7	CONSIDERAÇÕES FINAIS.....	123
	REFERÊNCIAS.....	130
	APÊNDICE A – ESTÍMULOS DO ESTUDO DE TEXTOS NÃO MANIPULADOS.....	138
	APÊNDICE B – ESTÍMULOS DO ESTUDO SOBRE AS PREFERÊNCIAS SEMÂNTICAS	142
	APÊNDICE C – ESTÍMULOS DO ESTUDO SOBRE AS PREFERÊNCIAS SINTÁTICAS.....	147
	APÊNDICE D – DADOS REFERENTES ÀS MEDIAS DO ESTUDO SOBRE AS PREFERÊNCIAS SEMÂNTICAS	151
	APÊNDICE E – DADOS REFERENTES ÀS MEDIAS DO ESTUDO SOBRE AS PREFERÊNCIAS SINTÁTICAS	175
	ANEXOS.....	182

1 INTRODUÇÃO

1.1 Por que processamento?

Programas sofisticados fazem parte do cotidiano das pessoas em 2013. Estão presentes em dispositivos comuns, como telefones e computadores portáteis, que carregamos nos bolsos ou na mochila. Tais dispositivos são praticamente robôs sem forma humana, pois estão carregados de sistemas inteligentes de automação, que podem monitorar nossa agenda, a segurança de nossos lares, o funcionamento de nossos carros, a iluminação das empresas onde trabalhamos, enfim, podem conter uma quantidade de aplicativos potencialmente infinita. Para desenvolver tudo isso, muitos cientistas e programadores trabalham intensamente, interessados em criar sistemas automáticos capazes de replicar, com maior ou menor grau de precisão, ações humanas individuais e coletivas. Os autômatos não se limitam a reproduzir as ações humanas ordinárias nem ao menos objetivam simplesmente substituí-las. Mais do que isso, à medida em que as máquinas produzidas pelos humanos se complexificam, passam a ampliar seu potencial, numa retroalimentação contínua.

Para realizar todo o conjunto de ações automáticas, os seres humanos criam e recriam linguagens formais de programação. As diversas linguagens de programação, empregadas para construir os mais diversos sistemas, estão baseadas sobretudo em lógicas de vários tipos que procuram não apenas projetar ao máximo a capacidade humana de calcular como também têm procurado simular todas as realidades concebíveis, até mesmo a da capacidade humana de interagir socialmente usando uma língua particular.

Em relação à linguagem oral, convive-se hoje com os programas de sintetização da fala, com os quais já podemos interagir minimamente. Os robôs que levamos em nossos bolsos, uma vez instruídos, podem, entre outras ações, ouvir uma cadeia de voz de uma língua particular, decodificá-la e realizar, a partir da decodificação, ações simples como fazer uma ligação para alguém, procurar informações sobre um assunto, iniciar um sistema de monitoramento a distância, instruir ou realizar qualquer outra tarefa cujo início possa ser dado de modo análogo à ação que nosso dedo executa ao mudar a chave de um interruptor da posição *off* para *on*, ou seja, é nosso comando de voz que faz as vezes de um comando mecânico.

Em relação à linguagem escrita, os computadores são fruto dela. Para dizer uma obviedade, não existiriam sem as línguas que falamos. No entanto, para muito além do auxílio à construção de uma linguagem de programação, a linguagem humana possui um imenso

espectro de finalidades. Como nos relembra Wittgenstein (1969), “é preciso não esquecer que o jogo da linguagem é dizer o imprevisível, isto é, não se baseia em fundamentos. Não é razoável (ou irrazoável). Está aí – como a nossa vida.”

Muito se avançou no que diz respeito à construção de códigos computacionais relativos à linguagem humana. Já temos atualmente programas que são capazes de realizar uma interação comunicativa mínima, de aprender em contexto, de executar tarefas linguísticas não previamente programadas, e de, até, traduzir com certa precisão documentos originais, orais ou escritos de grande extensão e complexidade. Falta, entretanto, uma conquista importante para que computadores consigam simular com eficácia a comunicação humana. Falta uma explicação formal e operativa do modo como a linguagem funciona em sua totalidade. É o que ainda falta que torna a tarefa intrigante e desafiadora para a maioria dos cientistas que trabalham na interface entre as áreas da Linguística Computacional e Psicolinguística.

Tomemos como exemplo as chamadas “Machine Translation”, ou simplesmente os códigos computacionais capazes de traduzir. Até 2008, os sistemas mais usados de tradução, como SYSTRAN e Translate API, eram baseados sobretudo em sistema de regras linguísticas e em dicionários. Hoje, o sistema com maior número de usuários, o Google Translate, inovou em relação aos modelos antigos e aparentemente revolucionou a área, concebendo seu próprio sistema que ignora as regras linguísticas. Transformou-se em uma ferramenta de sucesso, popular na Internet, e é hoje um serviço que fornece gratuitamente uma “machine-translation” capaz de traduzir textos longos entre pares de línguas.

Diferentemente do que se esperava, quando se está diante de mais de 50 anos de pesquisa sobre as regularidades da linguagem natural, os algoritmos atuais com maior índice de sucesso não estão baseados em regras linguísticas. Surpreendentemente, o algoritmo do Google Translate, que realiza traduções entre mais de 45 línguas, não se baseia em regras gramaticais, mas em análise estatística feita inicialmente a partir do imenso corpus de documentos da União Européia, um conjunto de textos oficiais versado em seis idiomas. O sucesso do algoritmo é atribuído a Och (2003), que desenvolveu um método chamado “statistical machine translation”. Atualmente, Och é o líder do setor de tradução da empresa Google.

A tradução do “Google Translate”, que aparentemente opera em pares de línguas, antes traduz da língua fonte para o inglês e posteriormente do inglês para a língua alvo, se diferente do inglês. O inglês, entretanto, que vem sendo usado como um passo intermediário,

é uma língua humana, repleta de ambiguidades e com regras específicas que podem não possuir contraparte nas demais línguas naturais. Para ilustrar, exemplifiquemos com uma regra do português que não tem contraparte em língua inglesa, e que é um dos temas analisados nesta tese: a correferência da categoria vazia *pro*. Em língua portuguesa pode-se estabelecer correferência entre um pronome nulo anafórico (*pro*) e um antecedente discursivo, correferência esta permitida porque o sistema da língua portuguesa apresenta as características de uma língua *pro-drop*¹, ou seja, uma língua que prescinde de atualização do sujeito pronominal diante de um verbo finito, como no exemplo (1a). Em língua inglesa, exemplo (1b), o sujeito é obrigatoriamente atualizado nesta situação.

(1a) Onde **vo**çê estava? Já **ø** comprou o pão?

(1b) Where were **you**? Have **you** already bought the bread?

O Google Translate, para os exemplos (1a) e (1b), realiza com 100% de precisão a tradução na direção Inglês → Português, mas tem menos de 70% de êxito na direção oposta Português → Inglês. Por quê? O que falta ensinar a este algoritmo para que consiga 100% de êxito na tradução de qualquer tipo de texto, para quaisquer pares de língua? Será preciso apenas aumentar o banco de dados? É possível fazer emergirem todas as regras a partir de grande quantidade de dados com amostragem representativa? Se for isso, qual é o tamanho ótimo deste banco para que possa operar com metodologia estatística? E por que precisaria deste imenso volume de dados? Seria econômica e elegante esta solução? E o que necessitam fazer para incrementar a precisão da tradução e fazer diminuir seu índice de erro a um intervalo aceitável de 0.5%?

Esta tese não foi pensada com o objetivo de fornecer respostas às perguntas que acabamos de formular, mas esta introdução procura demonstrar, com base na apresentação da argumentação acima, que a tarefa de simular artificialmente a linguagem ainda está longe de ser resolvida, se desejarmos obter resultados semelhantes aos que ocorrem na comunicação humana natural. Portanto, pretendemos com esta tese contribuir para os estudos da Psicolinguística Experimental que investigam a correferência em línguas naturais, avaliando certas pistas que condicionam preferências sintáticas e semânticas durante a resolução anafórica, a partir de estudos feitos com falantes do Português do Brasil.

Entre os problemas da Linguística Computacional e os da Psicolinguística, há

¹ Redução de “pronoun-dropping” que significa “supressão de pronome”.

muita interseção e acreditamos que as áreas, embora pareçam dissociadas e muitas vezes se desenvolvam separadamente, têm muito a aprender uma com a outra. A pergunta que resta da exposição acima e que muito nos motiva e motiva os pesquisadores do processamento da linguagem natural é a seguinte: em que medida a linguagem natural é governada por um conjunto mínimo de regras? Operariam estas regras de modo serial, modular, paralelo, em rede, ou misto? Em que momento e como aprendemos este conjunto de regras? Será que aprendemos? Será que nascemos com este conjunto? Será que o configuramos a partir de um conjunto prévio de regras internalizadas, geneticamente transferidas no curso da evolução?

Claude Piron (2002), por muitos anos tradutor das Nações Unidas e da Organização Mundial de Saúde, assim abordou a dificuldade do seu ofício de tradutor:

Por que um tradutor necessita de um dia de trabalho para traduzir cinco páginas e não apenas uma ou duas horas?[...]Cerca de 90% dos textos correspondem a simples condições. Mas, infelizmente, há outros 10%. É esta parte que requer seis (ou mais) horas de trabalho. Há ambiguidades que é preciso resolver. Por exemplo, o autor de um texto fonte, um médico australiano, citou o exemplo de uma epidemia que foi declarada durante a Segunda Guerra Mundial em um “Japanese prisoner of war camp”. Estava ele falando sobre um campo norte-americano com prisioneiros japoneses ou sobre um campo japonês com prisioneiros norte-americanos? No inglês, há dois sentidos. Para resolver este pequeno problema, preciso ainda pesquisar e até talvez fazer uma ligação para a Austrália...

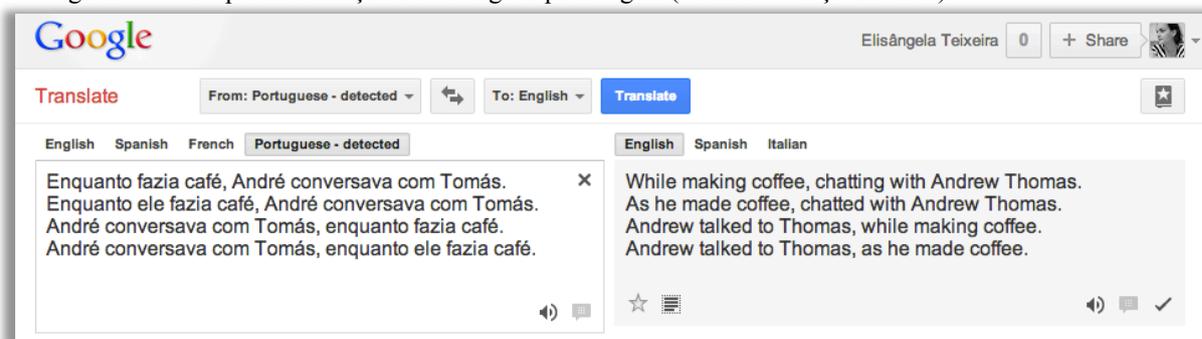
A dependência do contexto e a ambiguidade, características das línguas humanas, como bem diz Piron, não são um problema simples de resolver, o que nos faz persistir na dúvida sobre quão longe pode ir o método de Och (2003), caso não procure incorporar modelos híbridos de “machine translation”, baseados tanto em um sistema de regras linguísticas quanto em análise estatística.

Tanto o êxito quanto o índice de erro (em torno de 20%)² na qualidade das traduções do Google Translate nos leva a acreditar que a “statistical machine translation” nos informe bastante a respeito da natureza da linguagem. O tradutor do Google, apesar de ter sido e continuar a ser alimentado com bilhões de dados provenientes das mais diversas línguas, persiste em não elevar seus índices de acerto a níveis que o output (a tradução para a língua alvo) não se torne ridículo aos olhos humanos porque não reconhece características básicas das línguas tais como um sintagma nominal formado por um nome próprio, uma correferência com pronome nulo ou pleno, uma concordância dentro de um sintagma nominal ou entre sujeito e predicado, ou ainda porque não soluciona ambiguidades como a referida por

² Até agora, o Google Translate continua no topo dos sistemas com índice mais alto de precisão (entre 80% a 90% em pares como Francês <> Inglês e Italiano <> Inglês).

Piron (2002). Para exemplificar, colocamos no Google Translate um conjunto de frases usadas em um dos experimentos desta tese, cuja tradução para o inglês pode ser analisada observando-se a Figura 1:

Figura 1 – Exemplo de tradução do Português para Inglês (Feita em março de 2013).



A Figura 1 nos mostra a tradução do português para o inglês de um conjunto de quatro frases, cuja variação se dá apenas na posição da subordinada temporal e na presença ou ausência do pronome pleno para estabelecer a correferência com um dos argumentos da oração principal. Esse exemplo torna patente que o algoritmo não realiza com êxito duas das quatro orações introduzidas na língua fonte, desde que julguemos minimamente aceitáveis as demais. Tratando-se a língua fonte (o português) de uma língua *pro-drop*, ou seja, de uma língua que provavelmente permite a supressão do pronome por possuir rica morfologia verbal, e tratando-se a língua alvo (o inglês) de uma língua não *pro-drop* em orações finitas (com o verbo conjugado), o algoritmo atual ainda é incapaz de identificar e estabelecer essa característica na tradução entre este par de línguas.

A argumentação que vimos elaborando até este ponto pretende ilustrar o famoso “problema de Platão”, assim cunhado por Chomsky, para se referir ao argumento da pobreza de estímulo na aquisição da linguagem. Para Chomsky (1986), o sistema de conhecimentos específicos do adulto é tão complexo que não poderia simplesmente resultar da exposição da criança à língua nos anos iniciais da fase de aquisição. O paralelo com a “statistical machine translation” é óbvio. Os dados primários (os documentos com bilhões de exemplos em idiomas particulares) não são suficientes para explicar o sistema de conhecimento final (o output da tradução na língua alvo).

Aqueles que pensam que a mente humana opera na base de cálculos ou de um conjunto de regras pressupõem que exista um dispositivo, se não inato, ao menos representado na fisiologia cerebral, localizado espacialmente no cérebro em pontos específicos, conectados ou não em rede, cujo *modus operandi* é compartilhado por todos da

espécie humana. Esse dispositivo se assemelha ao que Hauser, Chomsky e Fitch (2002) definiram como Faculdade da Linguagem em sentido estrito. Assim, como nas primeiras definições da Gramática Gerativa, toda e qualquer investigação sobre custos de processamento da linguagem humana no cérebro parte de uma perspectiva mentalista e racionalista, que significa dizer que o objeto que está sendo aqui estudado é um sistema de regras e princípios baseados no cérebro e radicados na mente humana.

A discussão se a gramática universal é um dispositivo inato e se nascemos com um conjunto de células especializadas em executar as regras principais da linguagem, posteriormente combinadas aos parâmetros particulares de cada língua, será indiferente para os objetivos desta tese, cujo foco está unicamente nos custos de processamento, nas evidências empíricas encontradas, a partir de um conjunto de regras internalizadas, no caso, as regras da língua portuguesa. Isto porque acreditamos que o custo de processamento seja perceptível independentemente de a regra ter sido desenvolvida porque havia uma estrutura cerebral que o permitia ou porque foi registrada na matéria cerebral no curso da aprendizagem da linguagem.

O que será relevante para a discussão é o exame dos tempos de leitura evidenciados pelo comportamento ocular não-consciente (registrado experimentalmente nesta tese por meio de um rastreador ocular). Tempos que são interpretados em função da gramática ou do conjunto de regras que cada um possui para compreender e produzir interação comunicativa em uma dada comunidade linguística. O que procuramos investigar, como já anunciamos, é basicamente certas restrições e preferências na resolução da correferência, evidenciadas pelo comportamento de leitores ao estabelecerem correferência em contextos manipulados experimentalmente.

Como se sabe, muitas investigações sobre a linguagem estão baseadas em uma teoria da gramática que projeta as hipóteses de trabalho. A Teoria da Gramática Gerativa, que observa o fenômeno em si, social e individual, compartilhado em comunidade, fornece pistas de como investigar esse processo no cérebro, como relembra Raposo (1992):

[...] o sistema computacional (a gramática) actua em interação com outros sistemas conceptuais da mente humana (como, por exemplo, princípios de conhecimento do mundo, de formação e categorização de conceitos, de ‘senso comum’, etc.). Podemos conceber a gramática como um sistema nervoso central, que ordena, categoriza em padrões específicos e processa ‘a informação’ provinda destes outros módulos cognitivos. Para além destes, há ainda que ter em consideração sistemas fisiológicos de produção e percepção dos signos linguísticos (não necessariamente vocais, como no caso das línguas de sinais), sistemas de natureza social (princípios de conversação e de interação social), princípios gerais de articulação discursiva,

etc. Nesta perspectiva, qualquer fenômeno linguístico é o produto de uma interação complexa entre estes diversos módulos (incluindo o módulo computacional), sendo passível de uma análise complexa em cada um destes níveis.

Esta tese traz uma contribuição, em certa medida inovadora, porque faz uso do método de rastreamento ocular que consiste na gravação, com excelente grau de precisão, em torno de 8 milissegundos, da movimentação ocular de participantes enquanto lêem frases e textos em uma tela de computador. No Brasil, foi introduzido recentemente. Desde 2006 vem sendo utilizado na investigação do processamento da linguagem no Laboratório de Psicolinguística Experimental (LAPEX) da Universidade Federal do Rio de Janeiro (UFRJ), coordenado pelo Prof. Marcus Maia. No entanto, somente após o ano 2009, outros laboratórios brasileiros, como os da PUC-Rio, da UFC e da UFMG, passaram a fazer uso de rastreadores com maior precisão de posição e de tempo, quando adquiriram equipamentos mais acurados, que podem chegar a registrar as fixações e sacadas em valores abaixo do limiar de um milissegundo (ms).

A escolha por observar os fatores envolvidos na correferência advém do fato de, como já dito, este ainda ser um tema que requer esforços continuados que possam contribuir para sua melhor compreensão e que possam trazer esclarecimentos sobre seu papel para o processador sintático e para a resolução de restrições semânticas, pragmáticas e discursivas.

1.2 Objetivo

Esta tese tem como objetivo demonstrar experimentalmente que a posição sintática é o principal fator para a resolução anafórica em Português Brasileiro (PB), examinando a correferência anafórica em períodos complexos compostos por coordenação e subordinação. Para realizar este objetivo, procuramos evidências em um conjunto experimental formado por (i) dois experimentos de compreensão de períodos complexos, durante os quais os movimentos oculares dos participantes foram gravados; (ii) um experimento (sondagem) de produção de períodos complexos; e (iii) um estudo da movimentação ocular durante a leitura de textos não manipulados ou autênticos. O experimento com textos não manipulados objetivava estabelecer um parâmetro de comparação para os dois estudos das frases manipuladas experimentalmente. No conjunto experimental, pudemos examinar basicamente a influência de algumas preferências sintáticas e semânticas que podem produzir maiores custos de processamento durante o estabelecimento da correferência na leitura escrita. Em relação às preferências examinadas, foram observados

o paralelismo sintático e semântico, a relação de hierarquia semântica estabelecida entre antecedente e anáfora, a influência do foco ou centro, o contraste entre a resolução de pronomes nulos e plenos e a posição da expressão correferencial. As restrições obrigatórias não foram manipuladas nos experimentos. As restrições facultativas ou de preferência, por sua vez, foram manipuladas experimentalmente com o objetivo global de verificar se a entidade mais saliente ou na posição mais alta da frase, normalmente a de sujeito, era preferida na resolução anafórica em Português Brasileiro.

Maia (1997), em estudo seminal sobre o Português Brasileiro, realizou, entre outros, estudos de julgamento de gramaticalidade cujos resultados demonstraram uma correlação significativa entre o antecedente em tópico e o pronome nulo na posição de objeto. Nesta tese, pretendemos procurar evidências da correlação entre antecedente e correferente em posição de sujeito. Nossos experimentos de compreensão de frases testaram os efeitos de posição e hierarquia semântica em períodos complexos por coordenação (resolução intersentencial³) assim como foram testados os efeitos do processamento de um pronome nulo e pleno na correferência estabelecida entre o pronome e um dos dois antecedentes em competição, potencialmente correferentes, em posição de sujeito e de objeto na oração principal. Os pronomes nulos e plenos também foram testados em posição catafórica e anafórica em períodos complexos por subordinação (resolução intrasentencial). Já no experimento de leitura de textos não manipulados, observamos o tempo médio de duração das fixações oculares e calculamos o tamanho típico das palavras em língua portuguesa.

O experimento de produção presente nesta tese, - na verdade não pode ser caracterizado *stricto sensu* como um experimento, portanto, seria melhor chamarmos de sondagem -, tem como objetivo investigar qual o pronome escolhido pelos participantes para estabelecer correferência com o sujeito e objeto de uma oração principal. Levantamos a hipótese de que os participantes escolheriam pronomes demonstrativos para desfazer a ambiguidade proposta pela tarefa.

O estudo da resolução anafórica de expressões referenciais definidas em estruturas coordenadas paralelas é motivado por achados anteriores de Leitão (2005), aferidos a partir de experimentos de leitura automonitorada⁴, cuja hipótese norteadora estava baseada nos

³ Esta tese usa os termos sentença e oração como equivalentes. Os períodos complexos estudados aqui são intersentenciais (quando formados por coordenação) e intrasentenciais (quando formados por subordinação).

⁴ A leitura automonitorada é um experimento em que as palavras de uma frase são encobertas por traços ou asteriscos (correspondentes ao número de caracteres da palavra), que vão sendo desvendados à medida que o leitor avança na leitura do texto. É o leitor que controla o desvendamento de cada palavra, apertando, por exemplo, a barra espaço no teclado do computador de teste.

conceitos de centro catafórico e centro anafórico, pertencentes ao quadro teórico proposto por Grosz, Joshi e Weinstein (1995), conhecido por Teoria da Centralização.

O estudo da resolução anafórica de pronomes nulos e plenos na função de sujeito de orações subordinadas, em posição catafórica ou anafórica em relação à oração principal, encontra motivação em uma série de estudos feitos para línguas *pro-drop*, como o italiano (CARMINATI, 2002; SORACE; FILIACI, 2006), o espanhol (ALONSO-OVALLE *et al.*, 2002) e o português (CORREA, 1998; FONSECA; GUERREIRO, 2012; LUEGI, 2012; MORGADO, 2011) e na polêmica sobre a mudança pela qual o Português Brasileiro estaria passando, que sugere haver perda de morfologia verbal no PB, o que acarretaria a necessidade de atualização do pronome pleno e levaria o PB a ser classificado como uma língua parcialmente *pro-drop* (DUARTE, 1993, 1995).

O diferencial aportado pela metodologia experimental empregada nesta tese deve-se ao fato de que o uso de rastreadores oculares permite a observação de uma resolução temporal mais acurada. Através de medidas mais finas e precisas de movimentação ocular, o conjunto experimental desta tese procura contribuir com evidências empíricas para as conjecturas teóricas e experimentais já realizadas em torno da resolução anafórica em Português Brasileiro. Por meio desta metodologia experimental, pretendemos também acrescentar informações sobre os custos de processamento nas estruturas estudadas e nas estruturas vizinhas, com o objetivo de procurar estágios tardios de processamento durante o estabelecimento da correferência.

A investigação desenvolvida nesta tese se insere, como dissemos, no âmbito dos estudos psicolinguísticos que procuram prover uma explicação para os fatores que interferem no processamento anafórico, tendo como base teórica modelos de compreensão leitora, nos quais estão presentes processadores que modulam os níveis de compreensão da linguagem, que se articulam com um possível sistema de computação linguística (o *parser*), e que estabelecem forte relação com a memória de trabalho na compreensão do *input* linguístico.

Pesquisar sobre o problema da resolução anafórica significa também tentar responder aos problemas ligados à nossa capacidade de reter a informação recentemente lida em nossa memória e à nossa capacidade de integrar esta informação às demais que vão sendo ativadas durante a leitura de um texto. Os resultados de estudos experimentais acerca do problema da resolução anafórica podem corroborar ou não estudos teóricos prévios e contribuir para esclarecer uma série de questões teóricas sobre custos na integração das informações. Acreditamos que é relevante para a pesquisa na área de linguística a

incorporação de metodologias experimentais em sua agenda com a finalidade de encontrar evidências das argumentações teóricas em torno dos mesmos objetos de estudo.

A metodologia que empregamos para observar nosso objeto de estudo trabalha com um protocolo experimental mais natural, o que aumenta a confiabilidade dos dados, uma vez que os participantes se submetem aos testes diante de uma tela de computador semelhante às telas que costumam usar diariamente. Muito utilizado pela Psicolinguística, o rastreador ocular grava o comportamento natural do leitor que se sente livre para ler e reler a frase ou o texto, sem se preocupar em realizar concomitantemente qualquer outra tarefa. A naturalidade durante os testes é um fator importante tanto para a aquisição quanto para a interpretação dos dados sobre o processamento cognitivo da linguagem.

Os rastreadores oculares são instrumentos que permitem a gravação contínua dos movimentos oculares, possibilitando a identificação espaço-temporal do olhar durante a leitura de frases ou textos. A aferição do tempo que o olho permanece sobre uma palavra se torna um dado indicativo das operações cognitivas que ocorrem durante a leitura. O exame detalhado de diferentes grupos de fixações em relação ao tempo e ao espaço configuram o ponto de onde partimos para a construção de inferências acerca do modo como as estruturas estudadas são processadas. Nossos experimentos também podem solidificar a confiança no emprego de outras metodologias experimentais, apontando convergência com os resultados prévios sobre os mesmos objetos de estudo, realizados com metodologias *off-line*, como o uso de questionários, ou com metodologias *on-line*, como a leitura automonitorada.

A pesquisa em torno dos custos do processamento envolvido durante a resolução anafórica pode contribuir para responder a algumas das questões teóricas básicas sobre a arquitetura da linguagem, ajudando a melhor entender os processos cognitivos subjacentes durante a compreensão leitora.

1.3 Organização da tese

Esta tese está organizada da seguinte forma:

No capítulo 2, apresentamos as teorias e hipóteses sobre a resolução anafórica que compreendem o objeto no âmbito da gramática e do discurso, e que têm procurado evidências empíricas durante a resolução anafórica, fazendo uso de protocolos experimentais próprios da Psicolinguística. As teorias e os estudos apresentados no Capítulo 2 estão diretamente relacionados com os experimentos reportados nos Capítulos 5 e 6.

No capítulo 3, procedemos à exposição da metodologia de pesquisa experimental aqui adotada, o rastreamento ocular, explicando como as características da visão humana se tornam indícios do processamento mental da linguagem.

No capítulo 4, reportamos o estudo da movimentação ocular durante a leitura de textos não manipulados em língua portuguesa, que nos permite encontrar um padrão de tempo de fixação em função do tamanho típico das palavras da língua portuguesa.

No capítulo 5, nosso objetivo é avaliar algumas preferências semânticas e, para isso, descrevemos o experimento que manipula fatores semânticos (relação de hierarquia entre antecedente e anáfora) e sintáticos (posição estrutural do antecedente na oração) para avaliar os custos de processamento da resolução anafórica correferencial.

No capítulo 6, nosso objetivo é avaliar algumas preferências sintáticas e, para alcançá-lo, reportamos o experimento em que manipulamos o tipo de pronome na posição de sujeito de orações subordinadas e a posição da oração subordinada relativamente à oração principal. Neste capítulo também descrevemos a sondagem sobre as escolhas dos participantes na produção da correferência entre pronomes e antecedentes.

Por fim, no capítulo 7, nossas considerações finais são apresentadas a partir dos resultados encontrados pelos estudos conduzidos para esta tese. Confrontamos os resultados encontrados e nossa hipótese inicial e discutimos como as nossas análises podem lançar luz sobre o problema da resolução anafórica, esperando, com isso, contribuir para a explicação ou o delineamento de um modelo teórico que possa descrever como nosso cérebro funciona durante o processamento de expressões referenciais em língua portuguesa.

2 REPRESENTAÇÃO E PROCESSAMENTO CORREFERENCIAL

Neste capítulo, delimitamos o conceito de correferência anafórica utilizado nesta tese e apresentamos algumas das principais teorias e hipóteses, do âmbito da gramática e do discurso, que estão ligadas à resolução anafórica ou correferencial. As hipóteses e teorias aqui apresentadas estão diretamente relacionadas com as hipóteses de trabalho relativas aos experimentos discutidos nos capítulos 5 e 6.

A anáfora ou referência não é um tema novo nos estudos linguísticos ou filosóficos. No *Crátilo* de Platão, por exemplo, Sócrates, Hermógenes e o próprio Crátilo discutem vigorosamente o ato de nomear e a exatidão das palavras. É, em suma, uma obra majoritariamente voltada para a discussão da referência através da dúvida do que seja a ação de nomear. Neste diálogo, datado do século IV antes de Cristo, encontra-se o pensamento quase cético de Crátilo, discípulo de Heráclito, cuja conhecida máxima “um homem não se banha duas vezes no mesmo rio” é levada a uma profunda radicalização. Para Crátilo, o homem não chega nem a banhar-se uma única vez no rio, porque, segundo ele, a água que banha seu calcanhar já não é a mesma que molha a ponta de seus pés. O problema apontado neste diálogo refere-se à incognoscibilidade do real, à impossibilidade de nomear as coisas, porque uma vez nomeadas deixam de ser aquilo que foram antes da nomeação. A questão que coloca claramente o problema da anáfora ou da referência começa justamente em torno desta argumentação, porque, neste ponto do diálogo, Crátilo defenderá que com a linguagem é possível apenas apontar. Apontar, neste caso, significa referir-se a uma entidade dinâmica, constantemente (re)contruída no discurso, jamais atribuindo-a um correspondente fixo no mundo.

Diante deste debate, parece falsa, portanto, nossa idéia intuitiva de que a linguagem funciona como um rótulo para os objetos do mundo. Apesar de ser fácil constatar a dinamicidade da linguagem e sua escassa ancoragem na realidade, este não é de modo algum um fenômeno simples de ser compreendido nem filosófica nem cientificamente. Esta dificuldade existe, apesar de diversas explicações teóricas estarem difundidas em escritos tão antigos quanto os dos pré-socráticos ou problematizada em publicações relativamente recentes de alguns dos autores das correntes filosóficas mais influentes da segunda metade do século XX (ver FOUCAULT, 1966; DERRIDA, 1967; DELEUZE, 1953; WITTGENSTEIN, 1953).

Apesar de a discussão filosófica ser instigadora, nesta tese não trataremos do conceito de referência como já o fizeram na história da Filosofia e até mesmo da Linguística.

Antes, escolhemos, dentre os conceitos disponíveis, aqueles que pretendemos usar como ponto de partida para a nossa investigação. Na seção 2.1, definiremos o conceito de anáfora e correferência com o qual trabalharemos nesta tese. Mostraremos porque usamos os termos como sinônimos e exporemos quais são as classificações para as anáforas que serão investigadas por meio dos protocolos experimentais desta pesquisa.

Embora alguns teóricos das correntes de estudo mais próximas da Linguística Textual e da Análise do Discurso (MONDADA; DUBOIS, 1995; APOTHÉLOZ, 2001; CAVALCANTE, 2003) estejam contribuindo para uma discussão teórica profunda sobre o fenômeno da referência, esclarecendo por meio do conceito de referenciação as implicações que relevam dos estudos que compreendem o fenômeno de modo plural, considerando-o ao mesmo tempo em seus aspectos linguísticos, sociais e cognitivos, o entendimento da questão pelas pesquisas em Psicolinguística Experimental não se torna menor ou menos interessante porque circunscreve seu objeto dentro do escopo da sentença ou de sentenças adjacentes. O objetivo dos estudos experimentais em Psicolinguística é, na verdade, de natureza diversa. Como ficará claro nos capítulos de análise dos experimentos, o objeto último perseguido pela clássica e atual Psicolinguística Experimental é a descrição e explicação da arquitetura mental da linguagem a ser observada empiricamente por meio dos custos de processamento da linguagem. Tais custos são evidenciados nos tempos de reação, por meio dos registros da decisão comportamental, dos movimentos oculares não conscientes, dos potenciais de ação dos neurônios ou do fluxo sanguíneo decorrente de ativação de áreas cerebrais, captados por equipamentos cada vez mais sofisticados como os rastreadores oculares, os eletroencefalógrafos (EEG) ou os equipamentos de imagem por ressonância magnética funcional (IRMf). Portanto, sob a perspectiva da Psicolinguística Experimental, este trabalho procura avaliar, a partir de medidas de movimentação ocular, os custos de processamento durante a resolução anafórica. Tais medidas permitem inferir as relações estruturais dentro e fora do escopo da sentença e contribuem com evidências que auxiliam para o entendimento da arquitetura mental da linguagem.

Na pesquisa experimental sobre o processamento da anáfora, os pesquisadores têm selecionado, dentre as possibilidades do que possa ser entendido como expressão anafórica, uma compreensão restrita aos constituintes frasais ou quando muito a textos de pequena extensão. Expressão anafórica ou correferencial, para a pesquisa Psicolinguística, significa toda expressão substantiva, atualizada no nível da frase, que pode ocupar uma

função argumental na oração⁵. As expressões anafóricas, que serão investigadas nesta tese, são aquelas a que também chamamos de anáforas correferenciais, que constroem uma relação de co-indexação com um termo antecedente ou subsequente (no caso das catáforas), ou seja, termos que retomam a mesma referência ou entidade, como ficará mais claro a seguir. Deste modo, a retomada pode ser realizada por um sintagma nominal (SN), quer sob a forma de um item lexical (+ determinante) quer sob a forma de um pronome pleno⁶ ou nulo.

2.1 O conceito de correferência anafórica

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

A anáfora pode assumir muitas formas. Pode ser um SN idêntico ao antecedente, um SN que estabelece com o antecedente uma relação semântica de sinonímia, hiperonímia, hiponímia ou metonímia, pode ainda ser um SN completamente diferente do antecedente e pode assumir uma forma pronominal, seja a de um pronome pleno, demonstrativo ou nulo.

Como o fenômeno da anáfora está presente em todas as línguas conhecidas, sejam verbais ou visuais, acreditamos que o assunto mereça uma teoria psicológica e linguística para ser descrito e compreendido. Já há muitos anos que a anáfora vem sendo estudada por pesquisadores de diferentes escolas e abordagens, da Filosofia à Análise do Discurso. Nesse

⁵ Esta tese não usa os termos oração, período complexo e enunciado como sinônimos. O termo oração pode se referir a uma oração independente, coordenada ou subordinada, que preencha todos os espaços canônicos da frase, como sujeito e predicado. O termo período complexo se refere a um período formado por pelo menos duas orações. O período complexo pode ser composto por coordenação ou por subordinação. Já o termo enunciado é relativo a uma unidade do discurso, entendida como uma proposição que é usada em uma interação comunicativa com propósito reconhecido pelo produtor e/ou pelo interlocutor.

⁶ Na literatura da Psicolinguística, o pronome pleno também é referido como pronome lexical ou lexicalizado. Nesta tese, fazemos uso apenas do termo pleno, que corresponde ao conceito em inglês de “overt pronoun”.

espectro de pontos de vista, estão a Linguística Computacional e a Psicolinguística, por exemplo, que, apesar de serem disciplinas que trabalham à distância, seus objetivos se sobrepõem porque ambas pretendem descobrir quais os custos e os cálculos que o indivíduo ou um algoritmo precisam realizar para resolver uma anáfora encontrada no âmbito da sentença, do texto e do discurso.

A abordagem Psicolinguística é especialmente importante para esta tese porque vêm dela o objeto e a metodologia desta investigação. A Psicolinguística se interessa pela resolução anafórica porque quer compreender as representações mentais e os processamentos que o cérebro de um indivíduo realiza para produzir ou interpretar uma anáfora.

Para compreender como a anáfora se apresenta em língua portuguesa, examinaremos, a seguir, alguns tipos de anáforas em ordem de complexidade. Os exemplos são traduzidos⁷ de Nand (2012):

- (1) **Júlio_i** adora **seu_i** carrinho de brinquedo.
- (2) **Pedro e João_i** adoram **morangos_j**; e **eles_i** **os_j** comem frequentemente.
- (3) **Um garoto_i** e **uma garota** entraram no quarto. **O garoto_i** era alto e usava um chapéu.

A anáfora ilustrada pelo exemplo (1) é o caso prototípico e provavelmente de maior ocorrência nas línguas. Em Português Brasileiro está em distribuição com o uso dos pronomes possessivos “dele” e suas derivações. Em (1) a anáfora é o pronome possessivo “seu” que se refere à entidade previamente mencionada “Júlio”. Neste exemplo a resolução é trivial, uma vez que não há outro candidato para a anáfora e que existe uma relação direta entre pronome e antecedente. A complexidade aumenta no exemplo (2) porque há mais de um candidato para o pronome “eles” e “os”. O leitor se depara com os pronomes “eles” e “os” do gênero masculino e plural que, em potencial, podem estabelecer correferência com qualquer um dos dois SNs antecedentes, que aqui chamamos de competidores, e que no exemplo (2) são identificados na função de sujeito composto “Pedro e João” e na função de complemento verbal “morangos”.

A resolução anafórica na presença de competidores não é simples. Neste exemplo, a informação morfológica, de gênero e número, não é suficiente para a identificação. Muito

⁷ Frases originais de Nand (2012) (1) Jason loves his toy car. (2) Peter and John love apples and they eat them often. (3) A boy and a girl entered the room. The boy was tall and wore a hat. (4) I bought a Honda Civic. The car is red in color. (5) John bought a louse. The windows are wooden. (6) John drove into an electricity pole last night. The accident caused a blackout.

provavelmente a informação sintática forneça pistas importantes para a co-indexação entre “eles” e “Pedro e João”, nas posições mais altas da frase, de sujeito, e para a co-indexação entre “os” e “morangos”, já que ambos ocupam a posição de objeto. As estruturas coordenadas também são paralelas e fatores como frequência podem interferir para que o processador sintático realize a resolução sem maiores custos. A informação semântica do verbo “comer” também ativa os traços obrigatórios para o SN na posição de sujeito deste verbo. Esta posição rejeita o SN “morangos” como sujeito. No entanto, não é difícil de imaginar que a língua pode criar estruturas gramaticalmente aceitas que invertam a relação, fazendo com que o pronome “eles” retome “morangos” enquanto o pronome “os” poderia retomar o sujeito composto.

Além do uso de pronomes, como ilustram os exemplos (1) e (2), a anáfora pode se dar sob a forma de sintagmas nominais (SN) com núcleo lexical, como no exemplo (3), em que o SN “O garoto” se refere a “um garoto” anteriormente mencionado na primeira oração. No exemplo (3) o único antecedente possível para co-indexar “o garoto” é o SN “um garoto”. Neste exemplo, marca-se a diferença na retomada por um artigo definido. Note-se que o nome comum, introduzido na oração, sob a forma do SN “um garoto” é posteriormente retomado como anafórico, determinado por um artigo definido, já que a segunda menção ao referente confere à informação o status de dado e não de novo⁸. Estas aparentes sutilezas são características das anáforas, que em estudos que as examinam sob perspectivas não mentalistas, como a corrente funcionalista, seriam tomadas como relevantes. Para os algoritmos computacionais que procuram resolver a anáfora, é necessária a indexação dos nomes comuns, o que torna, portanto, o critério de definitude bastante útil para a identificação dos antecedentes. Os exemplos de (1) a (3) são prototípicos e relativamente simples de serem resolvidos tanto pelo processador da linguagem natural quanto por um algoritmo computacional, apesar da problematização exposta. Perceba-se que nos três exemplos há completa correspondência entre antecedente e anáfora, ou seja, ambos são correferenciais e co-indexados.

A correferência é um conceito que precisa ser esclarecido nos estudos que tratam da anáfora. Nesta tese trataremos basicamente da anáfora correferencial. A anáfora correferencial é a retomada de uma entidade anteriormente mencionada no discurso. Esta retomada pode se dar na forma de um pronome, de um demonstrativo, de um nulo ou na forma de um sintagma nominal. A anáfora correferencial que estabelece relação com uma

⁸ Para detalhes sobre os conceitos de dado e novo, ver Givón (1979).

entidade mencionada posteriormente também é chamada de catáfora. O importante a notar aqui é que para haver correferencialidade é preciso que anáfora e antecedente refiram-se a apenas uma entidade.

Até o exemplo (3), o estabelecimento da correferência acontece independentemente do conhecimento de mundo do leitor. As regras morfológicas, sintáticas e semânticas são suficientes para encontrar a correferência. Examinemos abaixo os próximos três exemplos, também extraídos de Nand (2012):

- (4) Eu comprei **um Honda Civic**. **O carro**_i é vermelho.
- (5) Alfredo comprou **uma casa**. **As janelas** são de madeira.
- (6) Carlos colidiu contra um poste de eletricidade ontem à noite. **O acidente** causou um apagão.

Ao ler o exemplo (4), percebemos que o SN “o carro” estabelece com o SN “um Honda Civic” uma relação de retomada. Sabe-se que “o carro” está co-indexado com “um Honda Civic”. No entanto, podemos perceber que o estabelecimento desta correferência ultrapassa o limite da informação morfológica (ambos são masculino e singular). Esta pista leva o leitor da frase a inferir que os SNs se referem à mesma entidade. É preciso, entretanto, de mais informação do que a que está na superfície da frase para estabelecer esta correferência. É preciso conhecer o significado de “Honda Civic”, o que nos parece fácil e imediato porque, dentro do conjunto de modelos de carro que muitas pessoas atualmente conhecem, está o modelo Honda Civic. Uma vez satisfeita a condição de o leitor conhecer o modelo Honda Civic como pertencente à categoria carro, a correferência se estabelece entre dois SNs que se referem à mesma entidade. Do ponto de vista da Semântica formal e de demais sistemas de processamento desta informação, por exemplo, este não seria um problema para resolver, já que bastaria a informação sobre “Honda Civic” pertencer ao mundo que contém a frase do exemplo (4), ou seja, o sistema linguístico precisaria ser alimentado com esta informação para operar sem problemas.

O conceito de entidade para a definição de anáfora correferencial torna-se ainda mais relevante quando examinamos o exemplo (5). O SN “uma casa” está claramente relacionado ao SN “as janelas”, no entanto não tratam da mesma entidade. “As janelas” estabelece com “uma casa” uma relação semântica de metonímia ou de ingrediência, que significa dizer que dentre as partes ou ingredientes que constituem o SN “uma casa” está o SN

“as janelas”. Veja-se que, no exemplo (5), falar de retomada constitui um problema porque “as janelas” e “uma casa” não são a mesma entidade, não possuem a mesma referência. Neste exemplo, aquilo que pode ser identificado como elemento anafórico é uma parte do antecedente. O SN “as janelas”, portanto, não retoma “uma casa”, antes faz uma remissão, configurando um tipo de relação que muitos pesquisadores denominam de anáfora indireta ou anáfora associativa para diferenciar da anáfora correferencial (KOCH; MARCUSCHI, 1998; MARCUSCHI, 2004; CAVALCANTE, 2004, 2003; KOCH, 2006; CIULLA, 2002, 2008). O caso do exemplo (5) é também resistente a tratamentos formais dentro da gramática gerativa, como nos alerta Raposo (1992), ao dizer que o problema ainda não havia recebido uma solução satisfatória no âmbito da Teoria de Regência e Ligação (CHOMSKY, 1981). Nesta tese, trataremos do processamento de anáforas correferenciais, ou seja, as que retomam o antecedente, as que se referem à mesma entidade.

O conceito de entidade também não se aplica ao exemplo (6). Neste caso, o SN “o acidente” retoma não uma entidade mas um evento declarado na oração precedente, ou seja, “o acidente” se refere à oração inteira e não a um SN. “O acidente” estabelece com a oração anterior uma relação de nominalização (KOCH, 2006), ou seja, o SN nomeia o evento declarado pela oração precedente.

A exemplificação acima não é exaustiva e muitos outros tipos de relações, que poderiam ser classificadas como anafóricas em sentido *lato*, existem na língua. Na interação comunicativa, é possível observar modos de remissão e retomada não discutidos com base nestes seis exemplos. A idéia era trazer apenas alguns exemplos aparentemente simples para demonstrar que mesmo estes se mostram resistentes a uma explicação satisfatória de como se dá a resolução anafórica. Os exemplos de (1) a (6) apresentam uma gradação de complexidade de resolução. Para um algoritmo ou a mente humana serem capazes de co-indexar dois SNs é preciso mais do que informação morfológica, sintática, semântica e pragmática, como deve ter ficado claro na problematização acima. É provável que o conhecimento de mundo e o papel da memória de trabalho, além de outros indícios, quer prosódicos, gestuais ou interacionais, sejam necessários para uma resolução eficiente da anáfora e para o incremento de informação necessária para a progressão da interação comunicativa ou textual.

Muitos estudos experimentais sobre a referência e a anáfora correferencial têm sido feitos, sobretudo para a língua inglesa (GERNSBACHER, 1989; SANFORD; LOCKHART, 1990; GORDON; GROSZ; GILLIOM, 1993; GORDON; HENDRICK, 1997;

ALBRECHT; CLIFTON, 1998; KAUP; KELTER; HABEL, 2002; KOH; CLIFTON, 2002; SANFORD *et al.*, 2004). A língua inglesa, diferentemente da língua portuguesa, é apenas parcialmente uma língua *pro-drop*, porque somente permite a supressão do pronome em alguns casos de subordinadas infinitas ou em algumas situações de oralidade. Já a língua portuguesa é considerada *pro-drop*, uma vez que pode suprimir o pronome sujeito em orações finitas. Contudo, algumas pesquisas apontam para a existência de indícios de perda desta característica (DUARTE, 1995; BARBOSA; DUARTE; KATO, 2005), alegando que na morfologia verbal houve neutralização de desinências número-pessoais e que consequentemente tem havido ganho na atualização do pronome sujeito diante de verbos conjugados. A língua portuguesa tem uma conjugação verbal complexa, em que não há a obrigatoriedade de atualização de um SN correferencial ou de um pronome na função de sujeito da frase, por exemplo. Os estudos, recentemente publicados, sobre correferência anafórica para a língua portuguesa tratam tanto da investigação da correferência lexical quanto da correferência pronominal (LEITÃO, 2005, 2008; LUEGI; COSTA; FARIA, 2007; COSTA; LUEGI, 2009; MORGADO, 2011; LUEGI, 2012, entre outros).

Como a língua portuguesa é um sistema que apresenta diversidade na produção da correferência anafórica, é mister investigar o papel da anáfora correferencial em suas formas de retomada do antecedente, seja por SN com núcleo nominal, seja por retomada com pronome pleno ou nulo.

Na próxima seção (2.2), passaremos a apresentar algumas teorias linguísticas que procuram explicar a resolução anafórica no âmbito da gramática. Na seção seguinte (2.3), apresentaremos os resultados de trabalhos teóricos e experimentais que investigaram os custos de processamento desta resolução no âmbito do discurso.

2.2 Teorias e hipóteses sobre o processamento correferencial na gramática

Na seção 2.2.1, apresentaremos brevemente alguns conceitos da Teoria da Ligação. Trataremos do conceito de anáfora, que é de natureza diferente do conceito usado nesta tese, dos princípios de ligação, como o Princípio B, que trata da ligação do pronome, e do conceito de categoria vazia [*pro*], estudada em um dos nossos experimentos. Em seguida, na seção 2.3.4, discorreremos sucintamente sobre a Hipótese da Posição do Antecedente, proposta por Carminati (2002), que observa as preferências de retomada em estruturas intra-

sentenciais, de um ponto de vista que atribui ao cálculo sintático a indexação da correferência para os pronomes nulos.

2.2.1 Teoria da Ligação

A Teoria da Ligação (CHOMSKY, 1981) descreve as condições das dependências referenciais entre SNs (nomes, pronomes pessoais, reflexivos e recíprocos), com o objetivo de obter um conjunto de regras que explique como se dá a interpretação destes elementos. A dependência referencial está ligada à situação linguística na qual é possível atribuir a um SN a referência de outro SN, que normalmente o antecede mas que também pode sucedê-lo. A esta relação de dependência referencial é que é dada o nome de correferência.

A correferência é estabelecida entre expressões linguísticas que designam entidades ou eventos⁹ (RAPOSO, 1992). No entanto, é importante observar, se lembrarmos dos exemplos da seção 2.1, que estabeleciam correferência entre orações justapostas ou coordenadas, que a Teoria da Ligação tratará da correferência dentro do escopo da oração¹⁰. Como nos lembra Raposo (1992), dentro da tradição gerativa, a Teoria da Ligação representa “o módulo da teoria gramatical que estuda as propriedades estruturais e semânticas das relações de dependência referencial”.

A teoria diferencia basicamente três tipos de sintagmas nominais que podem estabelecer correferência. São as anáforas¹¹ “se”, “si próprio(a)”, “ele mesmo(a)(s)”, “uns com os outros”; os pronomes “ele”, “ela”, “isto”, “seu”, “um”, “os”, etc. e as expressões-R(eferenciais) “nomes próprios”, “nomes comuns”.

Vejamos abaixo cada um dos tipos de SNs com potencial de referência, para em seguida examinarmos as restrições obrigatórias e as facultativas relativas aos exemplos que seguem:

(7) **Alberto_i** retirou-se_i da sala.

⁹ Por entidade, entende-se pessoas, animais, coisas, idéias, ou seja, as representações das entidades do universo discursivo de uma língua. O uso do termo evento pretende neste contexto dar conta de estados, ações, acontecimentos nomeados, descritos ou declarados na forma de SNs ou de orações.

¹⁰ Oração, neste contexto, está sendo usada como equivalente ao termo sentença, que é muito utilizado nos estudos de Psicolinguística. Neste ponto, também pretendemos chamar a atenção do leitor para diferenciar as teorias que constroem uma explicação para a correferência no domínio da sentença (frase, oração) daquelas que o fazem no domínio do discurso (entre orações, entre parágrafos, etc).

¹¹ O termo anáfora dentro da Teoria da Ligação designa apenas os pronomes reflexivos e recíprocos. Seu uso dentro da Teoria da Ligação tem um significado diferente do sentido lato com o qual estamos trabalhando nesta tese (ver seção 2.1).

- (8) **Alberto_i** pensou que **ele_{i;j}** estava errado ao fazer isso.
- (9) O chefe **de Alberto_i** deve querer conversar com **seu funcionário_{i;j}** em breve.

No exemplo (7), observa-se um caso de anáfora em que o pronome reflexivo “se” retoma obrigatoriamente o SN “Alberto” como antecedente. Ambos estão co-indexados¹². No exemplo (8), o caso tratado é o de pronome. Em (8), em contraste com a restrição obrigatória de (7), a relação entre “Alberto” e o pronome “ele” pode ser de correferência, mas não necessariamente. A co-indexação é representada pelo índice “i”, enquanto o índice subscrito “j” mostra a possibilidade de disjunção. Isto significa que os pronomes podem designar outras entidades que não estão presentes dentro do escopo da oração e que não necessariamente ligam-se a qualquer SN antecedente dentro da mesma oração, mesmo que concordem em gênero e número. No exemplo (8), podemos imaginar que “ele” pode ser um amigo de “Alberto”, chamado “Ricardo” ou qualquer outro ser com capacidade de decisão para poder potencialmente estar errado, desde que este outro ser seja uma entidade cujos traços concordem em gênero e número com o pronome “ele”.

No exemplo (9), assim como em (8) e em (7), a expressão referencial “Alberto” é o terceiro tipo de SN com potencial de referência. Em (9) também temos “O chefe de Alberto” e “seu funcionário” como exemplos de expressões-R. Tais expressões são caracterizadas porque possuem um núcleo lexical, com sentido próprio, e porque funcionam autonomamente. A autonomia das expressões-R refere-se ao fato de que sua referência está nas entidades do universo discursivo de uma língua. Sobre o estabelecimento de correferência dentro do escopo da oração, as expressões-R têm mais dificuldade para retomar como antecedente uma outra expressão-R à sua esquerda. Esta capacidade de retomada é mais própria dos pronomes, como pode ser visto no exemplo (8). Se colocássemos “Alberto” no lugar de “ele”, em (8) teríamos uma oração agramatical, como no exemplo (10).

- (10) * **Alberto_i** pensou que **Alberto_i** estava errado ao fazer isso.

Se imaginarmos que Alberto trabalha em uma empresa, que nesta empresa há dois funcionários chamados Alberto e que a oração (11) declara algo sobre os dois Albertos, então

¹² Co-indexar significa atribuir um índice arbitrário (no caso o “i” e o “j” subscritos) que simboliza a correferência entre dois SNs. Se dois SNs possuem índices distintos isto quer dizer que não são correferentes, isto é, que possuem referência disjunta.

o segundo “Alberto” seria outra entidade do universo discursivo, o que tornaria a oração gramatical. Seriam apenas caracterizados como homônimos.

(11) **Alberto**_i pensou que **Alberto**_j estava errado ao fazer isso.

Os três tipos de SN com potencial de referência, acima descritos, podem ser resumidos numa relação de gradação das restrições ou dependências que podem ir da obrigatória a uma dependência mais limitada ou fraca, representada pela figura abaixo:

<p>Dependência obrigatória → Anáfora</p> <p>Dependência possível → Pronome</p> <p>Dependência fraca → Expressões-R</p>

Os pronomes e as anáforas para a teoria estão em distribuição complementar, uma vez que anáforas e seus antecedentes devem estar contidos necessariamente em uma mesma oração simples, o que não se aplica aos pronomes e seus antecedentes, como pode ser visto nos exemplos abaixo, extraídos de Raposo (1992).

- (12) **[A Maria]**_i confia em **[si própria]**_i.
- (13) * **[A Maria]**_i pensa que o Luís confia em **[si própria]**_i.
- (14) * **[A Maria]**_i confia **n[ela]**_i.
- (15) **[A Maria]**_i pensa que o Luís confia **n[ela]**_i.

As orações estão dispostas em pares contrastivos. A oração (12) em contraste com a (13) demonstra que a anáfora “si própria” precisa estar contida na mesma oração simples e, ainda, precisa estar em uma posição mais baixa do que seu antecedente para poder estabelecer correferência com ele. Do contraste de (14) e (15) observamos que os pronomes podem estabelecer correferência com seus antecedentes desde que estejam em uma subordinada encaixada, ou seja, desde que sua posição não seja relativa à posição do seu antecedente.

Para entender melhor o que significa estar em uma posição mais baixa ou mais alta ou ainda em posição independente ou relativa, passemos a examinar um conceito importante, introduzido por Reinhart (1983) que é a noção de *c-comando*.

qualquer domínio¹³. O conceito de domínio e de domínio local (FIGURA 2), dependente da noção de c-comando e de ramificação, é relacionado ao conceito de ligação local, derivado, por sua vez, do conceito de ligação. A ligação local é particularmente diferente da ligação porque serve para caracterizar os casos em que há uma categoria interveniente entre a ligação de dois sintagmas nominais co-indexados. Em nosso estudo sobre a correferência de pronomes nulos e plenos, trabalhamos com orações em que há uma categoria interveniente, que perturba a ligação do pronome pleno ou nulo com o antecedente (ver Capítulo 6), causando ambiguidade.

Os Princípios A e B explicam porque as orações dos exemplos (13) e (14) são agramaticais. No entanto, como já vem sendo demonstrado, tais princípios, tal como estão hoje definidos, podem apresentar falhas na análise de estruturas sintáticas de vários idiomas (REINHART, 1983; GRODZINSKY; REINHART, 1993; REULAND, 2001, 2003), sempre que a restrição aplicada não resulta em estruturas agramaticais, além de não dar conta de todos os cálculos que o processador sintático precisa executar para resolver a correferência no escopo da oração.

2.2.1.1 As categorias vazias ou os pronomes nulos

Como vimos na seção precedente, as categorias estudadas pela Teoria da Ligação com propriedades referenciais são as anáforas, os pronomes e as expressões-R. A língua portuguesa, no entanto, licencia um quarto tipo de categoria referencial que são os pronomes nulos. As propriedades primitivas propostas por Chomsky (1988, *apud* RAPOSO, 1992) para caracterizar as expressões referenciais são os traços de [\pm anáfora] e [\pm pronome]. Com base nas combinações possíveis destes traços, as anáforas se caracterizariam como [+ anafórica] e [- pronominal], os pronomes como [- anafórico] e [+ pronominal], as expressões-R como [- anafórica] e [- pronominal], e, finalmente, a categoria vazia, tratada nesta tese como *pro* (ou pronome nulo), é caracterizada com os traços [+ anafórica] e [+ pronominal].

A categoria vazia seria uma categoria mista que precisaria obedecer ao mesmo tempo ao Princípio A e B. No entanto, a categoria vazia possui uma classificação que a torna paralela às categorias fonéticas estudadas. O paralelo fica, então, estabelecido entre a anáfora e o vestígio de SN, o pronome e o *pro* (pronome em posição argumental de sujeito de verbos

¹³ A realidade psicológica dos Princípios B e C foi estudada em Português Brasileiro por Maia, Garcia e Oliveira (2012).

finitos), as expressões-R e o vestígio de QU, e a categoria vazia e o PRO (pronome em posição argumental de sujeito de orações infinitas).

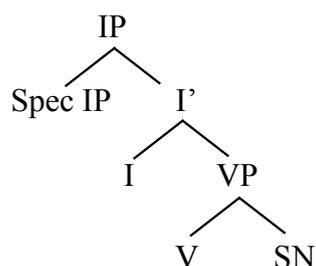
Em línguas românicas como o Português, o Italiano e o Espanhol, a categoria vazia *pro* é encontrada na posição de sujeito nulo. Um exemplo de uma contraparte fonética é o pronome pessoal. Sempre que uma língua possui uma morfologia verbal rica que marca a pessoa do verbo, pode-se encontrar a categoria *pro*, que teria como contraparte qualquer um dos pronomes pessoais da língua portuguesa, como “eu”, “tu”, “ele”, “nós”, etc. Em nosso estudo do processamento da correferência do pronome pleno e nulo apenas examinamos o *pro* cuja contraparte é o pronome de 3ª pessoa do singular “ele” ou “ela”.

A categoria *pro*, portanto, obedeceria ao Princípio B da Teoria da Ligação. Montalbetti (1984) conjectura que a categoria *pro* apresentaria uma preferência de interpretação. Diante de dois antecedentes potenciais, a categoria *pro* selecionaria preferencialmente o antecedente que a c-comanda. No caso de correferência com um antecedente que não estabelece a relação de c-comando, a língua preferiria sua contraparte fonética, ou seja, o pronome pessoal “ele”, que aqui designamos apenas por pronome pleno.

2.2.2 Hipótese da Posição do Antecedente

A Hipótese da Posição do Antecedente foi proposta por Carminati (2002), com base no pressuposto de que existe uma divisão de trabalho entre os pronomes em posição estrutural de sujeito sob a forma nula (categoria vazia *pro*) e sob a forma plena. Tal divisão de trabalho leva os pronomes nulos a preferirem estabelecer correferência com o antecedente mais proeminente em contextos intra-sentenciais. A partir do exame da língua italiana, Carminati (2002) afirma que a proeminência do antecedente anafórico em contexto intra-sentencial é determinada pela posição sintática de Spec IP. Esta é a posição pré-verbal do sujeito oracional, a mais alta e mais proeminente em relação às outras posições na estrutura bimembre da árvore sintática.

Carminati (2002) testou sua hipótese em uma ampla gama de experimentos, *on-line* e *off-line*, manipulando as condições dos antecedentes que ocupam a posição de Spec IP na estrutura da sentença. Abaixo, vemos uma estrutura oracional em árvore, que ilustra a posição Spec IP.



Entre os experimentos conduzidos por Carminati (2002), a autora investigou os sujeitos nulos e pronominais em orações subordinadas temporais, em contextos ambíguos com um ou dois referentes. A Hipótese da Posição do Antecedente, confirmada pela análise dos dados do conjunto de experimentos que conduziu, se sobrepõe às hipóteses de que o pronome nulo é correferenciado com o sujeito da oração por um princípio de economia ou porque estaria em oposição ao pronome pleno que tem como função desfazer a ambiguidade, ligando-se a posições mais baixas da oração. O que Carminati (2002) sustenta a respeito do pronome pleno é que a correferência que estabelece é menos estável e mais dependente de contexto do que a correferência estabelecida pelo pronome nulo.

Seus achados para a língua italiana, como já dito, uma língua *pro-drop* como a língua portuguesa, sugerem que a resolução anafórica é sensível aos fatores estruturais que conferem à posição de Spec IP uma proeminência que a torna a candidata preferencialmente escolhida como antecedente.

2.3 Teorias e hipóteses sobre o processamento correferencial no discurso

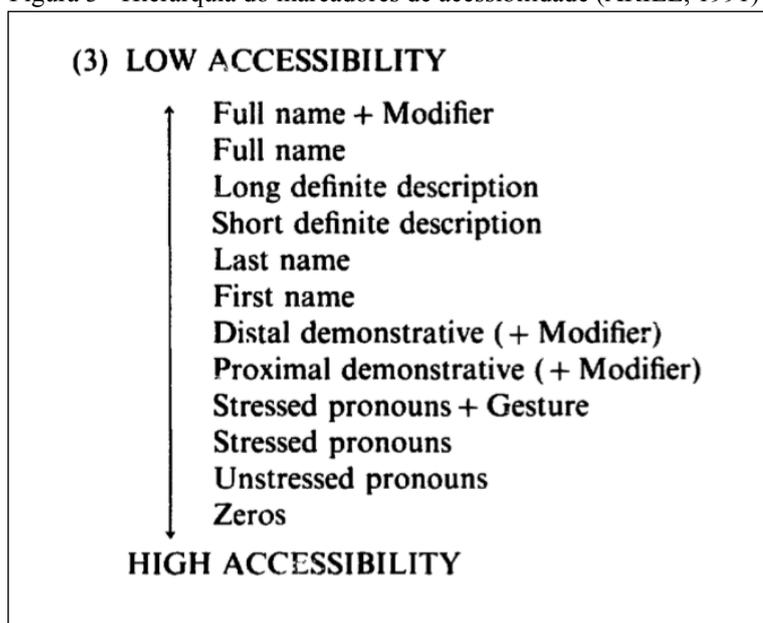
Nesta seção, examinaremos algumas teorias sobre a anáfora e alguns trabalhos em Psicolinguística que procuraram provar algumas das predições destas teorias. As teorias discutidas não se circunscrevem ao âmbito de uma teoria da gramática e não se atêm, portanto, a fatores exclusivamente sintáticos. A maioria delas leva em conta implicações discursivas e cálculos que ultrapassam as fronteiras da sintaxe. Na seção 2.3.1, apresentamos a Teoria da Acessibilidade de Ariel (1991) e discutimos sua implicação para o processamento anafórico. Na seção 2.3.2, apresentamos a Teoria da Centralização que, a partir do conceito de centro, estabelece parâmetros para compreender a coerência discursiva. Em seguida, explicamos sucintamente alguns achados provenientes de estudos experimentais que procuraram validar as predições da Teoria da Centralização. Na seção 2.3.3, apresentamos a Hipótese da Carga Informacional de Almor (1999), procurando identificar sua contribuição em relação às teorias discutidas.

2.3.1 Teoria da Acessibilidade

Ariel (2001) vem discutindo, desde seus trabalhos iniciados em meados da década de 1980, que a anáfora, entendida do ponto de vista discursivo e da oração (ARIEL, 1991), pode ser compreendida por meio da noção de acessibilidade aos conteúdos armazenados na memória de trabalho. Sua proposta, atualmente conhecida por Teoria da Acessibilidade, pressupõe que as representações mentais dos referentes discursivos (entidades) não são igualmente acessíveis pelo destinatário¹⁴ em qualquer estágio do discurso. Para Ariel (1991) as expressões anafóricas seriam, portanto, marcadores de acessibilidade que podem ser classificados segundo uma gradação hierárquica.

Baseada em evidência empírica de estudos que observaram, em línguas não relacionadas entre si, a distribuição dos tipos de expressões anafóricas em diversos textos e que realizaram estudos de julgamento de gramaticalidade, a principal contribuição de Ariel (1985, 1988, 1990, 1991, 1996, 2001) é afirmar que o uso de diferentes expressões referenciais, como os sintagmas nominais definidos, os pronomes demonstrativos, os pronomes plenos e nulos, é uma função do grau de acessibilidade dos seus antecedentes. A classificação de gradação de acessibilidade proposta para as anáforas correferenciais, que vai da baixa acessibilidade para a alta acessibilidade, pode ser melhor compreendida ao observarmos abaixo a reprodução de seu esquema original.

¹⁴ Ariel (1991, 2001) usa o termo “addressee” para o que se entende como a segunda parte do discurso, o leitor, o ouvinte, o destinatário, em oposição ao produtor do discurso, o qual ela denomina “speaker”. Nesta tese, optamos pelo termo destinatário para traduzir “addressee”, quando resenhamos as teorias sobre anáfora que a compreendem no nível discursivo.

Figura 3 - Hierarquia do marcadores de acessibilidade (ARIEL, 1991)¹⁵

As descrições definidas, como o sintagma nominal “o garoto” no exemplo (3) (seção 2.1) marcam baixa acessibilidade relativamente aos pronomes “seu” e “eles”, exemplos (1) e (2) (seção 2.1), que marcam maior grau de acessibilidade.

Com base em três princípios: (i) a informatividade, (ii) a rigidez e (iii) a atenuação, Ariel (1991) argumenta que a escala acima (FIGURA 3) de codificação das expressões referenciais em graus de acessibilidade é derivada da gradação de cada um dos princípios. Para a Teoria da Acessibilidade, as expressões definidas longas¹⁶, por exemplo, que procuram referentes menos acessíveis são as mais informativas, mais rígidas e menos atenuadas.

Os três princípios fazem com que a explicação da escala (FIGURA 3) demonstre a não arbitrariedade da gradação. “O grau de informatividade incorporada em um marcador linguístico é crucial para seu papel como um marcador de acessibilidade, quanto maior for a informação lexical de um marcador, melhor será a recuperação de um material menos acessível.” (ARIEL, 1991, p. 449)

O oposto também é verdadeiro na escala. Quanto mais vazio semanticamente é o marcador, maior é sua função em recuperar um antecedente com alto grau de acessibilidade.

¹⁵ Tradução da Figura 3. Escala que vai da baixa à alta acessibilidade: Nome completo (+ modificador) → Nome completo → Descrição definida longa → Descrição definida curta → Sobrenome → Prenome → Demonstrativo distante (+ modificador) → Demonstrativo próximo (+ modificador) → Pronomes tônicos (+ gesto) → Pronomes tônicos → Pronomes átonos → Zeros.

¹⁶ Uma expressão ou descrição definida longa é formada de mais de duas palavras de conteúdo (itens lexicais), enquanto uma curta teria até dois itens lexicais.

Para Ariel (1991), as consequências de aplicar o critério de informatividade é construção de um contraste entre funcionalidades, por exemplo o contraste entre o pronome nulo e o pleno. No entanto, a informatividade sem estar vinculada aos demais critérios não permite a gradação de todos os tipos de marcadores. A rigidez, que é o critério que aponta a dependência de contexto, ou seja a característica de um marcador se referir a apenas uma entidade, é usada como um critério que pode distinguir uma escala em que há marcadores mais ou menos rígidos. A rigidez, entendida como gradação, permite por exemplo distinguir os nomes próprios de descrições definidas.

As demais distinções são facultadas pela entrada do terceiro critério, que é a atenuação. Segundo Ariel (1991), o conceito de atenuação é bastante similar à proposta de tamanho fonológico de Givón (1983). Não seria, no entanto, o tamanho da palavra que desempenharia um papel na marcação de acessibilidade, mas sim a atenuação da pronúncia. A atenuação também se refere em certo aspecto à tonicidade, já que o critério serve aos propósitos da Teoria para distinguir entre pronomes átonos e pronomes tônicos. O critério da atenuação é controverso, mas essa discussão não é relevante para os propósitos desta tese.

Ariel (1991) também defende que sua teoria está baseada nas evidências empíricas de estudos psicolinguísticos que medem tempos de leitura em testes que manipulam distância, foco e classes de anáforas (ANDERSON *et al.*, 1983; LI; THOMPSON, 1979, PURKISS, 1978 *apud* ARIEL, 1991). Os estudos reportados por Ariel (1991) encontraram tempos de leitura menores para as expressões referenciais com alta acessibilidade. Em vista disso, a Figura 3 é sugerida como uma escala que pretende ser universal, uma vez que os estudos que a embasam foram feitos para línguas não relacionadas.

Além das evidências psicolinguísticas, os achados provenientes de estudos de julgamento de gramaticalidade por falantes nativos (BROADBENT, 1973 *apud* ARIEL, 1991) procuraram observar o critério de proeminência sintática e discursiva, através da medida de preferência na correlação entre anáfora e antecedente. Na observação, os antecedentes poderiam ou não ocupar o lugar de tópico. Mais uma vez nestes casos foi encontrada uma preferência para a correferência entre anáfora e tópico.

Para compreender bem a escala de acessibilidade proposta por Ariel (1991), além dos três critérios de codificação, das evidências psicolinguísticas e dos estudos de julgamento de gramaticalidade, os graus de marcação devem ser compreendidos relativamente uns aos outros. Isto quer dizer que as estruturas não possuem graus absolutos, mas que a escala varia em função das línguas, que podem comportar-se diferentemente. Ariel (1991) defende que as

expressões anafóricas só podem ser classificadas em relação aos seus antecedentes, já que é esta relação que vai determinar se há ou não correferência.

Tabela 1 - Reprodução da tabela de Ariel (1991) que apresenta a distribuição de expressões anafóricas no texto.

Expressão	Contexto			
	Mesma Oração	Oração prévia	Mesmo parágrafo	Entre parágrafos
Pronome	110 = 20.8%	320 = 60.5%	75 = 14.2%	24 = 4.5%
Descrição definida	4 = 2.8%	20 = 14.1%	65 = 45.8%	53 = 37.3%

Na tabela 1, podemos ver que os pronomes são correferentes de entidades muito próximas, quer dentro da mesma oração (20.8%), quer na oração prévia (60.5%). O percentual de presença dos pronomes como correferentes no contexto entre parágrafos, por exemplo, decai radicalmente, chegando a 4.5%. Ariel (1991) ainda faz um cálculo da distribuição dos pronomes, retirando dos dados os antecedentes que também exerciam a função de tópico discursivo. A queda passa a ser ainda mais abrupta, neste caso, porque o uso dos pronomes entre parágrafos vai para um índice de apenas 0.37%. Portanto, a teoria da acessibilidade afirma que a proeminência do antecedente é uma função que permite a maior velocidade de recuperação.

Já as descrições definidas, os casos dos sintagmas nominais com núcleo lexical, estão claramente distribuídas no âmbito do parágrafo ou entre parágrafos. Na Tabela 1, não reproduzimos as demais expressões referenciais porque não serão estudadas nesta tese, mas se confrontarmos apenas os pronomes e as descrições definidas poderíamos até pensar que se encontram próximo ao que seria uma perfeita distribuição complementar.

Os argumentos básicos que sustentam a Teoria da Acessibilidade giram em torno dos critérios acima expostos e de como é necessário considerá-los em conjunto para a ordenação dos marcadores em graus. Contudo, Ariel (1991) não deixa de considerar relevantes para a explicação da acessibilidade fatores outros como a característica de certas entidades no discurso serem *per si* mais salientes do que outras, como é o caso de uma entidade com o traço [+ animado], normalmente preferida na correferência anafórica.

2.3.2 Teoria da Centralização

Tomando por base as preferências discursivas, tanto de elocução quanto de compreensão, Grosz, Joshi e Weinstein (1995) propuseram um quadro teórico com o objetivo de tentar caracterizar a coerência local discursiva, construindo um modelo com o objetivo de servir a alguns propósitos do processamento computacional da linguagem natural. Esta teoria, conhecida por Teoria da Centralização, parte de trabalhos prévios (GROZS; SIDNER, 1986; GROZS; JOSHI; WEINSTEIN, 1983, 1986 *apud* GROZS; JOSHI; WEINSTEIN, 1995) que estabeleceram uma distinção entre três componentes da estrutura do discurso¹⁷: uma estrutura linguística, uma estrutura intencional e um estado de atenção. No nível da estrutura linguística, o discurso pode ser dividido em segmentos discursivos, dependentes entre si, que se organizam sob a forma de uma estrutura de constituintes de enunciados. A estrutura intencional se refere às relações entre as intenções, que são responsáveis pela conexão da argumentação racional do discurso. O estado atencional modela o foco de atenção dos participantes em qualquer ponto do discurso. As mudanças no estado atencional são uma função da estrutura intencional e das propriedades dos enunciados da estrutura linguística.

A Teoria da Centralização lida, portanto, com as relações estabelecidas entre o foco de atenção, as expressões referenciais escolhidas para dar continuidade ao discurso e a percepção de coerência entre os enunciados de um segmento discursivo. O objetivo principal é apresentar um modelo de componente local¹⁸ de um estado atencional.

Para explicar a Teoria da Centralização é necessário examinar detidamente as interações entre a coerência local e a escolha das expressões anafóricas. A qualidade de tais interações podem fazer variar a coerência, o que ocasiona sobrecarga ao sistema inferencial. Para os autores, a retomada de entidades anteriormente introduzidas é um dos fatores que determinam a coerência local. É importante ressaltar que outras teorias na época de publicação dos trabalhos de Grosz e colaboradores, já procuravam compreender a estrutura discursiva local (HALLIDAY; HASAN, 1976; KINTSCH; VAN DIJK, 1978).

Examinemos a variação dos segmentos de (1) a (3), extraídos de Grosz, Joshi e Weinstein (1995), para em seguida compreendermos as noções principais da teoria. Os enunciados que sobrecarregam o sistema inferencial estão marcados em itálico. Já os elementos correferentes estão indexados e marcados em negrito.

¹⁷ Grosz, Joshi e Weinstein (1995) definem discurso como sendo um conceito maior do que uma mera sequência de enunciados. Dizem que para uma sequência de enunciados ser um discurso, é preciso que tenham coerência.

¹⁸ Componente local é um segmento de discurso que deve possuir a propriedade de coerência local, coerência que se estabelece entre os enunciados do segmento.

Segmento de um discurso (1)

a. O Pedro_i às vezes comete gafes. **b. Ontem** foi um dia lindo e **ele_i** estava motivado para velejar em seu novo barco. **c. Ele_i** queria que **Tom_j** o acompanhasse em sua expedição marítima. **d. Ele_i** telefonou-**lhe_j** às 6 horas da manhã. **e. Ele_j** ficou doente e furioso por ter sido acordado tão cedo.

Segmento de um discurso (2)

a. O Pedro_i às vezes comete gafes. **b. Ontem** foi um dia lindo e **ele_i** estava motivado para velejar em seu novo barco. **c. Ele_i** queria que **Tom_j** o_i acompanhasse em sua expedição marítima. **d. Ele_i** telefonou-**lhe_j** às 6 horas da manhã. **e. Tom_j** ficou doente e furioso por ter sido acordado tão cedo. **f. Ele_j** disse um desaforo ao **Pedro_i** e desligou. **g. Claro que ele não tinha a intenção de chatear o Tom.**

Segmento de um discurso (3)

a. O Pedro_i às vezes comete gafes. **b. Ontem** foi um dia lindo e **ele_i** estava motivado para velejar em seu novo barco. **c. Ele_i** queria que **Tom_j** o acompanhasse em sua expedição marítima. **d. Ele_i** telefonou-**lhe_j** às 6 horas da manhã. **e. Tom_j** ficou doente e furioso por ter sido acordado tão cedo. **f. Ele_j** disse um desaforo ao **Pedro_i** e desligou. **g. Claro que Pedro não tinha a intenção de chatear o Tom_j.**

Os três segmentos acima mostram uma gradação de aceitabilidade da última oração de cada segmento, que vai da mais baixa aceitação (segmento 1) à aceitação máxima (segmento 3). Para entender estes níveis de aceitabilidade, a Teoria da Centralidade propõe dois conceitos-chave que explicam a escolha das expressões referenciais para a retomada de antecedentes mais salientes. A saliência é um fator importante para esta teoria que entende ser o sujeito mais saliente do que o objeto, que, por sua vez, é mais saliente que as demais funções sintáticas.

Os conceitos-chave carregam parte do nome da teoria, que é a noção de centro. Na teoria, propõe-se que existam dois tipos de centro: os centros catafóricos (C_{ca}), que seriam uma espécie de SN com núcleo lexical, que introduzem no segmento discursivo conteúdo referencial, e os centros anafóricos (C_{an}), que retomariam os centros catafóricos, fazendo uso de pronomes ou categorias vazias, constituindo no discurso a coerência e a coesão

discursivas. Os enunciados (E) podem ter mais de um centro catafórico, como é o caso de uma oração cujo verbo seleciona dois argumentos. Ambos os argumentos são potencialmente centros catafóricos, dependendo somente de como o discurso será continuado. No entanto, um enunciado terá somente um único centro anafórico, já que a noção de centro está estreitamente ligada à noção de intenção. Isto não quer dizer que um enunciado não possa retomar dois antecedentes sob a forma de dois pronomes. A existência de apenas um centro anafórico está relacionada ao fato de que apenas um dos dois pronomes de uma oração como “**Ele**_i telefonou-**lhe**_j às 6 horas da manhã” é escolhido para dar continuidade discursiva.

Entendida a noção de centro, vejamos como aplicá-la na interpretação dos segmentos (1), (2) e (3). Na oração (a), dos três segmentos, identificamos o SN “O Pedro” como o centro catafórico, que é posteriormente retomado pelo centro anafórico, expresso sob a forma do pronome “ele” nas orações (b), (c) e (d). A oração (c) introduz um potencial centro catafórico, o SN “Tom”, mas este potencial não se realiza nas orações (d) dos três segmentos. A oração (d) traz o pronome “lhe” como correferente de “Tom” introduzido na oração (c), mas não o torna centro anafórico, porque o centro anafórico da oração (d) é “Ele”, que retoma “O Pedro”.

A oração (e) do segmento 1, tem como sujeito o pronome “Ele” que correferente com “Tom” introduzido na oração (c). Há neste ponto do segmento (1) uma diminuição de aceitabilidade porque o pronome “Ele” retoma uma entidade que não é o centro anafórico do segmento, ou seja, não está no centro de atenção. A oração (e) provoca, portanto, uma ruptura da coerência local. O leitor, ao se deparar com o SN “doente”, é obrigado a reanalisar o segmento sob pena de não compreendê-lo. A ruptura interfere, portanto, no sistema de atenção.

Se a intenção era operar uma mudança de centro, a expressão referencial a ser selecionada para a oração (e) deveria ser a retomada do SN “Tom”. Assim, como podemos constatar na leitura do segmento (2), não há quebra de coerência local na oração (e) do segmento (2), que realiza a mudança de centro de atenção, atualizando o SN “Tom” na posição de sujeito.

Observemos que, após operar a mudança de centro na oração (e), nos segmentos (2) e (3), a oração (f) já retoma o centro catafórico de (e), que é o SN “Tom”, por meio de um pronome, sem causar sobrecarga ao sistema de inferência. Como a enunciação da oração (g) do segmento (2) acontece após a mudança de centro, espera-se que o pronome “ele” seja

correferente de “Tom”. Novamente, na oração (g) percebemos uma ruptura. O pronome “ele” da oração (g) retoma “O Pedro” que neste ponto do discurso não é mais o centro catafórico.

Para não haver ruptura na oração (g) do segmento (3), uma segunda mudança de centro tem de acontecer com a retomada do SN “O Pedro”, reintroduzindo-o no segmento discursivo como centro catafórico.

O segmento (3) apresenta duas mudanças de centro. Tais mudanças demonstram que, para a Teoria da Centralização, qualquer expressão referencial, esteja ela em posição de sujeito ou outra, pode ser o centro catafórico de um discurso, retomado por um centro anafórico que garante continuidade referencial.

Como vimos ao analisar os três segmentos acima, o pronome aparece como a melhor e mais frequente forma para desempenhar a função de centro anafórico. Nos exemplos explicados por Grosz, Joshi e Weinstein (1995), os centros anafóricos retomam em geral a entidade mais proeminente. O objetivo ao retomar o mais proeminente é manter a coerência local. Nos casos de mudança de centro que analisamos, o SN em posição de objeto é, em primeiro lugar, colocado em proeminência, para depois ser retomado em sua forma inalterada, estando apto a correferir com pronomes na sequência ou continuidade discursiva.

A descrição destas características não pretende ser exaustiva, antes quer sugerir que seriam as condições mais frequentes e, portanto, ótimas, para uma alta performance de um *parser* que tenha como função processar a correferência baseado em regras discursivas.

Para sistematizar a análise das rupturas e transições analisadas nos segmentos acima, a Teoria da Centralização propõe uma classificação dividida em três tipos de relações de transição que podem se dar entre pares de enunciados. São elas: (i) a continuação do centro, que seria a continuação do mesmo centro anafórico através das sequências de enunciados, como se observa na sequência das orações (b), (c) e (d) dos segmentos (1), (2) e (3); (ii) a retenção do centro; que significa a permanência do centro anafórico em posição menos saliente em um enunciado, mas que provavelmente não será mais retomado no enunciado seguinte, como entre as orações (f) e (g) do segmento (3); e (iii) a mudança de centro, que provoca a ruptura, uma vez que o centro anafórico não se atualiza na sequência discursiva, como se observa na oração (e) do segmento (1).

As rupturas, observadas nos segmentos (2) e (3), referem-se à violação de duas regras formuladas pela teoria sobre as restrições de realização e mudança de centro. Examinemos a regra 1, transcrita abaixo, sobre as restrições de realização de centro:

Regra 1: se qualquer elemento de um centro catafórico de um enunciado n $C_{ca}(E_n)$ é realizado por um pronome em E_{n+1} , então o centro anafórico de um enunciado posterior $C_{an}(E_{n+1})$ deve ser igualmente realizado por um pronome. (GROSZ; JOSHI; WEINSTEIN, 1995)

O que esta regra estipula, em síntese, é que não há qualquer elemento em um enunciado que possa ser realizado como pronome a não ser que o centro anafórico deste enunciado também seja realizado como pronome. A regra 1, parece óbvia para quem observa a função de um pronome no texto, mas sua relevância está na formulação de uma regra, com potencial de aplicação em algoritmos que procurem implementar um módulo de resolução anafórica. Para os estudos sobre o *parser*, é igualmente importante porque se revela uma função da referência pronominal, ou seja, o uso de um pronome para realizar um centro anafórico é um sinal para o ouvinte de que o falante está continuando a falar sobre a mesma coisa.

A segunda regra proposta pela teoria procura aplicar uma restrição de mudança de centro.

“Regra 2: sequências de continuidade são preferidas em relação às sequências de retenção, e sequências de retenção devem ser preferidas em relação às sequências de mudança. Em particular, um par de continuidade através de E_n e através de E_{n+1} , representado como $Cont(E_n, E_{n+1})$ e $Cont(E_{n+1}, E_{n+2})$ respectivamente, é preferido em relação ao par de retenções, $Ret(E_n, E_{n+1})$ e $Ret(E_{n+1}, E_{n+2})$. O caso é análogo para o par de retenções e o par de mudanças.” (GROSZ; JOSHI; WEINSTEIN, 1995)

A regra 2 é, na verdade, uma hipótese de que há preferência nas mudanças de sequências de centro, e não tem o mesmo poder de implementação da regra 1. Antes, pretende ser um reflexo de nossa intuição que procura minimizar o número de mudanças de centro, já que a mudança acarreta um custo inferencial maior do que a continuidade ou a retenção.

Pesquisas na área da Psicolinguística têm demonstrado a validade do centro anafórico, que em inglês é realizado pelo pronome, mas que, em línguas *pro-drop*, é realizado por suas contrapartes sem forma fonética, as categorias vazias. No caso da Língua Portuguesa é realizado pelo pronome nulo na posição de sujeito.

Em seguida, vamos examinar duas destas pesquisas da área da Psicolinguística que procuram verificar a eficácia da Teoria da Centralização por meio de evidências empíricas. A primeira a ser reportada é a que estabelece uma sobrecarga no sistema de inferência quando viola a regra 1, por não dar continuidade de centro por meio de um pronome, mas sim por repetição do centro catafórico. Esta violação que ficou conhecida como Penalidade do Nome Repetido (ou do SN repetido) foi inicialmente investigada por Gordon,

Grosz e Gillion (1993). A segunda é a que testa a teoria examinando a função do paralelismo estrutural para a coerência discursiva em experimentos conduzidos por Chambers e Smyth (1998).

2.3.2.1 Penalidade do Nome Repetido

Gordon, Grosz e Gilliom (1993) elaboraram um conjunto de cinco experimentos para investigar as principais previsões da Teoria da Centralização. Procuraram demonstrar que o pronome é a expressão referencial majoritariamente preferida, em detrimento da descrição definida ou do nome, para estabelecer a continuidade em um segmento discursivo. Seus resultados mostraram que há apenas um centro anafórico preferido, realizado sobretudo por pronome na posição gramatical de sujeito do enunciado. Os autores também constataram que, apesar de haver um conjunto de centros catafóricos hierarquizados em relação à sua proeminência, a determinação do centro anafórico não é alterada em torno da noção de proeminência.

Seus resultados fornecem uma evidência importante para dissociar as relações catafóricas e anafóricas dos processos de coerência discursiva.

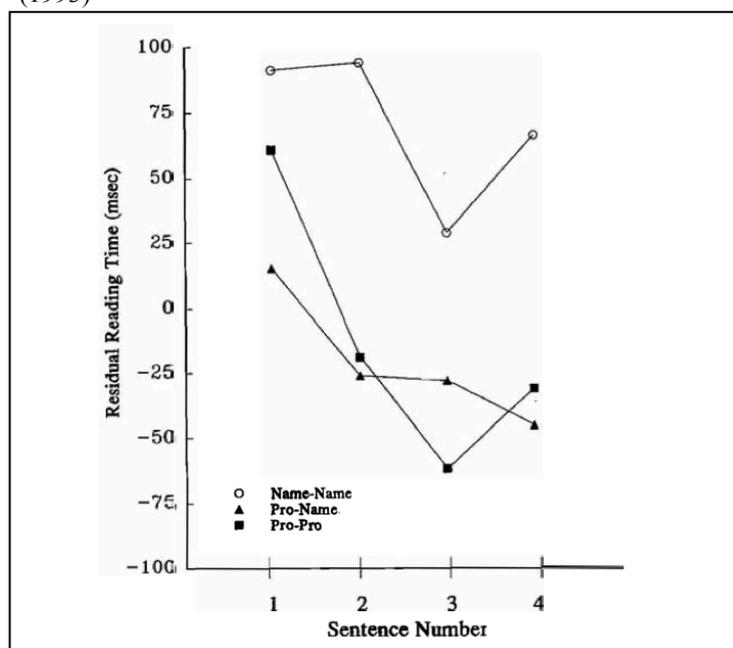
Os resultados de Gordon, Grosz e Gilliom (1993), reproduzidos na Figura 4, permitem, por meio dos tempos de leitura, a comparação de continuidade de centro entre as três estruturas testadas: (1) nome-nome; (2) pronome-nome; e (3) pronome-pronome. Pode-se notar um considerável aumento nos tempos de leitura para a condição “nome-nome”. A repetição do nome nesta condição, no lugar do pronome, preferido como elemento de ligação entre os enunciados, causa estranhamento ao leitor e leva provavelmente, em um primeiro momento, à identificação de uma referência disjunta, ou seja, à interpretação do nome repetido como a introdução de um novo centro catafórico. No entanto, esta hipótese não se sustenta para além da reanálise porque, nos testes conduzidos, o segmento do discurso tratava de promover a coerência entre as orações, como pode ser visto no conjunto de estímulos utilizados, reproduzidos abaixo em sua língua original. Os índices subscritos 1, 2 e 3, referem-se às três condições experimentais testadas.

Segmento de um discurso (4)

a. Bruno was the bully of the neighborhood. **b. He₁/He₂/Bruno₃** chased **Tommy** all the way home from school one day. **c. He₁/He₂/Bruno₃** watched **him₁/Tommy₂/Tommy₃** hide behind

a big tree and start to cry. **d. He₁/He₂/Bruno₃** yelled at **him₁/Tommy₂/Tommy₃** so loudly that all the neighbors came outside.

Figura 4 – Resultados do experimento 1 de Gordon, Grosz e Gilliom (1993)



Apesar de nesta tese não ser testado um segmento discursivo como apresentado pelo estudo acima, menos ainda na condição de nome repetido, o conjunto experimental proposto em nosso trabalho examinará a correferência entre uma expressão referencial na forma de descrição definida como correferente anafórico de uma outra expressão definida e que estabelece com a anterior uma relação semântica de hiperonímia ou hiponímia, tanto na posição de sujeito quanto na de objeto. Estas condições nos permitem verificar, ainda que sob diferentes designs experimentais, as predições dos trabalhos de Grosz e colaboradores para a língua portuguesa. Além disso, o experimento sobre a retomada por pronomes e nulos em condição de total ambiguidade, impõe-nos a importante tarefa de perceber se a retomada em língua portuguesa é preferencialmente realizada pela categoria vazia, o pronome nulo, e se o pronome pleno tenderia a ser correferenciado com um centro catafórico potencial, mas não proeminente sintaticamente.

2.3.2.2 Paralelismo Estrutural

Os efeitos do paralelismo estrutural na coerência discursiva foram investigados a fundo por Chambers e Smyth (1998), em um estudo que se tornou importante porque testou o paralelismo em estruturas sintáticas congruentes e verificou os custos do processamento de retomadas na posição de sujeito e na de objeto.

Alguns estudos prévios ao de Chambers e Smyth (1998), como os de Halliday e Hasan (1976), já haviam demonstrado que o uso do pronome para se referir a uma entidade proeminente é uma pista explícita de que o enunciado, em que ocorre a retomada por pronome, está claramente relacionado ao contexto. Halliday e Hasan (1976) sugerem ainda que o uso de uma expressão mais informativa, como um nome, para o estabelecimento da correferência, torna mais difícil a inferência do tipo de relação que se pretende estabelecer entre os enunciados.

Assim como o estudo de Halliday e Hasan (1976), várias investigações experimentais já estavam procurando demonstrar o enviesamento que a correferência entre pronome e antecedentes proeminentes, observados em estruturas paralelas, pode provocar na análise dos resultados de tempo de leitura (GARVEY; CARAMAZZA; YATES, 1976; CRAWLEY; STEVENSON; KLEINMAN, 1990; SMYTH, 1994; GORDON; SCEARCE, 1995).

O estudo de Chambers e Smyth (1998) não é, nesse aspecto, uma novidade, mas seu diferencial reside nos ganhos que sua análise de resultados aporta para a discussão do papel do paralelismo na produção da coerência discursiva. Os autores investigaram a hipótese de que haveria facilitação e, conseqüentemente, maior velocidade de processamento na compreensão de estruturas paralelas. Uma vez detectada esta facilidade, os autores procuraram entender como a Teoria da Centralização explicaria os efeitos encontrados.

Chambers e Smyth (1998) conduziram, então, três experimentos de leitura automonitorada, para testar, basicamente, os efeitos da estrutura de enunciados e expressões referenciais anafóricas na compreensão do discurso. Para isso, examinaram os fatores de paralelismo em pares de enunciados, tais como:

- (16) Leonardo entregou a Michel um sanduíche.
Depois, **ele**/Carla deu a Carla/**ele** uma maçã.
- (17) Martin disse para Liz verificar o óleo.

Depois, **ele**/*Martin*/*Dean*/**ela**/*Liz*/*Dean* disse a(o) *Dean*/**ela**/*Dean*/**ele**/*Martin* para verificar o radiador.

(18) Débora acertou o David no nariz.

Depois, *Débora*/**ela** golpeou-**lhe**/*David* (n)as costelas.

Os exemplos (16), (17) e (18) são os estímulos pertencentes a cada um dos três experimentos. As estruturas testadas são paralelas mas os antecedentes podem não ser paralelos (exemplo 17). Entre todos os experimentos, foram manipuladas as estruturas em torno da variação das seguintes condições: pronome na posição de sujeito (16); pronome na posição de não sujeito (16); anáfora do sujeito com e sem antecedente paralelo (17); anáfora de não sujeito com e sem antecedente paralelo (17); e orações que retomam os antecedentes na forma de nome-nome; pronome-nome; nome-pronome; pronome-pronome.

Os resultados analisados em conjunto sugeriram que, independentemente do papel ou posição gramatical do pronome, este estabeleceria coerência discursiva com antecedentes paralelos em enunciados paralelos. Já nas condições não paralelas, o uso de pronomes provocou tempos maiores de leitura das frases. De um modo geral, os resultados não confirmam a noção de hierarquia entre os centros catafóricos, como proposto pela Teoria da Centralização, porque os pronomes retomam com eficácia qualquer antecedente, indiferentemente da ordem em que são mencionados. Isto quer dizer que o pronome potencialmente retoma qualquer um dos centros catafóricos, quer este centro esteja na posição de sujeito ou em posição argumental menos saliente. Não haveria para a correferência entre pronome e antecedente maiores custos em função da acessibilidade dos antecedentes. O que os resultados demonstram amplamente é que o paralelismo estrutural entre enunciados facilita bastante a coerência semântica entre pronome anafórico e antecedente, desde que realizados na mesma posição. Os autores ainda encontraram tempos de leitura bem superiores para as condições em que a anáfora era a repetição do antecedente, caso que não foi atestado apenas em estruturas não paralelas e em estruturas em que a repetição não se dava na posição de sujeito.

Este trabalho de Chambers e Smyth (1998) critica a possível extensão da Teoria da Centralização, ao afirmar que há um viés (o paralelismo) na formulação das noções de centro anafórico e catafórico. No entanto, a teoria de Grosz e colaboradores não foi formulada com a intenção de abordar todas as situações e estruturas linguísticas. Parece quase óbvio que está contida em um conjunto muito pequeno de situações, o que chamam de segmentos de um

discurso. Acreditamos, ainda assim, que a Teoria da Centralização é bastante esclarecedora, pois percebemos o ganho que aporta ao identificar regularidades que se traduzem formalmente para explicar o processamento otimizado durante a resolução anafórica. Os demais problemas, identificados em estudos posteriores, como o de Chambers e Smyth (1998), que ainda não receberam tratamento formal, ficam à espera de uma modelização que possa ser implementada em sistemas automáticos inteligentes.

2.3.3 Hipótese da Carga Informacional

A Hipótese da Carga Informacional, formulada por Almor (1999), trata do sintagma nominal anafórico visto por uma perspectiva psicológica e não estritamente linguística. Almor (1999) procura investigar as relações que os sintagmas nominais anafóricos estabelecem com a memória de trabalho. Não delimita sua abordagem ao escopo da frase, mas se interessa pelas ativações e processamento da informação semântica no âmbito do discurso.

Sua hipótese de trabalho entende a anáfora como um processo de otimização baseado no custo de processamento, que define como sendo os custos de ativação da informação semântica.

Para Almor (1999), a anáfora tem uma função discursiva. Os experimentos que conduziu, fazendo variar o foco discursivo e as relações semânticas entre anáfora e seus antecedentes, o levaram a concluir que o custo de processamento dos sintagmas nominais anafóricos é um reflexo de sua justificativa funcional. Encontrou que as anáforas com menor custo de processamento possuíam maior justificativa funcional. Seus resultados corroboram a Teoria da Acessibilidade (ARIEL, 1991) que apresenta uma gradação de expressões anafóricas em função da acessibilidade do referente.

A Hipótese da Carga Informacional de Almor (1999), em suma, defende que o processamento das expressões anafóricas reflete a máxima da quantidade, um dos princípios pragmáticos griceanos (GRICE, 1975), segundo o qual a informação deve ser dada na medida em que é necessária, ou seja, a complexidade da informação deve ser proporcional ao seu propósito comunicativo.

Mas, segundo Almor (1999), para compreender sua teoria, é preciso lançar mão de mais duas afirmativas, além da máxima da quantidade de Grice (1975). A primeira diz respeito à complexidade. Na Hipótese da Carga Informacional, a complexidade é representada pela medida da carga informacional, uma noção que expressa as restrições da memória de

trabalho para armazenar e processar simultaneamente a informação verbal. A segunda afirmativa diz respeito à informação. A informação veiculada pela expressão anafórica pode ser de dois tipos ou a combinação de ambos: informação necessária para identificar os antecedentes e/ou informação nova.

O princípio pragmático que descreve o processamento anafórico de acordo com a Hipótese da Carga Informacional é, portanto, o seguinte: a carga informacional de uma anáfora em relação a um antecedente dado deve ter uma justificativa funcional, a de auxiliar a identificar o antecedente ou a de adicionar nova informação ao antecedente, ou ambas. (ALMOR, 1999)

O papel da memória de trabalho¹⁹ é uma noção fundamental para esta teoria. A memória de trabalho possui uma limitada capacidade, que é usada tanto para armazenar quanto para processar informação. Por isso, muitos pesquisadores acreditam que o uso de expressões anafóricas no discurso é um reflexo desta limitação. Já foi provado que o aumento da distância entre um antecedente e uma anáfora sem conteúdo semântico, como pronomes e categorias vazias torna o processamento extremamente custoso (SANFORD; GARROD, 1981). A superespecificação de uma expressão anafórica, que retoma um antecedente próximo ou altamente acessível, pode levar igualmente a uma sobreposição de informação que esgota os recursos da memória de trabalho.

No entanto, como esclarece Almor (1999), o uso de uma expressão anafórica do tipo “descrição definida”, que aporta informação nova ao antecedente em foco, nem sempre é desfavorecido, como pode ficar sugerido pelo modo como os resultados de Gordon, Grosz e Gilliom (1993) foram interpretados. Almor (1999) esclarece que, no trabalho sobre a penalidade do nome repetido (ver seção 2. 3.2.1), os autores encontraram tempos maiores na leitura de expressões anafóricas repetidas e concluíram, a partir destes achados, que se tratava da violação da Regra 1 da Teoria da Centralização, que diz que o pronome deve ser preferido para retomar um centro catafórico ou, em outras palavras, um antecedente em foco.

Segundo Almor (1999), se a interpretação dos dados de Gordon e colegas estivesse correta, deveria haver penalidade para qualquer expressão anafórica com conteúdo semântico, ou seja, penalidade para qualquer expressão anafórica que não fosse um pronome (nulo ou pleno).

¹⁹ A memória de trabalho é um sistema mental que permite a ativação e manipulação de múltiplas partes de informação transitória. Uma definição precisa sobre o conceito de memória de trabalho foge aos objetivos desta tese.

Almor (1999) defende, portanto, que não há penalidade do nome repetido, uma vez que a repetição de um nome para retomar um antecedente ou o uso de uma descrição definida como expressão anafórica pode ter uma justificativa funcional no discurso. Almor (1999) não contradiz os resultados de Gordon, Grosz e Gilliom (1993), que muito provavelmente têm razão quando se compara o custo de um pronome com o de um sintagma nominal repetido. Sua crítica se centra, sobretudo, na interpretação dos achados sobre os maiores custos no contexto experimental testado. Assim, o conjunto experimental de Almor (1999) sugere que o custo de processamento de uma expressão anafórica sob a forma de uma descrição definida deve ser avaliado em relação à sua funcionalidade no discurso.

Este capítulo apresentou as principais teorias e hipóteses que guiam nossa investigação. Definimos o conceito de correferência anafórica no âmbito da gramática e do discurso. Nosso interesse não foi o de discutir os conceitos e seus fundamentos, mas o de apresentar os conceitos que se tornam operacionais na investigação que empreendemos nesta tese. Optamos por apresentar neste capítulo de fundamentação os trabalhos a partir dos quais muitas investigações sobre o Português Brasileiro foram conduzidas. Nos capítulos 5 e 6, assim como em nossas considerações finais, procuramos articular nossa investigação com os trabalhos já realizados para a língua portuguesa.

3 METODOLOGIA EXPERIMENTAL

Neste capítulo procedemos à exposição da metodologia de pesquisa experimental aqui adotada, o rastreamento ocular, explicando como as características da visão humana se tornam indícios do processamento mental da linguagem. Discutiremos também as razões pelas quais a metodologia tem sido usada pela Psicolinguística para esclarecer diversas questões teóricas sobre a compreensão leitora, auditiva e visual de linguagens verbais e não verbais. Finalmente, apresentaremos as hipóteses que serão testadas nesta tese e o protocolo de coleta de dados empregado.

O objetivo deste capítulo é, em suma, prestar esclarecimentos sobre as decisões metodológicas por nós adotadas. Acreditamos que este esclarecimento seja importante porque estamos tratando de estudos que lidam também com os mais básicos processos mentais implicados no âmbito da compreensão leitora, ou seja, os processos biológicos. Isto não quer dizer que nossos processos biológicos não estejam permeados por processos culturais. Pelo contrário. Os estudos aqui apresentados não sustentam ou defendem a preponderância de um ou outro processo.

A abordagem experimental do processamento da linguagem escrita delimita o seu foco de interesse na compreensão de uma fase essencial: a percepção visual da palavra escrita. Posteriormente analisa o acesso lexical, a integração das palavras em uma oração, a integração do sentido entre orações e entre segmentos discursivos, até a compreensão total do enunciado completo.

Quando falamos de percepção visual da palavra escrita, estamos tratando de um processo psicológico de nível básico, que implica processamentos do tipo *bottom-up*. Treiman (2001) define o processo *bottom-up* na leitura como sendo aquele que recebe o estímulo de fora para dentro, ou seja, como um processo de nível básico que, em primeiro lugar, recebe o estímulo visual, para depois transformá-lo em imagens no córtex visual primário. Ainda no cortex visual primário, a informação das imagens procura as áreas de identificação de letras, para posteriormente ser enviada às regiões cerebrais de identificação de palavras, de onde são distribuídas para serem tratadas em áreas especializadas no processamento da linguagem, seja em nível oracional ou discursivo. Um processo *bottom-up* é, em suma, aquele que ocorre sem que anteriormente sejam demandados conhecimentos de mundo ou linguísticos para o início do tratamento da informação visual em forma de palavras.

O processo *top-down* se caracteriza pela demanda por informações

superordenadas, da ordem do conhecimento enciclopédico (mundo, linguístico, etc) durante o fluxo de leitura. Também chamado de antecipações, o processamento *top-down* é um movimento que prediz o que será percebido fisicamente pelo olho. Trata-se, em uma síntese superficial, do inverso do processamento *bottom-up*. O processo *top-down* se refere ao tratamento da informação guiado principalmente pelo conhecimento individual anterior à interação com o estímulo visual.

Processos psicológicos mais complexos, do tipo *top-down*, que sucedem ou mesmo antecedem a percepção visual, podem ser posteriores e anteriores à codificação da imagem visual em forma linguística. Na leitura, as antecipações, também conhecidas por preeditividade, são consideradas parte de um processamento *top-down*, que ocorre independentemente de haver um processamento *bottom-up* anterior ou posterior de decodificação do estímulo visual. São os processos *top-down* que fazem com que uma dada forma linguística se decodifique em sentido contextual e se relacione com as demais informações da memória de trabalho, ativadas em um dado evento comunicativo. O tempo total de processamento de um termo lido é entendido como o somatório do tempo dos processos *bottom-up* e *top-down*, já que a decisão do leitor de passar para o próximo termo ocorre somente depois de ele ter compreendido o sentido do termo lido e o integrado às informações previamente lidas e presentes em sua memória de curto prazo.

Há muitas controvérsias em relação a qual dos dois processos seria preponderante na ação de ler. Colocar-se de um lado ou de outro significa sustentar ou não a preeditividade na leitura (RAYNER; POLLATSEK, 1989). Apesar de ser importante fazer referência a este debate, porque é controverso e implica em pressuposições bastante distintas sobre como se processa a linguagem no cérebro, seu conteúdo, no entanto, não será objeto deste capítulo.

3.1 A fisiologia da visão

A visão é responsável por grande parte da cognição humana. Os olhos são os órgãos sensoriais que nos permitem ver. Para perceber as imagens, nossos olhos possuem uma parte chamada retina, que está repleta de dois tipos de células fotorreceptoras: (i) os cones, que são as células que percebem as cores; e (ii) os bastonetes, as células que regulam a entrada de luminosidade, responsáveis pelo processamento das imagens em preto e branco. Os cones se concentram na parte da retina chamada de fóvea, que ocupa um espaço pequeno de apenas 4 mm. Apenas neste espaço reservado à fóvea é que conseguimos ver com detalhes.

Aquilo que vemos em torno da fóvea, numa região conhecida como parafóvea, ainda guarda um grau bastante elevado de detalhe, mas não tem acuidade de cores nem de contorno de formas. Mesmo tendo uma menor precisão de detalhes, a parafóvea permite reconhecer objetos, desde que aquilo que ela processa seja uma imagem já conhecida por nosso cérebro. Corriqueiramente, os seres humanos pensam que percebem com riqueza de detalhes tudo o que está a seu redor. No entanto, o olho só vê com detalhes o que está em foco, ou seja, aquilo que está sendo percebido pela fóvea, esta minúscula região da retina, do tamanho da cabeça de um alfinete, que recebe a projeção das imagens percebidas a uma angulação de apenas dois graus. Nosso cérebro nos faz crer que estamos vendo uma grande área porque nosso olhar é não estacionário, significa dizer que se move a todo instante, e porque a cada nova posição produz vários focos em frações de segundos, escaneando as imagens de seu entorno e armazenando-as em uma memória de curto e longo prazo, que complementa ininterruptamente as visões que formamos do mundo.

São os impulsos nervosos dos cones e bastonetes que enviam a informação através do nervo ótico para o córtex visual primário que, por sua vez, gera a visão dos objetos. Esta geração da imagem, para aqueles que estudam a linguagem escrita, é de crucial importância, pois é o momento em que o cérebro reconhece a forma visual como linguística e a relaciona ao inventário de termos de uma dada língua, armazenados na memória de longo prazo, que, como se sabe, não tem uma localização específica no cérebro, mas que é provável que se localize próximo das áreas clássicas de processamento da linguagem, como as tão conhecidas áreas de Broca e Wernicke.

A percepção visual de caracteres linguísticos, portanto, requer que zonas especializadas do cérebro possam analisar e interpretar os sinais linguísticos que são dispostos ao leitor sob formas e cores diversas.

Durante a leitura, os olhos se movimentam de um modo particular. Investigar esse modo de movimentar-se é tentar encontrar sistematizações, ou seja, padrões recorrentes de movimentação ocular, que permitam inferir o modo como nosso cérebro realiza operações mentais complexas durante o processamento da informação linguística.

Nossos olhos realizam dois tipos diferentes de movimento: as sacadas e as fixações. O primeiro, as sacadas, são movimentos rápidos em que o olho parte de um ponto de fixação e passa a se fixar no ponto subsequente, sem parar em pontos intermediários. As sacadas são, portanto, pequenos saltos de um ponto a outro, cujo tempo varia entre 20 a 40ms. As fixações, também conhecidas como campo visual, são os espaços onde a informação pode

ser conscientemente processada, sempre quando o olhar aí permanecer por algo em torno de 250 ms. O tempo de permanência das fixações variará em função da tarefa cognitiva, mas está compreendido em média entre 150 a 400 ms (RAYNER, 1998). Esse é o intervalo médio de permanência para levar à consciência a informação visual e integrá-la às demais. Este tempo permite que o indivíduo chegue a construir sentidos. Segundo os estudos de Rayner (1998), uma fixação inferior a 150 ms não permite levar à consciência a informação visual, mas permite nos estudos psicolinguísticos realizar o que ficou conhecido como *priming*, que é a impressão inconsciente de um estímulo, geradora de uma resposta no indivíduo. Atualmente, trabalha-se com outra taxa, que seria a de 40 ms para o *priming* encoberto²⁰, ou preativação não consciente.

Deste modo, estudar o movimento dos olhos durante a leitura (as sacadas e as fixações) não significa estudar apenas um comportamento fisiológico; antes significa observar indicadores de como os indivíduos processam informação linguística e de eventuais dificuldades no processamento desta informação.

3.2 Rastreamento ocular

O rastreamento ocular é uma técnica que permite registrar o movimento dos olhos. Os rastreadores oculares²¹ são câmeras de vídeo que gravam os dois tipos de movimento acima descritos: as fixações e as sacadas. Nos rastreadores mais comuns, a gravação é feita com a utilização de uma câmera e de um emissor de luz infravermelha. A luz, uma vez refletida nas córneas, permite identificar suas coordenadas espaciais e precisar a localização dos rápidos movimentos sacádicos e das fixações oculares durante o rastreamento. As câmeras podem vir acopladas a telas de computador, no formato de óculos, ou como sensores de ambiente que rastreiam o olhar do indivíduo em uma sala. Em Psicolinguística, os mais usados são acoplados a telas de computador nas quais são projetados os estímulos (textos e/ou imagens) a serem percebidos em testes cognitivos.

A Psicolinguística utiliza tais técnicas para compreender como se dá a aquisição, o reconhecimento, a produção e a compreensão de textos e de frases com graus de dificuldade variáveis. Abordagens metodológicas que permitiram avanços no conhecimento dos processos cognitivos empregaram usualmente situações experimentais bem controladas, nas quais participantes são expostos a estímulos relativamente simples. Os experimentos procuram criar

²⁰ Van Rullen e Thorpe (2001) demonstraram ser possível categorizar imagens com exposição de apenas 20 ms.

²¹ *Eye-tracking* é o correspondente em inglês, cujo uso tem sido preferido em detrimento de sua tradução.

estímulos experimentais simples e altamente controlados para conseguir extrair das respostas comportamentais uma espécie de efeito puro, como quem quer entender a estrutura de cada passo do processo cognitivo. O estudo da compreensão de um texto de 2.000 palavras, por exemplo, ainda é muito pouco explorado por esta abordagem, porque implica o cruzamento de diversas variáveis que podem inviabilizar a análise dos dados, já que é difícil determinar o que está ocorrendo exatamente em cada instante de um processamento complexo, medido em uma escala espaço-temporal altamente acurada.

Pesquisadores, ao controlarem os estímulos dos seus experimentos, também podem deparar-se com um outro tipo de problema, o da validade ecológica de seus dados, que guarda certa semelhança com os problemas do debate entre uma visão holística e uma visão reducionista dos processos cognitivos. O reducionismo pode levar a um tal esquartejamento do objeto de análise que é difícil encaixá-lo na compreensão do todo, isto é, qualquer conhecimento que se alcance em um experimento isolado e circunscrito não explica ou pode até chegar a distanciar o pesquisador do seu objeto primeiro de investigação. O alerta é que as condições experimentais simples podem ser antes simplórias, além do fato de que é perigoso desconsiderar importantes variáveis que não são possíveis de serem controladas em laboratório. Há, por outro lado, o perigo inverso, em que experimentos não controlados, com inúmeras variáveis, podem não aportar praticamente conhecimento válido ou confiável, daí o pouco uso de textos complexos na investigação dos mecanismos responsáveis pelo processamento linguístico.

Em que medida, portanto, as técnicas experimentais contribuem para a compreensão do funcionamento de nossa capacidade linguística? Algumas delas são usadas há mais de quarenta anos pela Psicolinguística. Podemos citar, como já o fizemos anteriormente, as mais populares, que são o teste de leitura automonitorada, a medida dos tempos de reação, a movimentação ocular por meio das fixações oculares e dos movimentos sacádicos e os potenciais relacionados a evento que medem o tempo de reação neuronal. Há ainda os estudos de imagem cerebral, por ressonância magnética funcional e por magnetoencefalografia, bem mais dispendiosos, que somam os resultados dos tempos de reação aos estudos dos correlatos cerebrais com o objetivo de propor modelos mais precisos do ponto de vista espacial e temporal.

3.3 O estudo experimental da correferência anafórica e a movimentação ocular

Desde meados da década de 1990, o estudo da cognição humana, de um modo geral, e da linguística, em particular, passou por uma segunda revolução. Se a entrada em cena do Gerativismo, no final da década de 1950, representou uma primeira revolução cognitivista, a capacidade de investigar o processamento cerebral sem a necessidade de recorrer a métodos invasivos representa, ao menos, uma segunda fase desta revolução. Com o avanço da tecnologia, o estudo da linguagem ganhou a possibilidade de ser combinado com o estudo de outros processos perceptivos e sensorio-motores. Atualmente pesquisadores combinam diversas metodologias *on-line*. O estudo da movimentação ocular, por exemplo, é muitas vezes combinado com técnicas que medem a atividade cerebral.

O uso de métodos não invasivos tem permitido que sejam analisados em tempo real diversos processamentos linguísticos. Trata-se de um salto muito importante, que vem iluminando vários caminhos que levarão ao encontro de respostas sobre o modo como seres humanos compreendem e produzem a linguagem, considerando questões desde a especialização cerebral do hemisfério esquerdo até esclarecimentos acerca da genética da linguagem.

O estudo dessa capacidade humana da linguagem se insere dentro de uma busca bastante antiga sobre o que é especificamente humano. A linguagem é um sistema muito complexo e sabe-se hoje que não se desenvolve do mesmo modo em todos os seres humanos. Diz-se, comumente, que desenvolvemos habilidades: habilidade de falar, de compreender, de escrever e de ler. Dentre estas quatro habilidades, podemos dizer que duas dizem respeito à produção (fala e escrita) e duas outras à compreensão (audição e leitura).

O estudo da referência e da correferência é um tópico relevante na investigação sobre a arquitetura da linguagem porque permite compreender como falantes realizam a integração das informações indicadas pelo texto com seu conhecimento prévio, atribuindo coerência e significado aos enunciados.

O custo de processamento das expressões referenciais e correferenciais, que constroem a coerência discursiva, é talvez um dos problemas mais instigadores na investigação do processamento linguístico. Examinar o custo do processamento dessas expressões significa procurar indicadores das complexas operações mentais realizadas no ato de compreensão. Em resumo, para realizar esta tarefa, o leitor precisa, em primeiro lugar, compreender a palavra lida a partir do sinal gráfico impresso em algum suporte; em segundo,

armazená-la em sua memória; e, por fim, recuperar esta informação para relacioná-la com as demais no fluxo da leitura (GORDON; CAMBLIN; SWAAB, 2004). Já que o ato de ler exige esta concatenação de operações mentais, podemos dizer que a primeira e principal razão para estudar experimentalmente o fenômeno da referência, por meio de um rastreador ocular, apoia-se na necessidade da investigação de nossa capacidade cognitiva de resolver o problema da referência.

Mas o problema da referência se expressa em muitos tipos. Segundo Cavalcante (2003), as expressões referenciais podem pertencer a duas classes: as expressões que possuem como função a continuidade referencial e as que não possuem esta função. Estas últimas são comumente nomeadas introdutores referenciais, enquanto as primeiras enquadram-se no que chamamos expressões correferenciais. As expressões correferenciais ou anafóricas são, por sua vez, classificadas em muitos subtipos (KOCH, 2006; MARCUSCHI, 2004; CAVALCANTE, 2003; MONDADA; DUBOIS, 1995; APOTHÉLOZ, 2001; APOTHÉLOZ; CHANET, 1997; entre outros).

Pesquisar sobre nossa capacidade de resolução anafórica, portanto, significa tentar responder aos problemas ligados à nossa capacidade de retenção da informação recentemente lida em nossa memória e à nossa capacidade de integrar esta informação às demais que vão sendo ativadas durante a leitura de um texto. É possível verificar em tempo real, por meio dos tempos das fixações oculares, quais os termos que se referem a outros e em que pontos, da leitura ou da audição, os participantes se detêm mais tempo para integrar a informação e armazená-la na memória de longo prazo.

O estudo experimental do problema da correferência anafórica vem esclarecer uma série de questões teóricas sobre custo de processamento e dificuldade na integração das informações. A precisão dos resultados fornecidos pelo rastreador ocular permite, por meio do controle do experimento, identificar em que momento os leitores estão resolvendo as anáforas do texto e diferenciar este momento de outros fatores linguísticos envolvidos nesta resolução, como pode ser o *priming* léxico-semântico. Garrod e Sanford (1977) demonstraram que na resolução anafórica não havia influência do *priming* lexical por meio do controle entre frases, onde estava presente aquilo que chamaram de tipicidade do antecedente, isto é, a presença de termos pertencentes ao mesmo campo semântico da anáfora correferencial.

Entretanto, foi demonstrado por Garnham (2001) que a tipicidade do antecedente é um fator que influencia a facilidade com a qual é lida uma anáfora. O autor mostrou que, quando a anáfora correferencial é um item lexical formado por um hipônimo ou hiperônimo

típico em relação ao seu antecedente, a leitura se torna mais fluida. Caso contrário, há um acréscimo no tempo da leitura da anáfora correferencial. Parafraseando o estudo de Garnham, observemos o exemplo.

(19) Um pardal costumava entrar na casa. O pássaro era atraído pela despensa.

(20) Um ganso costumava entrar na casa. O pássaro era atraído pela despensa.

Em (19), o antecedente “o pardal” é um tipo bastante típico de pássaro; já em (20), “um ganso” é um pássaro menos típico (BATTIG; MONTAGUE, 1969; ROSCH, 1975). A tipicidade do antecedente altera o tempo de processamento da expressão correferencial “o pássaro”, neste caso um hiperônimo previsível. Constatar a previsibilidade de um hiperônimo pode também oferecer pistas sobre o modo como construímos categorias de palavras, além de procurar entender como chegamos a recategorizar referentes em hiperônimos não previsíveis, por exemplo. É possível modelar os processos de análise e interpretação das sentenças, ou seja, o cálculo das fixações nos diversos níveis de análise da frase permite investigar passo a passo as pistas que nos auxiliam na construção do sentido que atribuímos a um texto.

O estudo da referência, entendida como uma atividade discursiva em que os sujeitos da interação realizam operações de ativação, reativação ou desativação de objetos do discurso (KOCH, 2006), não necessita estar circunscrito ao estudo da leitura de textos. Fato é que, para o estudo da compreensão da fala, é usado o paradigma do mundo visual para compreender como indivíduos processam os referentes textuais, sejam introdutores, sejam operadores de progressão textual (anáforas). Este protocolo experimental foi pensando a partir da clássica constatação dos estudos psicológicos de Cooper (1974), que percebeu que as pessoas costumam olhar para os referentes do mundo (objetos físicos) em seu entorno ao ouvirem palavras que se referem a objetos com representação física na realidade (como objetos concretos, tais como lápis, borracha, caderno). O paradigma do mundo visual²², desenvolvido primeiramente por Tanenhaus, Spivey-Knowlton, Eberhard e Sedivy (1995), consiste em medir os movimentos oculares dos participantes de uma pesquisa enquanto estão ouvindo instruções para manipular objetos em tempo real ou para tomar decisões sobre imagens (ou simplesmente observá-las) projetadas em uma tela. Seus resultados apontam para o fato de que o tempo de processamento implicado na ação de olhar para os objetos pode servir como indício do processamento da informação fonológica, além de sugerirem o modo

²² Tradução de “visual world paradigm”.

como os participantes constroem as estruturas sintáticas, interpretam a função de agente e objetos típicos de verbos, resolvem o problema do pronome no contexto, entre outros resultados (ALLOPENNA; MAGNUSON; TANENHAUS, 1998; TANENHAUS; SPIVEY-KNOWLTON; EBERHARD; SEDIVY, 1995; ARNOLD; EISENBAND; BROWN-SCHMIDT; TRUESWELL, 2000).

É também possível medir, tanto por meio do paradigma do mundo visual quanto por experimentos de leituras de frases, qual o efeito do conhecimento de mundo no processamento da compreensão da linguagem. Como já dito acima, na ação de recuperar informação prévia, os leitores variam em relação ao tempo de fixação na leitura de uma palavra.

Os rastreadores oculares atualmente utilizados pelos pesquisadores possuem uma resolução temporal impressionante e permitem, por meio dos *softwares* de análise, o fracionamento dos tempos de leitura em diversos componentes (variáveis dependentes). O pesquisador pode observar os dados das fixações iniciais, das refixações em uma palavra, das regressões a palavras precedentes, da releitura de uma palavra após regressão, da distância entre sacadas etc. Todos estes dados são potenciais indicadores dos processos cognitivos que ocorrem em cada instante de tempo, já que a gravação possui um ótimo nível de detalhamento, cujas margens de erro são muito pequenas, praticamente desprezíveis, da ordem de 0.5 ms nos mais modernos rastreadores.

Por fim, é importante salientar que a metodologia da movimentação ocular é mais livre de enviesamento, permitindo os participantes agir com naturalidade nos testes, um fato importante na aquisição e na interpretação dos dados sobre o processamento cognitivo da linguagem.

A técnica é, portanto, muito eficaz para a investigação das relações semânticas, sintáticas e discursivas estabelecidas entre os constituintes frasais e entre estruturas oracionais.

Os estudos sobre a trajetória ocular em tarefas cognitivas relativas à linguagem têm oferecido uma importante contribuição para a pesquisa acerca dos tempos de reação, medidas que indicam de forma indireta como o cérebro processa a informação recebida pelo olhar. Os resultados das pesquisas experimentais estão trazendo novos esclarecimentos acerca dos custos de processamento das estratégias cognitivas de compreensão da linguagem. O uso dessas técnicas experimentais estão, portanto, no cruzamento de teorias psicolinguísticas de processamento da leitura e de teorias neurofisiológicas sobre a movimentação ocular.

3.4 Hipóteses

Com base nas teorias e hipóteses apresentadas para explicar a resolução anafórica no âmbito da gramática e do discurso e com base no conhecimento sobre a metodologia experimental on-line de rastreamento ocular, propusemo-nos investigar nesta tese as hipóteses descritas na seção seguinte.

3.4.1 Hipótese básica

Em períodos complexos, o sintagma nominal que ocupa a função de sujeito da oração principal é o antecedente preferencialmente retomado para estabelecer correferência com a primeira expressão referencial anafórica do período.

3.4.2 Hipóteses secundárias

1. Em períodos complexos por coordenação, as expressões anafóricas preferem estabelecer correferência com os antecedentes na posição de sujeito da primeira oração coordenada.
2. Em períodos complexos por coordenação, não há custos adicionais de processamento quando o objeto direto da primeira oração torna-se o centro catafórico potencial a ser retomado por uma expressão anafórica com conteúdo semântico na segunda oração.
3. Em períodos complexos por coordenação, há custos adicionais para o estabelecimento de relação semântica de hiponímia na correferência entre antecedente e expressão referencial anafórica.
4. Em períodos complexos por coordenação, o paralelismo estrutural das orações é um fator relativamente mais fraco do que a posição sintática de sujeito para o estabelecimento da correferência.
5. Em períodos complexos por subordinação, a categoria vazia *pro* prefere o estabelecimento de correferência com o sintagma nominal em posição de sujeito da oração principal.
6. Em períodos complexos por subordinação, o paralelismo estrutural das orações é um fator relativamente mais fraco do que a posição sintática de sujeito para o estabelecimento da correferência.

7. Em períodos complexos por subordinação, há custos adicionais de processamento para a correferência entre pronome pleno e objeto da oração principal.
8. A categoria vazia *pro* e o pronome pleno do Português Brasileiro estão em distribuição complementar, desempenhando funções anafóricas especializadas.

3.5 Procedimentos experimentais para coleta dos dados

Participantes

Cada estudo examina grupos de pessoas com no mínimo 20 participantes. Os grupos foram compostos por adultos, falantes nativos da língua portuguesa, leitores proficientes com visão normal ou corrigida, cujo grau de escolaridade fosse de, no mínimo, 12 anos ou ensino médio concluído. Os participantes foram selecionados para constituir um grupo em certa medida homogêneo quanto ao conhecimento de mundo compartilhado. Sempre que houve dúvida em relação a dados coletados de algum participante, por questões técnicas e outras, optamos pela exclusão dos dados do participante em questão. Os estudos reportam apenas o número válido de participantes.

Protocolo experimental

Cada experimento segue um rigoroso protocolo experimental. De maneira geral, cada participante é convidado a colaborar voluntariamente de um experimento que consiste na leitura silenciosa de um conjunto de sentenças ou textos experimentais, dos quais pelo menos um terço é composto de distratores. Estas frases e textos são projetadas em um tela de computador equipada com uma câmera infra-vermelha integrada, que registra o movimento ocular do participante. Para recrutar os participantes, foi feito previamente um teste de sua proficiência de leitura, que consiste em medir a velocidade de leitura em função da compreensão. Trata-se de um pré-teste aplicado a todos os participantes antes de se submeterem aos experimentos desenhados para esta tese. Uma vez tomadas as medidas dos participantes, o experimento propriamente dito é iniciado. Os participantes passam por uma fase de treinamento, que tem como objetivo dar instruções sobre o experimento e familiarizar os participantes com o protocolo experimental. Entretanto, a fase pré-teste não informa aos participantes o objetivo da pesquisa, porque é necessário que estes sejam ingênuos para não enviesar os dados. Nos protocolos desenhados para esta tese, após a leitura das frases ou textos, faz-se uma pergunta do tipo múltipla escolha, que é projetada na tela, com o objetivo

de testar a compreensão do texto. O participante é convidado a responder, selecionando com o mouse uma das duas respostas sugeridas.

Na construção das frases e textos que foram usados nos experimentos desta tese, com exceção do experimento com o objetivo exploratório, foram controlados o tamanho das palavras, a distância entre as anáforas e os antecedentes (potenciais), o gênero das anáforas e dos termos antecedentes. As variáveis de movimentação ocular que foram estudadas são: (i) o número de fixações nas regiões de interesse, (ii) o tempo médio de duração de uma fixação, (iii) o tempo total de fixação em uma região de interesse, (iv) o tempo da primeira fixação e (v) o tempo de exposição ao estímulo, chamado aqui também de tempo de reação. Uma vez feitas as medidas, os resultados são calculados para se elaborar médias entre os participantes e/ou entre os itens experimentais.

Procedimento de coleta de dados

Os procedimentos são descritos em cada um dos capítulos que tratam dos estudos experimentais. As participações são todas individuais e sempre na presença do experimentador, no caso a autora da tese, que monitora o fluxo da leitura e a distância do participante através de um outro monitor.

Para recolher dados comportamentais de seres humanos, esta pesquisa foi submetida e aprovada pelo Conselho de Ética em Pesquisa da Universidade Federal do Ceará (ver ANEXOS) e solicitou de cada um dos participantes a assinatura de um termo de consentimento. Todos os documentos referentes ao consentimento, assim como os dados experimentais, estão guardados, arquivados e disponíveis para consulta.

Antes de começar o experimento, cada um dos participantes é submetido a um *baseline* individual. No rastreador o *baseline* é uma calibração, um procedimento obrigatório na pesquisa com rastreadores oculares e tem como objetivo realizar um cálculo da distância do participante da tela do computador e da distância que separa os olhos de cada um dos indivíduos, porque o resultado fornecido pelo programa calcula além da fixação de cada um dos olhos, a média da distância entre as duas córneas de cada participante e este ponto de fixação projetado na tela.

4 ESTUDO DA LEITURA DE TEXTOS NÃO MANIPULADOS

Neste capítulo, reportamos o estudo da movimentação ocular durante a leitura de textos não manipulados em língua portuguesa. Tomando por base os estudos sobre a movimentação ocular durante a leitura em língua inglesa (JUST; CARPENTER, 1980; RAYNER, 1998), conduzimos um estudo com o objetivo de investigar um padrão de tempo de fixação durante o fluxo de leitura em língua portuguesa. Além disso, procuramos determinar um tamanho típico de palavra dos textos estudados, para que os resultados da duração média de fixação possam estar estreitamente relacionados com a característica física dos textos lidos, característica examinada aqui em número de letras e tamanho médio das palavras.

4.1 Introdução

Para compreender um texto, não é suficiente que o leitor somente reconheça as palavras individualmente e analise a estrutura gramatical de cada sentença. É preciso que o leitor mantenha em sua memória as representações de cada uma das entidades e dos eventos mencionados e relacione-os às informações ou representações que se somam no fluxo da leitura. Como nos dizem Staub e Rayner (2007), implicada neste processo de compreensão do texto, também está a capacidade de determinar quais são os antecedentes das anáforas pronominais ou dos sintagmas nominais com descrições definidas, e quais são as inferências que devem ser feitas sobre as relações entre eventos e entidades, incluindo as relações cronológicas, causais, explicativas, etc, presentes em um texto.

Poucos estudos, entretanto, têm examinado como a movimentação ocular se relaciona com o processamento do discurso. Esta área pode ser bastante explorada e alguns estudos, como o da Teoria da Centralização (ver Capítulo 2), fornecem pistas de como realizá-los para investigar custos de processamento. No entanto, é a construção de uma metodologia experimental coerente que fará o diferencial para conseguir extrair das movimentações oculares informação sobre o processamento do discurso.

No discurso há muitas variáveis interrelacionadas. Isolá-las para realizar um estudo experimental é um desafio a ser solucionado, como nos lembram Mondada e Dubois (1995), ao defenderem a tese de que a anáfora, no lugar de ser definida como simples representação de referentes extramentais, pode ser compreendida por meio da noção de

referenciação, na qual está contida a idéia de que os indivíduos, não apenas retomam antecedente previamente mencionados, mas ancoram suas práticas discursivas muito mais em suas atividades sociais do que em um mundo previamente concebido e discretizado em objetos e entidades fora-de-contexto.

Diante de visões controversas em relação ao que é o discurso ou sobre quais estruturas cognitivas são necessárias durante a interação comunicativa, constata-se que está aberto um vasto campo de pesquisa, que, acreditamos, deva ser explorado com os métodos disponíveis e com aqueles que possam ser especialmente criados para investigar este fenômeno.

A compreensão leitora, seja no âmbito do estudo da frase, do período ou do texto, vem sendo, há pelo menos quarenta anos, investigada regularmente com o uso de rastreadores oculares. As respostas comportamentais do movimento ocular tem sido interpretadas como pistas fundamentais na construção de inferências sobre o processamento mental da linguagem (JUST; CARPENTER, 1980; RAYNER, 1998).

Entre as primeiras descobertas acerca da leitura, estão os resultados de Rayner (1998), que, dentre outros achados, demonstrou não haver simetria para a percepção das palavras entre leitores de línguas cuja escrita vai da direita para a esquerda (caso do português e inglês) e da esquerda para a direita (caso do hebraico e árabe). Da evidência, o autor concluiu que nossa movimentação ocular é reflexo dos nossos processos cognitivos, construídos por padrões biológicos e culturais.

Constatou também que o tempo de fixação varia em função da dificuldade de processamento, ou seja, que as durações das fixações apresentam variações consideráveis entre sujeitos e intrassujeitos, dependendo da complexidade da tarefa realizada. Rayner (1998) faz uma revisão bibliográfica da pesquisa realizada, nas décadas de 1980 e 1990, sobre o custo de processamento informacional para várias tarefas, cujos tempos médios estão sintetizados na Tabela 2 reproduzida abaixo.

Tabela 2 – Distribuição dos tempos médios de duração da fixação de Rayner (1998)

Tarefa	Duração da fixação (ms)	Tamanho da sacada (graus)
Leitura silenciosa	225	2 (8 caracteres)
Leitura em voz alta	275	1.5 (6 caracteres)
Busca visual	275	3
Percepção de imagens	330	4
Leitura de música	375	1
Datilografia	400	1 (4 caracteres)

O comportamento ocular durante a leitura foi exemplarmente descrito por Just e Carpenter (1980), em um estudo seminal que avaliou as variações intra e intersujeitos. Suas conclusões apontaram dois importantes princípios, que, eventualmente, se tornaram clássicos. O primeiro, denominado “*immediacy assumption*”, defende que uma palavra é interpretada assim que é fixada, ou seja, no fluxo da leitura, o leitor interpreta uma palavra após a outra, mesmo correndo o risco de realizar hipóteses erradas. O segundo, denominado “*eye-mind assumption*”, revela que a permanência do olhar em uma palavra é a medida do tempo de seu processamento ou compreensão. As conclusões deste estudo sustentam que são os itens lexicais (ou palavras de conteúdo) aqueles majoritariamente fixados durante a leitura, ou seja, que as palavras funcionais, apesar de serem igualmente processadas, são fixadas em apenas 35% das vezes. Para a investigação do processamento da linguagem, este estudo trouxe ainda muitas outras contribuições.

Uma delas foi a de ter estabelecido que a frequência de uma palavra é um fator alterador do tempo de fixação ocular. Por exemplo, se um leitor de PB se depara com termos muito frequentes, o tempo de fixação nestas palavras diminuirá significativamente. Este efeito se dá por cascata, já que as palavras subsequentes às mais frequentes também tendem a ser por menos tempo fixadas. Ainda, se o contexto em que a palavra surge é bastante conhecido do leitor, é provável que este faça uma previsão ou antecipação sobre a palavra que virá expressa no texto, diminuindo assim o tempo de fixação nesta palavra. Vejamos um exemplo. Suponhamos que um leitor esteja diante de uma frase como: “A capital da França é Paris”. Seria praticamente certo que será quase nulo o tempo em que seus olhos se fixarão no termo “Paris”. Este efeito é conhecido por *preditividade*²³.

O contrário também se provou verdadeiro, já que, quando leitores estão diante de assuntos ou termos não usuais e inesperados, estes costumam aumentar o tempo de fixação nestas palavras e nas subsequentes, além de realizarem também uma maior quantidade de movimentos sacádicos regressivos. A média de movimentos regressivos é de 15% do total de movimentos progressivos na leitura (CARREIRAS; CLIFTON, 2004). Por movimento regressivo, entende-se o retorno para um ponto do texto anteriormente fixado.

A extensão de uma palavra também altera o tempo de fixação e variará em função do idioma, uma vez que as línguas humanas possuem tamanhos de palavras característicos. Como foi dito acima, a fixação compreende uma região muito pequena, que significa para a leitura do texto escrito algo como o espaço em torno de 4 a 8 caracteres. A amplitude de um

²³ Ver Capítulo 3 para a definição de processo do tipo *top-down*.

movimento sacádico gira em torno de 7 a 9 caracteres em média. Portanto, tanto a fixação quanto a amplitude das sacadas na leitura devem estar correlacionadas com a extensão da palavra e com os conhecimentos prévios do leitor, ou seja, com o seu conhecimento de mundo, linguístico e do contexto.

Esta influência marca um caráter subjetivo que pode se refletir na motivação do leitor. A motivação pessoal pela leitura é uma variável que precisa ser controlada para não haver equívoco na análise dos dados. Em função dos objetivos e da familiaridade com o tema dos textos, os tempos de leitura e processamento de um indivíduo podem variar substancialmente.

Finalmente, o estudo de Just e Carpenter (1980) ainda contribuiu para decompor a leitura em quatro estágios de processamento, que seriam os seguintes: em primeiro lugar, o leitor codifica a palavra, ou seja, processa visualmente o *input*, transformando-o em uma forma linguística; em segundo lugar, passa ao acesso lexical, que é o pareamento entre a forma visual e a linguística; em seguida, o leitor determina qual é o papel semântico deste item lexical; para, finalmente, integrá-lo à informação da sentença na qual está presente. A compreensão do texto como um todo é atribuída à capacidade de o indivíduo relacionar e integrar todas as sentenças do texto.

Para entender o que significa o tempo de fixação ou a duração do olhar em uma expressão é preciso saber que este tempo é relativo à soma de todas as fixações consecutivas sobre uma palavra antes de os olhos do leitor deixarem esta palavra. A fixação ocular é assumida como o reflexo do tempo de processamento de uma palavra em particular. Aqui cabe ainda lembrar que o termo fixação é uma aproximação, porque o olhar humano nunca é estacionário. Há sempre uma pequena variação, diríamos ínfima, que faz com que essa precisão metodológica seja necessária. Por isso, um dos resultados importantes no estudo com rastreadores oculares é a duração total do olhar em uma palavra, obtida por meio do somatório de todas as microfixações no espaço de 2° (dois graus) delimitado pela fóvea (ver seção 3.2).

Com o objetivo de examinar como seria feita a leitura de textos em língua portuguesa, conduzimos um estudo de leitura de textos não manipulados, retirados de uma coluna semanal do jornal Folha de São Paulo. A coluna pertence ao escritor e físico Marcelo Gleiser. Procuramos avaliar o tempo médio da duração da fixação ocular durante a leitura silenciosa. Também fizemos o estudo do tamanho característico das palavras destes textos, para correlacionar tamanho da palavra, tempo de fixação e posteriormente amplitude das sacadas.

4.2 Materiais e Método

4.2.1 Participantes

Participaram deste estudo voluntários que eventualmente também passaram por outros testes experimentais reportados nos Capítulos 5 e 6. Escolhemos os dados de 20 participantes para analisar a média das fixações oculares. A Tabela 3 abaixo resume os dados demográficos do grupo examinado. O grupo foi majoritariamente formado por alunos de graduação e pós-graduação dos Departamentos de Física e de Letras Vernáculas da Universidade Federal do Ceará.

Tabela 3 – Dados dos participantes do estudo da movimentação ocular em textos não manipulados.

Dados demográficos

N	20
Sexo	12 M/ 8 F
Lateralidade	19 D/ 1E
Idade	27,3 (6,2)
Escolaridade	17 (1, 2)

Desvio padrão entre parênteses

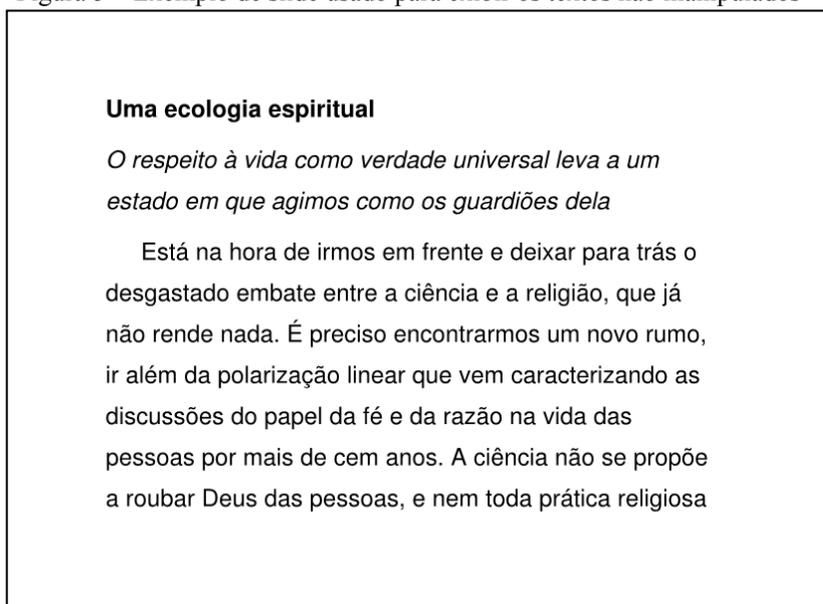
4.2.1 Procedimento experimental e materiais

Os participantes foram convidados a ler textos autênticos na tela do rastreador ocular Tobii T120Hz (para detalhes sobre o equipamento ver seção 5.2.4). Todos consentiram por escrito que seus movimentos oculares poderiam ser gravados durante a leitura. Antes de iniciar o protocolo experimental, a calibração, procedimento obrigatório que configura o equipamento para obter as medidas correspondentes àquele indivíduo, foi realizada seguindo as recomendações do fabricante do rastreador ocular. Uma vez concluída a fase de calibração, os participantes passaram por uma fase de treinamento em que leram um texto curto extraído da revista *Veja* (ver APÊNDICE A, Texto 4). Em seguida, começaram a ler os textos não manipulados que iam sendo exibidos em slides, com uma média de 10 linhas de texto. Para passar para a próxima página, os participantes precisavam dar um clique no mouse.

Os participantes leram três textos de divulgação científica (ver APÊNDICE A, Textos 1, 2 e 3), originalmente publicados na coluna de Marcelo Gleiser, na Folha de São Paulo. Apesar de se tratar de uma publicação em jornal de ampla circulação nacional, o texto

é de certa forma temático, pois o autor se centra em temas próprios da ciência, e muito frequentemente circunscreve suas crônicas a temas próprios da Física, sua área de formação principal. Os textos foram escolhidos porque poderiam apresentar um certo grau de complexidade sem se tornarem técnicos, complexidade esta que manteve os participantes atentos durante a leitura. Os slides, que foram usados para a exibição dos textos completos, podem ser visualizados na Figura 5, abaixo.

Figura 5 – Exemplo de slide usado para exibir os textos não manipulados



Como o rastreador ocular registra o movimento dos olhos dos participantes a uma distância de 60 cm, fizemos o cálculo do número de caracteres que seriam vistos pelo campo visual a esta distância. Um grau (1°) do campo visual corresponde a uma média de 2.29 caracteres em uma fonte Arial de 28 pontos. Foi este o tamanho e a fonte escolhida para exibição dos textos nos slides.

Todos os participantes receberam a mesma instrução que os informava que, ao final da leitura, seriam feitas perguntas para conferir a compreensão do texto. As perguntas, na verdade, eram pequenas questões elaboradas pelo examinador, que iniciava uma conversa com o participante sobre o tema da coluna. A conversa tinha a intenção de detectar falhas na compreensão global do texto. Como isto não foi identificado, todos os dados coletados, de todos os participantes, foram usados para a análise.

4.3 Medidas examinadas

4.3.1 Duração média da fixação ocular

A duração média da fixação ocular corresponde ao tempo em que o olho permanece em uma área da tela. Usamos o filtro “Tobii Fixation Filter”, presente no software do fabricante do rastreador ocular. Este filtro elimina ruídos e considera a fixação se compreendida dentro de um diâmetro de 30 pixels, desde que aí permaneça por pelo menos 100 milissegundos (ms). Isto quer dizer que o limiar mínimo de uma fixação deste filtro é de 100ms, mas que não há máximo estabelecido.

4.3.2 Característica dos textos apresentados

Os textos apresentados foram descritos em relação às suas características de número de palavras e de tamanho médio de palavras. Esta medida é importante para avaliarmos os tamanhos das sacadas e os tempos de fixação dos participantes no texto. Como visto na introdução (seção 4.1), a amplitude da sacada pode compreender entre 7 a 9 espaços de letras.

4.4 Resultados

Os resultados que obtivemos para as fixações oculares estão sumarizados na Tabela 4. A coluna da esquerda é a codificação que recebeu cada participante. A coluna da direita mostra as médias de fixação obtidas para cada participante. Entre parênteses reportamos o desvio padrão das médias relativas a cada participante. A Tabela 4 permite observar a variação inter sujeitos dos tempos médios de fixação de cada participante. A média geral obtida foi de 212 ms com desvio padrão de 26 ms. Se retomarmos o tempo obtido por Rayner (1998) para a leitura silenciosa, de 225 ms, constatamos que os tempos de leitura que encontramos para a língua portuguesa não se distanciam dos tempos da leitura em inglês. Observamos, portanto, que os resultados são similares. Note-se que o nível de dificuldade imposto pelo texto era um nível médio, correspondente aos demais textos publicados em jornal de grande circulação nacional. Se não o fosse, nossos tempos médios de fixação provavelmente seriam maiores.

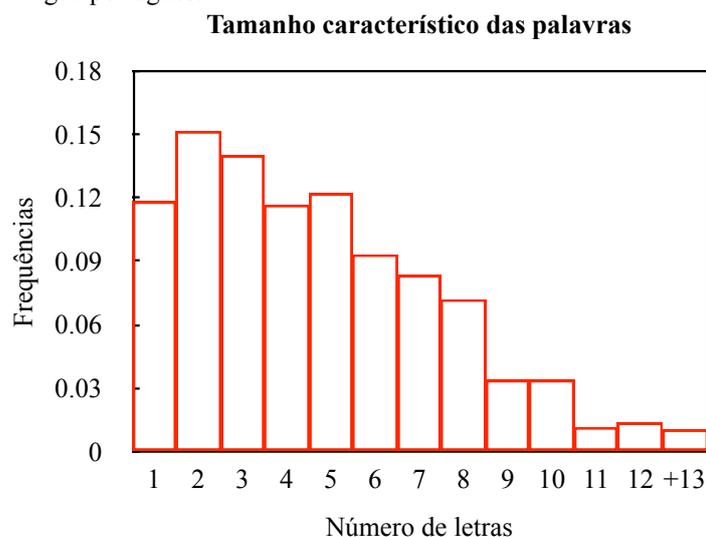
Tabela 4 – Tempo médio das fixações encontradas na leitura dos textos não manipulados.

Tempo médio da fixação em segundos	
P01	0.192 (± 85)
P02	0.188 (± 115)
P03	0.192 (± 86)
P04	0.212 (± 95)
P05	0.198 (± 163)
P06	0.176 (± 78)
P07	0.227 (± 97)
P08	0.195 (± 112)
P09	0.200 (± 94)
P10	0.245 (± 137)
P11	0.201 (± 88)
P12	0.211 (± 82)
P13	0.209 (± 116)
P14	0.278 (± 93)
P15	0.185 (± 138)
P16	0.227 (± 79)
P17	0.214 (± 82)
P18	0.265 (± 78)
P19	0.228 (± 86)
P20	0.197 (± 121)
Média	0.212 (± 26)

Desvio padrão entre parênteses.

O estudo do tamanho médio das palavras da língua portuguesa presentes nos textos lidos, extraídos da Folha de São Paulo, pode ser visualizado no Gráfico 1, abaixo. Por palavra, este estudo entendeu qualquer forma da língua que esteja separada por espaços no texto. Portanto, este gráfico apresenta a distribuição das formas presas e livres dos textos de Marcelo Gleiser. Da esquerda para a direita, vemos a frequência das palavras com uma letra (±12%), duas letras (15%), três letras (14%), e assim sucessivamente, até chegar às palavras com 13 ou mais letras, que foram condensadas em uma só barra (1%).

Gráfico 1 - Distribuição do tamanho característico da palavra em língua portuguesa



Já a Tabela 5, abaixo, resume todos os dados característicos dos textos, mas apresenta mais duas colunas que caracterizam as palavras com 4 ou mais letras e as palavras com 3 ou menos letras. Isto foi feito porque queríamos calcular o tamanho médio da palavra em língua portuguesa separando-se as palavras curtas das longas, considerando para isso as definições de Quaresma e Pinho (2007). Quaresma e Pinho (2007) encontraram que o tamanho médio característico da palavra em língua portuguesa é 4.64. Nós encontramos um resultado praticamente idêntico, com média de 4.70. No entanto, calculamos a mediana para estabelecer um parâmetro com número inteiro para caracterizar o número médio de letras nas palavras dos textos lidos neste estudo. A mediana de todas as palavras é 4. Se considerarmos palavras, como itens lexicais livres e potenciais núcleos de sintagmas nominais, com mais de 3 letras, o que elimina a grande maioria dos artigos e preposições da língua, obtemos como tamanho mediana o número 6.

Tabela 5 – Características das palavras nos textos estudados

Tamanho característico das palavras			
	Todos os tamanhos	Maior ou igual a 4 letras	Menor ou igual a 3 letras
Percentual	100%	59%	41%
Número de palavras	1835	1083	752
Média	4.70 (± 2.83)	6.54 (± 2.22)	2.05 (± 0.79)
Mediana	4	6	2

4.5 Conclusões

O estudo das características das palavras dos textos autênticos lidos no rastreador ocular e o estudo das fixações médias nos leva a conclusão de que as palavras livres em língua portuguesa possuem em média 6 letras e que o tempo de fixação ocular durante a leitura de textos é de 212 ms. Comparando nossos resultados com os resultados de Rayner (1998), obtivemos:

Tabela 6 – Padrão de duração de fixação e tamanho médio da palavra em português.

Duração da Fixação em milissegundos		
	Inglês	Português
Leitura silenciosa	225	212
Tamanho médio da palavra ²⁴	5	6

²⁴ Para o tamanho da palavra em inglês consultamos o site <http://www.wolframalpha.com/>

5 ESTUDO DE PREFERÊNCIAS SEMÂNTICAS

Neste capítulo, descrevemos o experimento que observa as preferências de correferência ao manipular a relação de hierarquia semântica entre os SNs potencialmente co-indexados e a saliência sintática do antecedente anafórico. Como os estímulos foram divididos em dois grupos, dizemos que o Grupo 1 foi testado com os estímulos do experimento 1 e o Grupo 2, do experimento 2.

5.1 Estudo da hierarquia semântica entre anáfora e antecedente: introdução

Apesar dos esforços empreendidos para compreender como se dá a resolução anafórica, tanto no campo da pesquisa linguística quanto no âmbito da pesquisa sobre computação da linguagem natural, o tema ainda suscita a necessidade de esforços continuados para ser mais bem entendido. O conhecimento dos custos implicados na resolução anafórica pode contribuir para a explicação de como funciona o processador linguístico e de como poderia ser implementado um algoritmo computacional que obtivesse taxas de êxito, suficientemente satisfatórias, para identificar, corretamente, termos co-indexados em frases e textos. Na comunicação humana, a anáfora²⁵ é uma regra que serve a pelo menos dois grandes propósitos: referir-se a uma informação dada no discurso e contribuir para a construção da coerência e coesão discursivas.

A resolução da correferência anafórica, além de obedecer às restrições sintáticas obrigatórias (CHOMSKY, 1981; REINHART, 1983), sofre influência de fatores relativos à semântica dos termos anafóricos e de seus antecedentes, sobretudo quando se fala de sintagmas nominais anafóricos. Fatores tais como a tipicidade do termo antecedente em relação ao anafórico (GARROD E SANFORD, 1977; ALMOR, 1999; VAN GOMPEL, LIVERSEDGE E PEARSON, 2004) têm sido manipulados em experimentos que procuram contribuir para a compreensão das computações sintáticas que interferem nos custos e, conseqüentemente, na velocidade de resolução anafórica.

Numa revisão da literatura sobre o assunto, é possível distinguir pelo menos três teorias que têm procurado explicar os efeitos da tipicidade do antecedente. A primeira delas foi proposta por Garrod e Sanford (1977) e ficou conhecida como a Teoria da Sobreposição Semântica, cuja principal conclusão é a de que o efeito de tipicidade ocorre porque é mais fácil identificar um antecedente quando este possui um alto grau de sobreposição semântica,

²⁵ Neste trabalho, o termo anáfora é empregado em seu sentido lato.

isto é, quando possui traços semânticos e/ou conceituais comuns em relação ao termo anafórico. Os autores não observaram um efeito facilitador para a resolução quando a sobreposição semântica entre anáfora e antecedente era pequena. Eles examinaram os tempos de reação em experimento de leitura automonitorada de duas sentenças que estabelecem entre si uma correferência intersentencial.

A Teoria da Sobreposição Semântica foi bastante questionada porque havia uma sugestão de que o efeito pudesse antes resultar de um *priming* semântico-lexical. O estudo de Garrod e Sanford (1977) foi replicado usando-se outras medidas, como o rastreamento ocular. Apesar das críticas, os resultados destes trabalhos permaneceram consistentes com as predições da Teoria da Sobreposição Semântica (RAYNER; KAMBE; DUFFY, 2000; MYERS; O'BRIEN, 1998; O'BRIEN *et al.*, 1995; O'BRIEN; ALBRECHT, 1992; O'BRIEN *et al.*, 1997).

A segunda teoria que gostaríamos de citar (ver Capítulo 2), ficou conhecida como Hipótese da Carga Informativa (ALMOR, 1999). Segundo o autor, toda anáfora possui uma carga ou um custo determinado por sua informatividade em relação ao seu termo antecedente. Para determinar esta carga, Almor (1999) propõe a idéia de um espaço semântico-conceitual em que a informatividade da anáfora possa ser avaliada em relação ao seu antecedente. Almor (1999) realizou um experimento de leitura automonitorada, em que manipulou a funcionalidade anafórica, colocando o antecedente em foco, clivado, e fora de foco, em sentenças não clivadas, além de manipular também a relação conceitual entre anáfora e antecedente. Seus resultados sugerem que o tempo de leitura dos sintagmas nominais anafóricos são um reflexo da sua função no texto, semelhante ao que propõe Ariel (1991). As anáforas com menor custo de processamento possuíam uma maior justificativa funcional dentro da frase lida. Seus resultados vão, portanto, na direção oposta das teorias que compreendem os sintagmas nominais anafóricos como uma classe homogênea. Segundo os achados do autor, os sintagmas nominais anafóricos precisam ser estudados em função do seu papel discursivo e do seu custo de processamento.

A terceira teoria a que vamos nos referir é oriunda do trabalho comparativo realizado por Van Gompel, Liversedge e Pearson (2004). Os autores elaboraram um experimento que reavaliasse as contribuições de Almor (1999) e os estudos prévios feitos por Garrod e Sanford (1977) porque os resultados de ambos entravam em conflito entre si. Apoiados pelo fato de que havia diferenças nas variáveis dos estudos em questão, realizaram uma investigação, fazendo uso de rastreadores oculares, com o objetivo de rever tais

resultados. Basicamente, o que fizeram foi observar os efeitos da tipicidade do antecedente em grupos de frases cujo antecedente estava em foco e fora de foco. O diferencial de Van Gompel, Liversedge e Pearson (2004) foi propor uma análise do custo de processamento em várias regiões da frase. Estudaram quatro regiões: (i) a região que precede a anáfora; (ii) a região onde está a anáfora; (iii) a região imediatamente posterior à anáfora (*spillover*) e (iv) a região final da sentença. Suas conclusões apontaram para a possível presença de dois tipos diferentes de processamento para a resolução dos sintagmas nominais anafóricos durante a leitura. Haveria um primeiro processamento, que os autores denominaram de estágio inicial, na região onde está a anáfora, com custos consistentes com a Teoria da Sobreposição Semântica. Em um estágio final, mensurado na região final da sentença, haveria um segundo processamento, com custos mais alinhados aos encontrados por Almor (1999), quando propôs a Hipótese da Carga Informacional. Assim, os autores desta terceira hipótese ou teoria parecem querer explicar os resultados discrepantes, buscando integrar os estudos anteriores em um única teoria, que aponta para a presença de dois estágios na resolução anafórica.

Apenas para complementar vamos nos lembrar dos interessantes achados de Gordon, Grosz e Gilliom (1993) que indicam haver maior custo de processamento quando o correferente é idêntico ao seu antecedente, no efeito que ficou conhecido na literatura como *Penalidade do Nome Repetido* (ver Capítulo 2). Este seria mais um fator que causa influência na resolução do correferente anafórico na forma de sintagma nominal. Esse achado está intrinsecamente relacionado com a manipulação da tipicidade do antecedente e com a funcionalidade da anáfora no discurso. O trabalho de Gordon, Grosz e Gilliom (1993) sugere que a penalidade do nome repetido acontece porque também violaria a máxima de quantidade de Grice, superespecificando uma informação já dada e ativa na memória de curto prazo do interlocutor. A sobreposição semântica seria, neste caso, total. Almor (1999), no entanto, demonstrou que tal acréscimo no custo de processamento seria observado somente se o antecedente estivesse em foco. Em contextos em que não está em foco, a repetição do sintagma nominal pode apresentar uma funcionalidade, como a de introduzir um novo tópico discursivo.

Diante do exposto acima, podemos observar que a resolução anafórica de sintagmas nominais é suscetível à variação em função de diversos fatores. Para efeito didático e de investigação, é possível dividir em dois grupos os fatores influentes na resolução da correferência. No primeiro grupo, encontram-se as restrições obrigatórias (CHOMSKY, 1981;

REINHART, 1983) e no segundo, podem-se agrupar todas as restrições brandas ou facultativas, as preferências discursivas (NAND, 2012).

As restrições obrigatórias precisam ser satisfeitas para haver correferência. São, basicamente, as morfológicas (concordância de gênero, número), as sintáticas (*c-command*) e as semânticas (compatibilidade semântica, animacidade). Já as restrições facultativas dizem respeito às preferências encontradas em diversas línguas para identificar o antecedente anafórico, como o paralelismo sintático e semântico, o centro catafórico, a recência, a frequência, a segmentação e a coerência (GARNHAM, 2001).

Entre as preferências que têm sido muito estudadas estão o estatuto do foco do antecedente e a relação semântica entre anáfora e antecedente, como descritos acima. Um sintagma nominal em foco é mais rapidamente selecionado como antecedente de uma expressão correferencial do que aquele que está fora de foco. Por foco, entende-se aqui o sintagma nominal na função e posição canônica de sujeito da frase ou o sintagma nominal clivado em início da oração. Já um antecedente fora de foco encontra-se em outra posição sintática, comumente na posição de complemento do verbo ou de objeto. As relações semânticas estabelecidas entre o elemento correferente e seu antecedente são muitas, sendo as mais estudadas as relações de tipicidade, de hierarquia, de sinonímia e de meronímia.

Com o objetivo de verificar nossas hipóteses de trabalho, realizamos uma investigação sobre a resolução anafórica em Português Brasileiro manipulando, em períodos complexos por coordenação, as relações de hierarquia semântica estabelecidas entre anáfora e antecedente e a proeminência sintática do termo antecedente. Nas próximas seções, iremos apresentar o estudos da movimentação ocular durante a leitura de um conjunto de frases construído e estudado previamente em experimento de leitura automonitorada feito por Leitão (2005).

O estudo de Leitão (2005) manipula a relação entre anáfora e antecedente, em estruturas coordenadas, contrastando uma relação de hipo-nímia e de hiperonímia com o objetivo de observar diferenças no custo de processamento na região do sintagma nominal anafórico e na região subsequente à anáfora, aqui chamada de *spillover*. Além da relação semântica, na presente investigação construímos um experimento, realizado em dois grupos distintos, em que procuramos analisar no primeiro grupo a relação semântica acima descrita na posição mais alta da frase, a de sujeito, enquanto no segundo grupo investigamos a mesma manipulação semântica na posição de objeto, ou seja, numa posição de menor saliência da oração quando comparada à posição de sujeito.

Além da hipótese básica norteadora desta investigação, que seria a preferência em estabelecer correferência com a posição sintática mais proeminente, o sujeito da oração, uma hipótese secundária projeta que haveria menor custo de processamento nas retomadas realizadas por sintagma nominal em relação de hiperonímia com seu antecedente, independentemente da posição sintática dos elementos correferentes. Outra de nossas hipóteses secundárias prediz que o custo de processamento do correferente em relação de hiperonímia com seu antecedente seria maior na posição de objeto do que na posição de sujeito da oração. No entanto, em ambas as posições, nossa hipótese prediz que o sintagma nominal correferente anafórico em relação de hiperonímia seria menos custoso para o processador sintático do que seu correspondente em relação de hiponímia.

Nosso estudo investigou, portanto, duas variáveis independentes: (i) hiperonímia na correferência e (ii) hiponímia na correferência, em dois grupos distintos: (i) antecedente e correferente na posição de sujeito (Grupo 1) e (ii) antecedente e correferente na posição objeto (Grupo 2).

5.2 Materiais e Métodos

5.2.1 Participantes

Testamos vinte e um participantes saudáveis, divididos em dois grupos, com visão normal ou corrigida ao normal, sem antecedentes neurológicos. A Tabela 7 apresenta os dados demográficos divididos por grupo em número, sexo, idade, escolaridade (em número de anos de estudo)²⁶ e lateralidade. O grupo 1 recebeu os estímulos em que manipulamos a correferência realizada por hipônimos e hiperônimos, na posição de sujeito tanto na retomada quanto no termo antecedente. Já o grupo 2 foi submetido a um conjunto de estímulos em que estava presente o mesmo tipo de manipulação, mas tanto a retomada quanto o termo antecedente se encontravam na posição de objeto.

Tabela 7 - Sumário dos dados demográficos do experimento do Capítulo 5.

Dados demográficos

	Todos	Grupo 1 Cond. Sujeito	Grupo 2 Cond. Objeto
N	21	10	11
Sexo	10 F / 11 M	7 F / 3 M	3 F / 8 M
Lateralidade	20 D / 1 E	9 D / 1 E	11 D / 0 E

²⁶ Estabelecemos os seguintes valores para escolaridade: ensino médio, 12 anos; graduação, 16 anos; mestrado, 18 anos; e doutorado, 22 anos.

Dados demográficos

	Todos	Grupo 1	Grupo 2
		Cond. Sujeito	Cond. Objeto
Idade	27,9 (5,5)	30,9 (5,1)*	25,1 (4,4)*
Escolaridade	17,9 (1,3)	18,1 (1,1) ‡	17,7 (1,3) ‡

Desvio padrão entre parênteses

* $F=3.864$ $p=0.081$

‡ $F=0.238$ $p=0.638$

5.2.2 Procedimento experimental e Estímulos

As frases experimentais foram distribuídas em quatro diferentes listas de estímulos (ver APÊNDICE B). Cada lista era composta por 10 frases experimentais entre 40 frases distratoras não relacionadas²⁷, totalizando quatro listas com 50 frases cada, sequenciadas aleatoriamente. Das 10 frases experimentais pertencentes a cada lista, cinco estavam na condição de hiponímia e cinco na condição de hiperonímia. Em duas listas, as frases experimentais apresentavam o antecedente e o co-referente na posição de sujeito e nas outras duas na posição de objeto. No total, registramos a movimentação ocular em 40 frases experimentais (10 por lista) e em 40 distratoras (que se repetiram em todas as listas). Cada lista iniciava com uma tela de instruções antes da fase de treinamento. Uma segunda tela de instruções era apresentada antes de iniciar a fase experimental propriamente dita.

Os participantes foram sentados em uma cadeira fixa com seus olhos a uma distância de aproximadamente 60 cm do monitor do rastreador ocular. Os testes ocorreram em dois espaços, em uma sala do Laboratório de Sistemas Complexos do Departamento de Física, no Campus do Pici, e em uma sala do Departamento de Letras Vernáculas, no Campus do Benfica. Ambos os departamentos pertencem à Universidade Federal do Ceará. Todos os participantes passaram pela calibração de acordo com os procedimentos padronizados do rastreador ocular Tobii T120. A calibração foi repetida sempre que necessário para atingir os níveis satisfatórios ao procedimento experimental. Participantes que não atingiram o nível desejado na fase de calibração não foram considerados válidos para este estudo.

Depois de ler a tela de instruções, os participantes precisavam dar um clique no *mouse* para iniciar a fase de treinamento. Duas frases distratoras foram lidas no início para que os participantes passassem por uma fase de aprendizagem. Após o treinamento, a segunda tela de instruções aparecia e os alertava sobre o início do experimento propriamente dito. Após a leitura das frases, o leitor precisava dar um clique no *mouse* para responder à pergunta

²⁷ As frases distratoras possuíam estrutura sintática não semelhante às frases experimentais, como pode ser conferido no Apêndice B.

de controle. Os participantes não foram informados sobre os objetivos do experimento e nenhuma instrução foi dada após a leitura da segunda tela de instrução. O experimento foi rodado individualmente no rastreador ocular Tobii T120, usando o software Tobii Studio 2.3.2.

Cada grupo de participantes foi convidado a ler uma das quatro listas. Após a leitura de cada frase, o participante visualizava uma pergunta de controle de atenção e, para o caso das experimentais, de controle de correferência. Os materiais experimentais foram previamente usados no estudo de Leitão (2005). O grupo 1 leu frases experimentais que consistiam em duas sentenças coordenadas pela conjunção aditiva “e” enquanto o grupo 2 leu sentenças coordenadas pela conjunção adversativa “mas”. Antes da apresentação de cada frase, uma cruz de fixação no canto esquerdo da tela, correspondente ao local do início da frase, era apresentada por dois segundos. Passados os dois segundos, era exibido na tela o slide contendo uma frase, disposta em uma única linha, escrita em fonte Calibri com tamanho de 25 pontos. Após a leitura da frase, o participante respondia à pergunta controle. Após a pergunta, a cruz de fixação reaparecia por mais dois segundos (ver exemplo desta sequência na FIGURA 6). Partimos do pressuposto que os participantes poderiam não estabelecer relação de correferência com um sintagma nominal antecedente. Se o participante estabelecesse a correferência, responderia afirmativamente à questão do tipo sim ou não, que lhe era exibida após a leitura de cada frase, como no exemplo (21).

(21) **[O macaco]**_i subiu na árvore mais próxima e depois **[o chimpanzé]**_{i ou j} avidamente comeu os frutos maduros.

Pergunta de controle de correferência:

(q.1) O chimpanzé subiu na árvore e comeu os frutos maduros?

Sim ou Não

Nas ocorrências em que o participante respondeu afirmativamente à pergunta, consideramos que este leitor estabeleceu a correferência entre o sintagma nominal [o chimpanzé] e o sintagma nominal [o macaco]. No grupo 2, procuramos realizar o mesmo tipo de sugestão na posição de objeto, conforme exemplo (22):

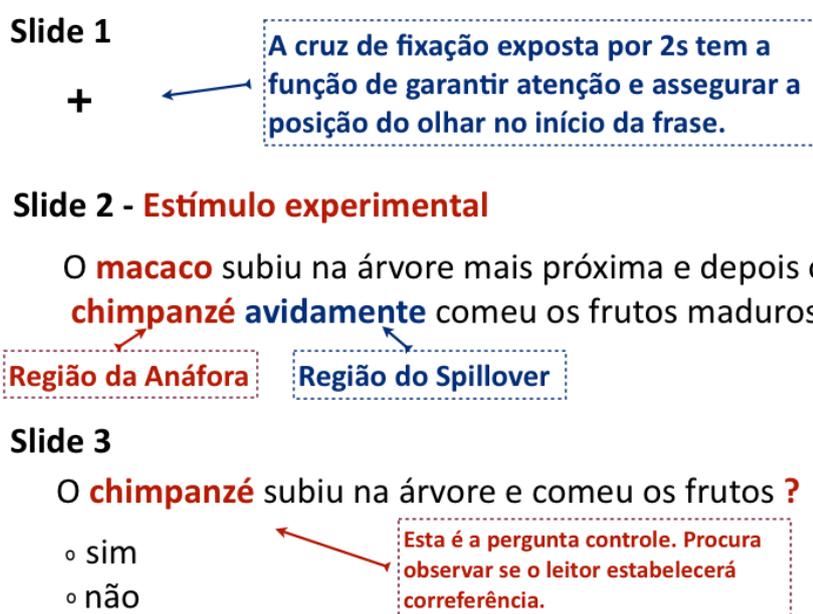
(22) Os biólogos avistariam **[um réptil]**_i no rio mas depois assustaram **[o animal]**_{i ou j} na margem.

Pergunta de controle de correferência:

(q.2) Os biólogos avistaram e depois assustaram o animal?

Sim ou Não

Figura 6 - Exemplo de uma sequência de estímulos apresentados ao participante e das regiões de interesse consideradas neste estudo. Em primeiro lugar, uma tela com apenas a cruz de fixação (1) é exibida por dois segundos, seguida da frase experimental (2) que fica na tela até o participante decidir dar um clique no mouse para passar para a próxima tela, onde será exibida a pergunta controle (3). Pode-se ver nesta figura, as duas regiões de interesse que foram consideradas para cálculo da duração da fixação, da duração da primeira fixação e do número de fixações.



5.2.4 Equipamento

A movimentação dos olhos foi gravada em um rastreador ocular da marca Tobii T120. Esse sistema registra a cada 8ms os movimentos oculares, sacadas e fixações, enquanto os participantes lêem as frases. O rastreador ocular é composto de um monitor de 17" que integra duas câmeras de alta velocidade, capazes de obter até 120 imagens por segundo. O sistema compensa pequenos movimentos de cabeça, eliminando assim a necessidade de um apoio para o queixo ou para a testa. Embora as câmeras rastreiem ambos os olhos, algumas vezes os dados são melhores para apenas um dos olhos. Este estudo considerou os dados obtidos a partir da média do olho esquerdo e direito. A vantagem deste sistema é que não há para o participante nenhum aparato visível, ele vê apenas o monitor, o que diminui a chance

de um viés causado pelo incômodo na posição da cabeça. Todos os participantes foram alertados antes do experimento de que o movimento de seus olhos seria gravado durante a sessão. A posição dos olhos foi gravada em 120 Hz.

O software de análise Tobii Studio 2.3.2 foi usado para calcular várias latências e medidas que descrevem o comportamento básico do olhar, as fixações. Usamos o filtro I-VT que calcula a velocidade de movimentação a partir de 20 ms, com um limiar de 30° por segundo. O tempo e o ângulo máximos permitidos entre as fixações foi de 75 ms e 0,5°, respectivamente. As fixações inferiores a 100 ms e superiores a 600 ms foram descartadas. Para assegurar as fixações nas áreas de interesse, desenhamos nas frases duas regiões conforme Figura 6. Em nossa análise, procedemos ao cálculo apenas dos dados capturados na leitura das frases em que os participantes estabelecem a correferência.

5.3 Variáveis dependentes

5.3.1 Performance comportamental

O resultado comportamental foi calculado como a média de acertos do grupo para cada condição experimental. Primeiro, determinamos o índice de acerto às questões do tipo sim ou não para cada participante e depois determinamos a média para cada tipo de correferência: hiponímia e hiperonímia.

5.3.2 Tempo total de leitura

Esta é a medida que representa o tempo total que o participante usou para ler a frase inteira. Conta-se a partir do início da apresentação do estímulo até o momento em que o leitor clica no mouse para passar ao próximo slide, que era a ação esperada para que ele pudesse responder à pergunta de controle. Contempla a leitura em toda a frase, não apenas nas regiões de interesse.

5.3.3 Tempo total de fixação

Esta é a medida que representa a soma das durações de todas as fixações dentro de uma região de interesse, isto é, contempla todas as leituras. As regiões de interesse estudadas foram a região da anáfora (crítica), correspondente ao sintagma nominal correferente, e região

do *spillover*, correspondente ao termo posterior ao correferente, normalmente um sintagma adverbial nos dois grupos. Caso o participante tenha feito mais de uma leitura daquela região, o que normalmente acontece, todas estas serão somadas para calcular o tempo total de fixação naquela região de interesse.

5.3.4 Número de fixações

Esta é a medida que representa o número de vezes que o participante fixa a região de interesse. Caso o participante tenha feito mais de uma leitura daquela região, todas estas fixações serão somadas para calcular o número de fixações daquela região.

5.3.5 Duração da Fixação

Esta é a medida que representa a média de todas as fixações dentro da região estudada. Caso o participante tenha feito mais de uma leitura daquela região, duas, três ou quatro, todas as suas fixações serão somadas para serem divididas pelo número de vezes que o participante fixou o olhar na região. Caso o participante não tenha fixado a região, seus dados não terão sido usados para efeito de cálculo.

5.3.6 Duração da Primeira Fixação

Esta medida é o cálculo da duração da primeira fixação em uma região de interesse, o que significa dizer que é a primeira fixação em qualquer das regiões de interesse estudadas. Se não houve fixação de um participante dentro da região de interesse, este dado não era computado para calcular a média e os efeitos estatísticos.

5.4 Resultados

5.4.1 Performance comportamental

Os participantes, de ambos os grupos, obtiveram uma média alta, atingindo 90% de acertos, se considerado o grupo como um todo. Este resultado já era de certa forma esperado, porque acreditávamos que os participantes responderiam afirmativamente à

pergunta de controle da correferência, ou seja, que estabeleceriam a correferência independentemente da relação semântica e sintática entre anáfora e antecedente. No entanto, houve maior índice de acerto na condição de hiperonímia, sugerindo que o estabelecimento da correferência por um termo superordenado poderia ser mais intuitivo. Percebemos um índice ligeiramente inferior no estabelecimento da correferência com sintagmas nominais hiponímicos (TABELA 8), tanto para a condição de sujeito quanto para a de objeto, mas os testes estatísticos, contudo, não revelaram diferença significativa ($F [1, 19] = 1.137, p=0.297$).

Tabela 8 - Média da performance comportamental para as diferentes condições.

		Hiperônimo	Hipônimo
Grupo 1	Cond.		
	Sujeito	0.90 (0.15)	0.88 (0.23)
Grupo 2	Cond.		
	Objeto	0.97 (0.07)	0.87 (0.27)

Desvio padrão entre parênteses

A aleatoriedade e o grande número de distratores, na proporção de quatro para cada item experimental, afastou qualquer efeito de aprendizagem ao longo do teste. Mesmo assim, observamos os dados à procura de tal efeito, o que não foi encontrado. Para nossa surpresa, um dos participantes, do grupo 1 (posição sintática de sujeito) não estabeleceu a correferência na condição de hipônimo em nenhum dos itens, demonstrando certa coerência na sua tomada de decisão. Nos demais participantes, o não estabelecimento da correferência se deu de forma pontual, muito provavelmente devido a variáveis difíceis de serem controladas, como conhecimento de mundo, por exemplo. Não foram encontrados efeitos principais na comparação das performances comportamentais para todas as condições experimentais, intra e inter grupos.

5.4.2 Movimentação ocular

Como apresentado na seção 5.3, iremos reportar os resultados encontrados em cinco diferentes medidas de movimentação ocular. A primeira medida que apresentamos na tabela 3 é o tempo total de leitura entre as condições (ver 5.3.2). Encontramos que o estabelecimento da correferência nos casos de hiperonímia na posição de sujeito é em torno de 10% mais rápido do que no caso de hiponímia, diferença estatisticamente significativa quando rodada a análise da variância (ANOVA) para medidas repetidas, com o grupo de

fatores Hiperônimo vs. Hipônimo, para os resultados obtidos no Grupo 1 (n=10), por participante (F [1, 9]=14.694, p=0.004) (GRÁFICO 2) e por item (F [1, 9]=5.164, p=0.049) (GRÁFICO 3). Encontramos um resultado marginalmente significativo, quando rodada a ANOVA para a combinação de todos os resultados, inter e intra grupos, ou seja, das quatro condições experimentais estudadas (F [1, 39]=2.293, p=0.092) . Para o Grupo 2, que leu a correferência na posição de objeto, nenhum efeito principal, nem por item nem por participante, foi encontrado para a medida de tempo total de leitura.

Tabela 9 - Tempo médio total de leitura da frase em segundos.

Tempo total de leitura da frase (em segundos)		
	Hiperônimo	Hipônimo
Por item		
Cond. Sujeito	7,17 (1,21)	8,53 (1,36)
Cond. Objeto	7,65 (2,24)	8,17 (1,96)
Por participante		
Cond. Sujeito	7,31 (2,65)	8,33 (2,78)
Cond. Objeto	7,61 (2,63)	7,65 (3,75)

Desvio padrão entre parênteses

Gráfico 2 - Tempo total de leitura por participante.

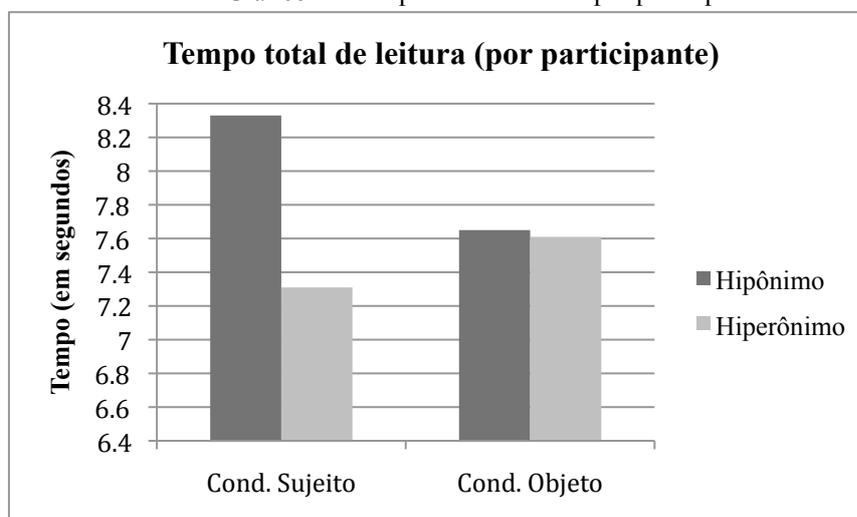
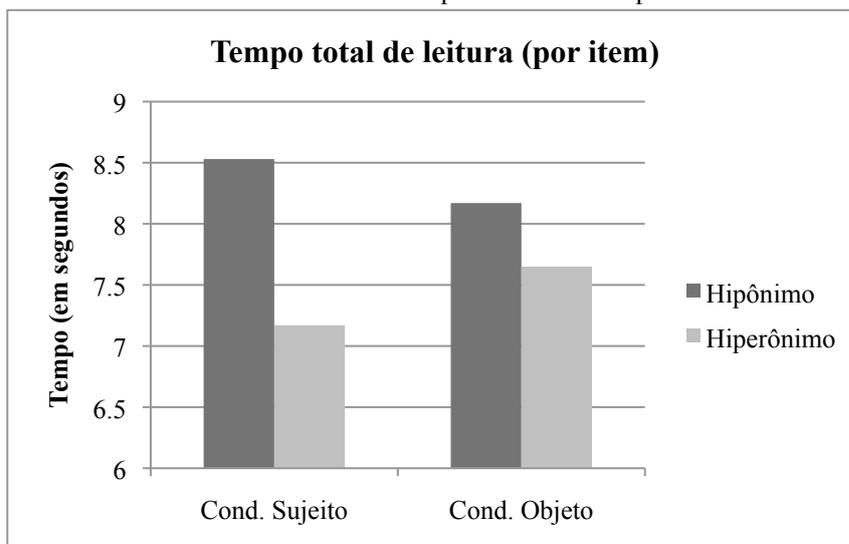
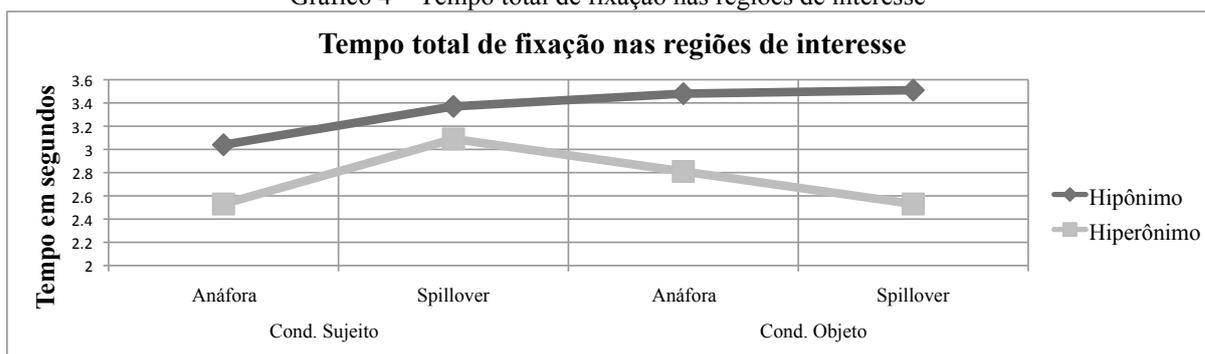


Gráfico 3 - Tempo total de leitura por item.



A medida do tempo total de leitura das estruturas coordenadas é reportada aqui porque entendemos como importante para avaliar, ainda que indiretamente e sem a precisão de medidas, como o tempo da primeira e da segunda leituras, o número de releituras do estímulo ou de partes do estímulo, dados que podem ser mais bem observados, quando o contrastamos com o tempo total de fixação (GRÁFICO 4) e com o número médio de fixações (GRÁFICO 5), nas regiões de interesse estudadas, região crítica (anáfora) e região posterior à anáfora (*spillover*).

Gráfico 4 – Tempo total de fixação nas regiões de interesse

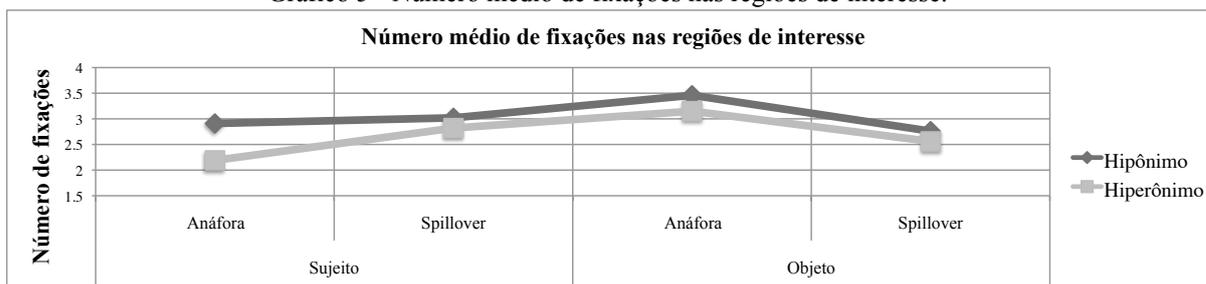


No Gráfico 4, a linha escura representa os tempos totais de fixação (a soma de todas as fixações) para a retomada por hipônimo e a linha clara representa a retomada por hiperônimo.

Pode-se observar neste gráfico que os tempos para a condição hiperônimo são sistematicamente menores do que para a condição de hipônimo, assim como nos mostram as medidas de número médio de fixações, representadas pelo Gráfico 5, abaixo.

Calculamos a ANOVA por participante para comparar a retomada por hipônimo e por hiperônimo no tempo total de fixação (GRÁFICO 4). Encontramos diferença significativa apenas para o Grupo 2 (n=11), em que se contrasta a relação semântica na posição de objeto. O efeito principal encontrado se refere às duas regiões, a da anáfora ($F [1, 10]=5.097$, $p=0.048$) e a região do *spillover* ($F [1, 10]=5.722$, $p=0.038$). Importante notar que, no caso da condição de objeto, a região do *spillover* coincide com o final da sentença coordenada, o que não pode deixar de ser levado em consideração, uma vez que estudos anteriores apontam tempos maiores nos finais das sentenças, em um efeito conhecido na literatura como encapsulamento²⁸, o que torna difícil avaliar se o efeito é devido ao custo de processamento da variável analisada ou se é devido a um momento em que há a compreensão total da frase.

Gráfico 5 - Número médio de fixações nas regiões de interesse.



Em relação ao número de fixações (GRÁFICO 5), apenas encontramos diferenças significativas para o Grupo 1 (n=10), na região da anáfora ($F [1, 9]=9.494$, $p=0.004$). Se observado o gráfico, comparando-se as linhas do hiperônimo e do hipônimo, percebe-se que é constante o menor número de fixações para a condição de hiperônimo, em qualquer das regiões estudadas. Tais diferenças mínimas mas repetidas nos levam a hipotetizar que uma amostra maior poderia gerar efeitos significativos no contraste dos dados para todas as variáveis estudadas neste trabalho.

Abaixo, vemos, sumarizados na Tabela 10, os dados referentes às medidas de duração média e de duração da primeira fixação. Todos os dados reportados abaixo foram calculados por item, tendo sido considerados para o cálculo apenas os itens em que os participantes estabeleceram a correferência. Estas duas últimas medidas da movimentação

²⁸ O termo equivalentemente em inglês é “wrap-up”.

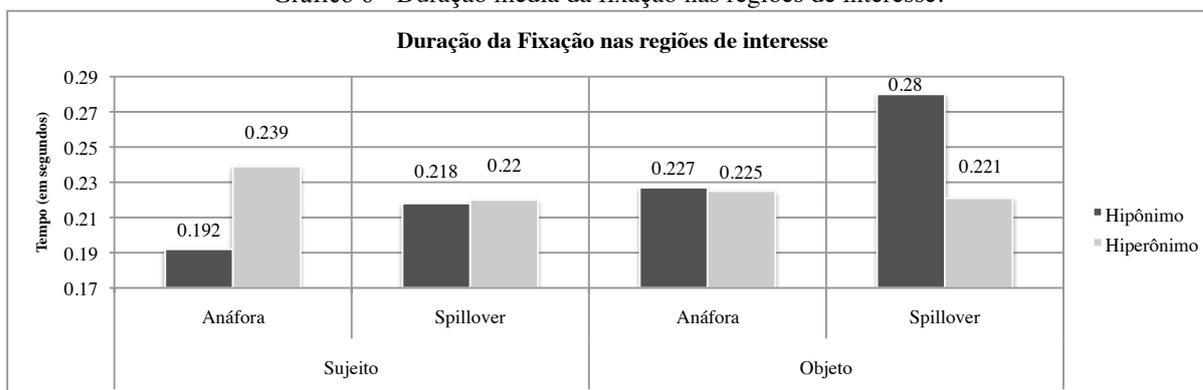
ocular que descrevemos abaixo são as que mais nos intrigaram neste trabalho. Diferentemente das projeções feitas a partir do trabalho de Leitão (2005), encontramos efeito significativo, oposto ao que esperávamos, para os tempos de duração média e de primeira fixação tanto para o Grupo 1 quanto para o Grupo 2. O custo de leitura foi maior ou equivalente para a condição de hiperônimo em todas as regiões estudadas, com exceção do grupo 2 na região do *spillover*.

Tabela 10 - Sumário dos tempos encontrados para duração média da fixação e duração da primeira fixação

	Duração média da fixação		Duração da primeira fixação	
	Hiperônimo	Hipônimo	Hiperônimo	Hipônimo
Região da Anáfora				
Grupo 1 - Sujeito	0,239 (0,05)	0,192 (0,07)	0,258 (0,06)	0,209 (0,07)
Grupo 2 - Objeto	0,225 (0,07)	0,227 (0,08)	0,255 (0,07)	0,224 (0,08)
Região do <i>Spillover</i>				
Grupo 1 - Sujeito	0.220 (0,06)	0.218 (0,07)	0.254 (0,08)	0.253 (0,09)
Grupo 2 - Objeto	0.221 (0,08)	0.280 (0,07)	0.234 (0,08)	0.249 (0,09)

Desvio padrão entre parênteses

Gráfico 6 - Duração média da fixação nas regiões de interesse.

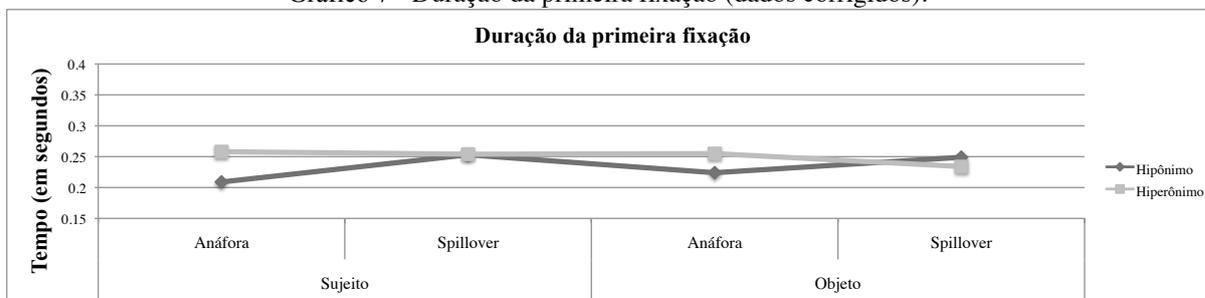


A barra escura representa no Gráfico 6 a duração média da fixação nas regiões de interesse para a condição de hipônimo e a barra clara representa a duração média da fixação nas regiões de interesse para a condição de hiperônimo. O Gráfico 6 não apresenta uma tendência clara de maior ou menor custo em nenhuma das variáveis independentes estudadas.

No Gráfico 7, a linha escura representa a duração da primeira fixação para a condição de hipônimo e a linha clara representa a duração da primeira fixação para a condição de hiperônimo. Pode-se observar (gráfico 7) que a duração da primeira fixação é estatisticamente significativa para as condições de hipônimo e hiperônimo nas posições de sujeito ($F [1, 31] = 4.242, p = 0.048$) e objeto ($F [1, 25] = 9.060, p = 0.006$) na região da co-referência. No entanto, tal medida é indiferente às condições experimentais nas regiões

posteriores ao SN co-referente. As médias foram obtidas considerando apenas os itens corretos, ou seja, cuja resposta à pergunta controle atestava o estabelecimento da co-referência.

Gráfico 7 - Duração da primeira fixação (dados corrigidos).



A medida da duração da primeira fixação é vista como importante no estudo do *parsing*²⁹ sob qualquer variável estudada e, no nosso caso em particular, no estudo do processamento da anáfora correferencial, porque é a medida *on-line* mais confiável para aferir custo de processamento. As medidas do tempo de uma segunda leitura já não são tão confiáveis para falar de *parsing*, pois podem ser interpretadas como tempos relativos à reanálise. Isto quer dizer que, para se falar de fatores intervenientes na resolução anafórica, as medidas totais ou aquelas que somam dados de releituras não são tão confiáveis quanto os tempos encontrados durante a primeira leitura. É-nos conveniente, portanto, esclarecer aqui a diferença entre a medida da duração da primeira fixação e a medida, não reportada neste estudo, da duração da primeira leitura. A primeira leitura é uma medida muito usada pela Psicolinguística para analisar o *parsing*. Para calculá-la é preciso somar todas as fixações (incluindo a primeira fixação, a medida que descrevemos na Tabela 10) que ocorrem em uma palavra antes de o olho realizar uma sacada e fixar uma palavra em outra posição da frase.

Apesar de as medidas de duração média e de primeira fixação (TABELA 10) não corroborarem os dados anteriormente obtidos por Leitão (2005), em estudo de leitura automonitorada, as três primeiras medidas da movimentação ocular acima comentadas, o tempo total de leitura, o tempo total de fixação e o número de fixações, sugerem que o custo de processamento do correferente em relação de hiperonímia com seu antecedente tende a ser mais baixo do que na relação de hiponímia. O que ainda nos resta compreender é em que momento da leitura ocorre a resolução anafórica, uma vez que nossos resultados não nos autorizam dizer que é no momento da leitura do sintagma nominal anafórico ou na região do

²⁹ Parsing é o termo em inglês para processador sintático, muito usado como termo técnico nos estudos publicados em língua portuguesa.

spillover, nos quais não mensuramos sobrecarga de processamento em todas as condições investigadas.

5.5 Discussão

Este trabalho procurou observar, em estruturas coordenadas, duas preferências semânticas e sintáticas que poderiam interferir na resolução da correferência anafórica de sintagmas nominais, na linha dos estudos anteriormente comentados no Capítulo 2 e na seção 5.1, que investigaram a sobreposição semântica dos termos correferentes, o custo de processamento do centro anafórico e o custo dos sintagmas nominais correferentes em situação de alta e baixa funcionalidade discursiva. Descrevemos, nas seções 5.2, 5.3 e 5.4, dois experimentos relacionados, em que foram investigadas as relações de hiponímia e hiperonímia do sintagma nominal anafórico com seus antecedentes (sobreposição semântica) em grupos de frases cujos termos correferentes estavam em posição sintática de sujeito e de objeto.

Para tanto, usamos os mesmos estímulos estudados por Leitão (2005) em experimento de leitura automonitorada. Avaliamos a performance comportamental, ou seja, a capacidade dos participantes de estabelecer correferência, além de analisar os dados obtidos da movimentação ocular de 21 participantes durante a leitura de frases. As medidas analisadas foram o tempo total de leitura, o tempo total de fixação, o número de fixações, a duração média da fixação e a duração da primeira fixação, na frase e em duas regiões de interesse: a região da anáfora e a região do *spillover*.

No primeiro experimento, relativo ao Grupo 1, que leu a correferência na posição de sujeito, encontramos resultados que confirmam os achados de Leitão (2005) para o objeto, que observou haver sistematicamente um tempo 20% menor na leitura das frases com retomada por hiperônimo. Diferentemente do estudo de 2005, a investigação com o rastreador ocular, que oferece maior precisão, encontrou uma diferença nos custos que gira em torno dos 15% (TABELA 10). Para o grupo 2, no entanto, não foi possível observar diferenças significativas em relação à retomada por hipônimo e hiperônimo.

A explicação dos resultados pode ser vista sob a perspectiva que compreende serem as retomadas com funcionalidade discursiva aquelas que acontecem na maioria dos casos, ou seja, nos casos em que o antecedente está muito acessível. Na manipulação dos estudos de Psicolinguística, costuma-se investigar todas as condições, mas parece haver um

efeito do hábito e da frequência nos resultados que são registrados por meio de rastreadores oculares. Queremos dizer que é provável que conheçamos uma regra sintática para a correferência de entidades mencionadas no discurso em posição de alta acessibilidade, e que é provável que não tenhamos internalizada uma regra para a correferência de sintagmas de baixa acessibilidade ou que não constituem foco discursivo, devido à baixa frequência com que acontece, ainda mais quando se trata de correferência estabelecida por sintagma nominal e não por pronome nulo ou pleno, o que nos leva igualmente a pensar que a distância semântica (ALMOR, 1999) entre correferentes pode ainda interferir na resolução anafórica.

Os resultados, referidos como intrigantes quando descrevemos os dados da duração média e da primeira fixação, talvez possam ser analisados posteriormente em uma amostragem maior, composta de grupos mais homogêneos. Observamos, ao fazer o pareamento dos dados demográficos, que há uma diferença entre os grupos de participantes, que poderia implicar em mudança de comportamento e variação nos resultados. Não nos foi possível fazer o pareamento dos grupos em relação à idade e à escolaridade, além de termos observado que o grupo 1 é formado mais por mulheres e que o grupo 2 é constituído, em sua maioria, por indivíduos do sexo masculino. O sexo dos participantes, em princípio, não deveria ser um fator a considerar, mas já é conhecido na literatura, sobretudo na que diz respeito aos estudos comportamentais em Psicologia, que os sexos têm comportamentos distintos. Como a medida ocular durante a leitura é uma medida comportamental, é possível que tenha havido interferência deste fator, inclusive nos dados reportados de duração da primeira fixação.

Na análise dos resultados do grupo 2, encontramos diferenças significativas de custo de processamento na região posterior à anáfora, a região do *spillover*. Nesta região era necessário o tratamento de sintagmas preposicionados adverbiais e normalmente era coincidente com a posição final das estruturas coordenadas estudadas. Os resultados, no entanto, apontam a sobrecarga somente para a condição de hipônimo, tanto na medida do tempo total de fixação (GRÁFICO 4) quanto na medida da duração média de fixação (TABELA 11), indicando que a resolução anafórica pode acontecer em dois estágios distintos (VAN GOMPEL, LIVERSEDGE E PEARSON, 2004).

Como sugestão para trabalhos posteriores que queiram investigar a relação semântica entre elementos correferentes, acreditamos que é interessante pesquisar tal relação em períodos complexos por subordinação, em casos de antecedentes indefinidos, retomados por sintagmas nominais definidos, além de procurar investigar se as relações estabelecidas

pelos participantes entre os termos correferenciados é, de fato, uma relação de pertencimento, de classe, ou se em alguns casos pode ser tratada como sinonímia ou como uma sobreposição total dos traços semânticos, sem acréscimo de informação nova.

O confronto entre as hipóteses levantadas para este estudo e os resultados que reportamos será amplamente discutido no Capítulo 7, em nossas considerações finais.

6 ESTUDO DE PREFERÊNCIAS SINTÁTICAS

Neste capítulo, iremos apresentar os resultados da pesquisa feita sobre a resolução anafórica de pronomes nulos e plenos, na condição de catáfora e anáfora, em períodos complexos por subordinação. Para realizar a pesquisa, dois experimentos foram conduzidos: um de compreensão e outro de produção. O experimento de produção é, na verdade, uma sondagem preliminar e com modesto alcance, que objetivava verificar quais escolhas os participantes fariam diante da tarefa de estabelecer correferência com um antecedente em posição de sujeito ou de objeto. Os resultados podem ser conferidos na seção 6.6.

O experimento de compreensão, foco principal deste capítulo, consiste na leitura de frases manipuladas experimentalmente, formadas por uma oração principal, cujo verbo necessitava de dois argumentos, e por uma oração subordinada temporal, cujo sujeito poderia ser uma categoria vazia [*pro*] ou um pronome pleno.

Nosso objetivo, ao realizar o estudo de compreensão, foi testar a hipótese de que há uma especialização para cada uma das formas pronominais do Português Brasileiro. Sabe-se que a especialização de funções nas línguas resulta em uma distribuição complementar. Portanto, observar se o PB possui um sistema pronominal em distribuição complementar é uma das hipóteses tanto do estudo de compreensão quanto do estudo de produção. Uma conjectura adicional é a de que, qualquer que seja a expressão anafórica pronominal, em línguas com sujeito nulo e pleno, os pronomes nulos tendem a estabelecer correferência com a entidade mais saliente no discurso, que, muito frequentemente, ocupa a posição de Spec IP da oração (ver seção 2.2.2), possui proeminência sintática de posição de sujeito e satisfaz a noção de tópico discursivo. Os estudos que procuram compreender as razões da especialização (de sua falta ou presença) têm observado padrões ligeiramente distintos para o Português Brasileiro e Europeu, o que justifica a necessidade de novos estudos que investiguem melhor o fenômeno.

Nosso estudo se insere no âmbito destes esforços e, particularmente, dos esforços que buscam por evidências de processamento *on-line*, a partir de protocolos experimentais que observem a resolução anafórica do sistema pronominal em curso. Estudos desta natureza, que investiguem o sistema do PB, ainda são contribuições necessárias que aportam evidências empíricas com o objetivo complementar estudos teóricos.

Se o sistema pronominal do PB possui especialização funcional, poderíamos hipotetizar que há uma gradação dos marcadores de acessibilidade (ARIEL, 1991) na

retomada de antecedentes em diferentes posições estruturais da oração. Se esta hipótese não se confirma, é preciso continuar a investigar as razões que levam línguas a manterem, na mesma sincronia, em seu sistema, a presença concomitante de uma categoria vazia e uma plena. Se dentro do âmbito da Linguística não encontrarmos uma explicação plausível para estes sistemas pronominais, é preciso investigar se as razões da existência de um sistema duplo não especializado reside em outros aspectos cognitivos, como a capacidade de armazenamento da memória de trabalho.

6.1 Estudo do processamento de pronomes nulos e plenos: introdução

Para uma anáfora ser resolvida é preciso identificar corretamente seu antecedente. A anáfora pronominal representa o problema de modo prototípico, já que compõe a base da escala de acessibilidade (ver Capítulo 2). Isto porque o pronome não traz informação semântica nenhuma que permita identificar um único antecedente quando vários possíveis coexistem. As informações carregadas por um pronome em língua portuguesa são basicamente morfossintáticas, porque se restringem à concordância de gênero e número. Diferentemente de anáforas sob a forma de SNs que fornecem informação sintática e semântica, a identificação do antecedente de uma anáfora pronominal, em contextos ambíguos, está longe de ser um problema trivial.

A escolha entre um pronome nulo e um pronome pleno em línguas *pro-drop* tem sido investigada nos últimos anos em diferentes línguas, relacionadas e não-relacionadas³⁰ (ALONSO-OVALLE *et al.*, 2002; CARMINATI, 2002; CORREA, 1998; COSTA; FARIA; MATOS, 1998; MAIA, 1997; FOLEY; VAN VALIN, 1984; FONSECA; GUERREIRO, 2012; HUANG, 1989; LUEGI, 2012; MERIDOR, 2006; MORGADO, 2011; MELO; MAIA, 2005).

Carminati (2002), em pesquisa de doutorado, apresentou uma série de evidências experimentais que sustentam fortemente a hipótese de que as línguas que possuem um sistema pronominal composto por dois pronomes, um nulo e um pleno, atribuem funções especializadas para cada uma deles, que operam em distribuição complementar. Tais evidências foram comprovadas no escopo da oração. A hipótese proposta por Carminati (2002), em estudo que examinou a língua italiana, ficou conhecida como Hipótese da Posição do Antecedente (ver Capítulo 2), que seria uma estratégia do *parser* de atribuir para os

³⁰ Línguas relacionadas são línguas provenientes do mesmo ramo.

pronomes nulos os antecedentes em posição estrutural de sujeito e para os pronomes plenos os antecedentes em posição estrutural mais baixa, relativamente à posição do sujeito.

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

Nos dados coletados por Carminati (2002) também foi observado que há uma diferença de força entre pronome nulo e pleno. Dito de outro modo, significa que, enquanto a preferência do pronome nulo pelo antecedente em posição de Spec IP é muito robusta, o pronome pleno apresenta certa flexibilidade quanto retém a posição estrutural de objeto. Esta flexibilidade pode ser observada nos resultados de Sorace e Filiaci (2006), na Tabela 11, mais adiante, e em nossos resultados, analisados nas seções seguintes deste capítulo.

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

Diante deste quadro teórico, conduzimos um experimento para testar se haveria especialização sintática da categoria vazia *pro* para falantes do Português Brasileiro. A idéia do experimento surge para observar o efeito desta categoria vazia em PB, contribuindo com evidências de processamento *on-line*, um esforço que se soma aos trabalhos recentemente concluídos sobre o Português Europeu e o Português Brasileiro (MORGADO, 2011; LUEGI, 2012). Luegi (2012) investigou o estatuto do pronome nulo comparando as variedades brasileira e portuguesa, mas o fez sob variáveis independentes distintas das nossas, porque seu objetivo era primordialmente controlar o estatuto do foco e observar se mudanças na ordem canônica da língua portuguesa ocasionavam efeitos na atribuição de antecedentes a sujeitos

nulos, ou seja, se a correferência forte entre nulo e posição estrutural em Spec IP era uma restrição sintática ou uma preferência discursiva.

Com base na restrição do pronome pleno de Montalbetti (1984), que diz que em várias línguas românicas os pronomes plenos, diferentemente dos pronomes nulos, não podem ser ligados localmente, Maia (1997) estudou em Português Brasileiro a realidade psicológica da categoria vazia em posição de objeto, característica particularmente mais frequente em Português Brasileiro do que em Português Europeu. Seus resultados apontam na direção de uma forte correlação para a correferência entre antecedente proeminente (tópico e posição estrutural) e categoria vazia em posição de objeto. Maia (1997) também estudou o preenchimento da posição de objeto por um pronome pleno, “ele”, construção permitida para Português Brasileiro, mas considerada agramatical em Português Europeu.

Duarte (1995) advoga que o Português Brasileiro vem preenchendo a posição de sujeito na primeira e segunda pessoa do verbo, já que se constata perda da riqueza morfossintática número pessoal. Sugere, no entanto, que o preenchimento da terceira pessoa parece ser uma tendência, mas que a frequência de atualização da terceira pessoa verbal observada foi relativamente bem menor à encontrada para as primeiras e segundas pessoas verbais.

O objeto dos dois experimentos aqui reportados, um de compreensão e outro de produção, é a resolução anafórica em período complexo, formado por uma oração principal e uma oração subordinada temporal. Na oração principal há dois argumentos do verbo, um SN na posição de sujeito e um outro SN na posição de objeto, cujo núcleo é um nome próprio frequente em PB, como “a Maria”, “a Lúcia”, “o David”, “o André”. As características morfológicas dos nomes próprios foram equiparadas, porque ambos podem ser do gênero masculino ou feminino. A oração subordinada temporal é formada, por sua vez, de uma conjunção temporal “quando” ou “enquanto” e de uma oração cujo sujeito pode ser um pronome pleno “ele” ou “ela” ou um pronome nulo (a categoria vazia *pro*). A subordinada temporal ainda poderia anteceder ou suceder a oração principal, criando, com o antecedente, uma relação catafórica (quando antecede a oração principal) ou anafórica (quando sucede a oração principal), como nos mostram os exemplos (23) e (24).

(23) [A Maria]_i conversava com [a Joana]_j enquanto [pro]_{i/j}/[ela]_{i/j} cozinhava.

(24) Enquanto [pro]_{i/j}/[ela]_{i/j} cozinhava, [a Maria]_i conversava com [a Joana]_j.

O que objetivamos examinar foi o processamento e a decisão do falante de PB diante de uma ambiguidade, que envolviam competidores, tanto na compreensão dos períodos quanto na tarefa de produção de períodos complexos. Com o registro da movimentação ocular na tarefa de compreensão, procuramos indícios de reanálise ou de elevação de custos de processamento durante a resolução das ambiguidades.

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

Tabela 11 – Dados do experimento de Sorace e Filiaci (2006)
 Percentual de respostas comportamentais de atribuição do antecedente

	Anáfora		Catáfora	
	Pronome nulo	Pronome Pleno	Pronome nulo	Pronome Pleno
Sujeito	51%	8%	85%	12%
Objeto	44%	82%	11%	24%
Outro	5%	11%	4%	64%

Os percentuais mais altos encontrados por Sorace e Filiaci (2006) foram marcados em negrito. Note-se que a catáfora parece causar um efeito na atribuição da sua categoria vazia ao primeiro competidor apresentado. Perguntamo-nos se seria este um efeito devido mais às limitações da memória de trabalho, como nos sugere Almor (1999), ou um efeito de restrição do processador sintático. Se observarmos o percentual de correferência entre o sujeito em posição anafórica e o pronome nulo, constatamos que os resultados de Sorace e Filiaci (2006) ficam em torno de valores aleatórios, ou seja, em torno dos 50%, o que sugere a não capacidade do participante de língua italiana de desfazer a ambiguidade sugerida durante a leitura do período complexo. O que se percebe também na leitura da Tabela 11 é que, embora Sorace e Filiaci (2006) tenham apresentado três opções (sujeito, objeto e outro), as decisões dos participantes se concentraram entre sujeito e objeto.

O estudo de Fonseca e Guerreiro (2012) apresenta resultados semelhantes aos de Sorace e Filiaci (2006), mas diferencia-se sobretudo na correferência entre antecedente e

pronome pleno em posição catafórica, como se pode perceber ao analisar a Tabela 12 abaixo, com os dados encontrados em sua pesquisa.

Tabela 12 - Dados do experimento de Fonseca e Guerreiro (2012)

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

Se compararmos os resultados do estudo de Fonseca e Guerreiro (2012) e de Sorace e Filiaci (2006), perceberemos que houve em PB uma preferência clara de correferência entre o pronome nulo em posição anafórica e o sujeito da oração principal, diferentemente do que encontraram Sorace e Filiaci (2006), cujos resultados não permitem afirmar qualquer preferência uma parecem aleatórios.

Os experimentos, reportados a seguir, nos fazem acreditar que a pesquisa que empreendemos, sobre a compreensão e produção de períodos complexos por subordinação, investiga, ao mesmo tempo, as preferências de co-indexação e as restrições obrigatórias resolvidas por nosso processador. Nossa pesquisa procura, com isso, os indícios que nos levem a definir melhor quais são essas restrições obrigatórias e facultativas no âmbito do processador sintático e no que diz respeito ao processamento de pistas semânticas e pragmáticas que condicionam a resolução da correferência pronominal em PB.

6.2 Materiais e Métodos

6.2.1 Participantes

O grupo testado foi composto de vinte e dois (22) participantes saudáveis, adultos, com visão normal ou corrigida ao normal, sem histórico neurológico prévio e livres da ingestão de medicamentos que afetam o sistema nervoso central. Não foram escolhidos, para participarem dos experimentos, participantes que apresentavam astigmatismo, mesmo corrigido pelo uso de óculos. Esta decisão foi tomada após a verificação de que havia maior perda de dados dos participantes astigmatas. Os dados demográficos, na Tabela 13, abaixo, descrevem N (o número de participantes), sexo, idade, nível de educação e lateralidade. Os participantes foram convidados a colaborar com o estudo de modo voluntário. Assinaram

consentimento formal (ver modelo em ANEXO) e estavam cientes de que suas identidades estariam preservadas e de que teriam sua movimentação ocular registrada durante a leitura dos estímulos.

Tabela 13 – Dados demográficos do estudo sobre processamento de pronomes nulos e plenos

Dados demográficos	
N	22
Sexo	20 Masculino / 2 Feminino
Idade	21.8 (± 1.28)
Nível educacional	14 anos de escolarização ³¹
Lateralidade	21 Destros / 1 Canhoto

Desvio padrão entre parênteses

6.2.2 Estímulos e Procedimento experimental

O conjunto de estímulos experimentais (ver APÊNDICE C) foi distribuído em quatro diferentes listas. Cada lista foi composta por 16 frases experimentais entre 48 frases distratoras relacionadas³², totalizando quatro listas com 64 frases cada, sequenciadas aleatoriamente. As quatro condições experimentais (anáfora, catáfora, pronome nulo e pronome pleno) foram distribuídas entre as 16 frases de cada lista. Cada participante leu um conjunto formado por quatro frases na condição de pronome nulo em posição anafórica (PNAN), quatro na condição de pronome pleno em posição anafórica (PPAN), quatro na condição de pronome nulo em posição catafórica (PNCA) e quatro na condição de pronome pleno em posição catafórica (PPCA), configurando um design em quadrado latino do tipo 2 x 2.

Somando-se todas as listas, registramos a movimentação ocular em 352 frases, com a participação de 22 voluntários que leram, cada um, 16 frases experimentais, divididas em 4 listas. Em suma, o design de nosso experimento pode ser visualizado abaixo:

$$(22_P * 4_L * 2_{V1} * 2_{V2})^{33} = 352$$

³¹ Os participantes eram alunos matriculados no 4º semestre de graduação da Universidade Federal do Ceará.

³² Neste experimento, as orações distratoras eram estruturalmente relacionadas.

³³ Há diferentes modos de visualizar um design. Escolhemos ressaltar os números de P (participantes), L (listas), V1 (variável independente 1 – pronome nulo ou pleno) e V2 (variável independente 2 – catáfora ou anáfora) para evidenciar o total de frases experimentais analisadas.

O número de 22 participantes é o resultado de decisão metodológica por excluir dois participantes cuja movimentação ocular não chegou aos 90% de captura, o limiar mínimo estabelecido para considerar os dados válidos para análise.

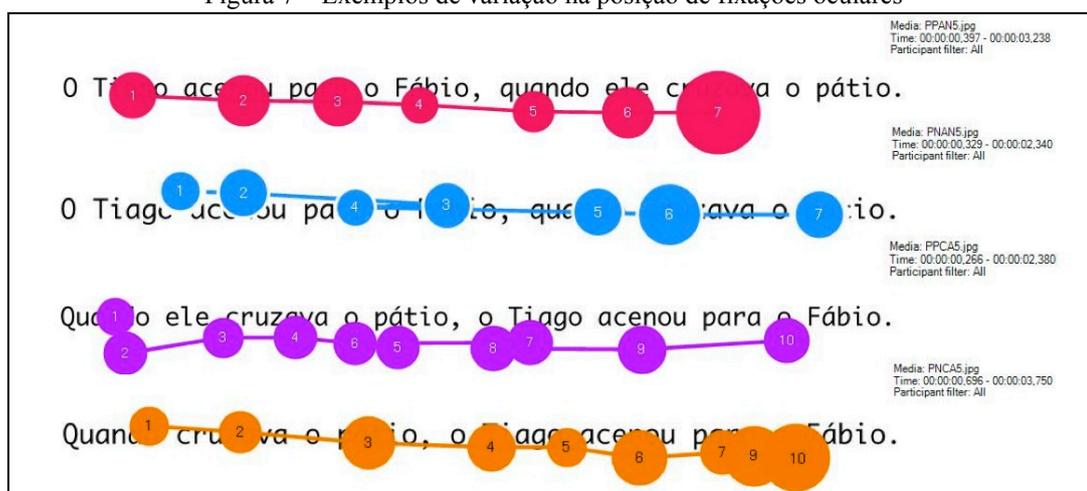
Cada lista iniciava com uma tela de instruções antes da fase de treinamento. Na fase de treinamento, frases distratoras eram apresentadas aos participantes para que pudessem se familiarizar com o procedimento experimental. Após a fase de treinamento, uma segunda tela de instruções era apresentada antes de iniciar a fase experimental propriamente dita.

Os procedimentos de coleta são os mesmos do estudo reportado no capítulo anterior. Consistia do convite aos participantes para passarem por um teste que demoraria em torno de 20 minutos, compreendidas as fases de treinamento e experimentais. Assim como no experimento anteriormente relatado, os participantes tiveram seus movimentos oculares gravados a uma distância de aproximadamente 60 cm do monitor do rastreador ocular Tobii T120, controlado pelo procedimento de calibração do equipamento. Os testes foram conduzidos individualmente. O programa usado para apresentação e registro dos estímulos foi o Tobii Studio 2.3.2. Todos os testes ocorreram em uma sala do Laboratório de Sistemas Complexos do Departamento de Física da Universidade Federal do Ceará.

Os procedimentos padronizados de calibração foram aplicados a todos os participantes. A calibração foi repetida sempre que necessário para atingir os níveis satisfatórios ao procedimento experimental. Participantes que não atingiram o nível desejado na fase de calibração não chegaram a ser testados. É raro, mas pode acontecer, por razões que não cabem serem explicadas aqui, casos em que os olhos não são capturados adequadamente pelo equipamento, por isso a importância de uma calibração bem feita para assegurar a confiabilidade dos dados capturados.

As frases experimentais foram exibidas no centro da tela do rastreador ocular, posição em que percebemos maior acurácia do ponto de fixação do olhar. As pequenas variações de captura, como se pode ver na Figura 7, são compensadas pelo desenho de regiões de interesse que se ajustam às variações no ângulo de captura dos olhos. A figura 7 é apenas ilustrativa, não representa a forma como os estímulos eram apresentados aos participantes.

Figura 7 – Exemplos de variação na posição de fixações oculares



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

Os materiais experimentais foram baseados nos estímulos produzidos por Carminati (2002), mas foram inteiramente reconstruídos e adaptados para a língua portuguesa. Com a intenção de neutralizar, ao máximo, o viés que pode ocasionar os aspectos de frequência, aplicamos vários controles para a construção dos estímulos experimentais. Em relação aos verbos da oração principal, privilegiamos os verbos bastante frequentes, cuja ação realizasse uma aproximação do sujeito e do objeto, uma espécie de ação a ser experienciada concomitantemente pelos dois argumentos verbais, como acontece com os verbos “conversar”, “falar”, “morar”, etc.

Em relação aos competidores, escolhemos nomes próprios com mais de um milhão de ocorrências no buscador Google, como pode ser visto na Tabela 14. Todos os nomes próprios das frases experimentais possuem cinco letras. As frases experimentais possuem em média 57 caracteres com espaço³⁴.

³⁴ A maior frase experimental tinha 63 caracteres com espaço. A menor frase continha 52 caracteres com espaço.

Tabela 14 – Controle de frequência dos nomes masculinos e femininos usados como competidores.

Resultados de ocorrências dos nomes próprios no buscador Google

Nomes masculinos		Nomes femininos					
Tomás	177 x10 ⁶	Jorge	319 x10 ⁶	Ester	50.7 x10 ⁶	Érica	78.1 x10 ⁶
Chico	136 x10 ⁶	César	116 x10 ⁶	Maria	945 x10 ⁶	Luísa	102 x10 ⁶
Tiago	30.6 x10 ⁶	Pedro	730 x10 ⁶	Joana	23.3 x10 ⁶	Júlia	281 x10 ⁶
Fábio	200 x10 ⁶	Lucas	230 x10 ⁶	Marta	246 x10 ⁶	Carla	115 x10 ⁶
Diego	496 x10 ⁶	Bruno	311 x10 ⁶	Paula	195 x10 ⁶	Sofia	141 x10 ⁶
Artur	52.2 x10 ⁶	Oscar	305 x10 ⁶	Alice	316 x10 ⁶	Vanda	15.2 x10 ⁶
Lúcio	57.7 x10 ⁶	André	302 x10 ⁶	Leila	68.1 x10 ⁶	Célia	9 x10 ⁶
Mário	482 x10 ⁶	David	151 x10 ⁷	Neide	3.69 x10 ⁶	Lígia	9.54 x10 ⁶

Os valores estão representados com o fator que deve ser multiplicados por 10⁶ (um milhão).

Todos os controles aplicados nos oferecem mais segurança para que a ambiguidade construída não pudesse ser resolvida em nenhuma fase da leitura, ou seja, para minimizar o efeito de qualquer pista, quer sintática, semântica ou pragmática, que viesasse os dados. Os exemplos (25), (26), (27) e (28), abaixo, representam um conjunto experimental. Cada uma destas frases era colocada em uma lista diferente para que o participante não lesse nenhuma frase repetida.

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

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

(q.1) Quem olhava pela janela?

O André - O David

6.2.4 Equipamento

Os movimentos oculares foram gravados no rastreador ocular de 120 Hz, da marca Tobii. O sistema faz uma amostragem e gravação dos movimentos e fixações das pupilas dos participantes enquanto escaneiam e lêem as sentenças, as questões e as alternativas de respostas. O rastreador ocular é embutido em uma tela de 17". Duas câmeras de alta velocidade fotografam os olhos em uma amostragem de 120 imagens por segundo para cada olho. O equipamento usado possui um software que compensa os movimentos leves de

cabeça, o que elimina a necessidade de um repouso para queixo e testa. Neste estudo, consideramos apenas os dados válidos para ambos os olhos, que foram posteriormente processados usando o software do Tobii, o programa Excel e o programa de estatística SPSS. O filtro utilizado para identificar uma fixação ocular foi configurado com o limiar máximo de 35 pixels de velocidade por 35 pixels de distância. Foram examinados todos os segmentos das frases, mas as regiões relativas aos competidores foram preferencialmente analisadas.

6.3 Variáveis dependentes

6.3.1 Performance comportamental

A resposta comportamental avaliada é uma medida *off-line*. Trata-se de um julgamento, feito após a leitura da frase, que procura verificar a compreensão do leitor. No caso, as respostas indicavam se o leitor havia estabelecido correferência com a entidade proeminente da oração principal ou com a entidade menos proeminente, o objeto da oração. Reportamos três diferentes visualizações dos resultados. Foi feita a soma dos resultados e foi calculado o percentual de atribuições para todas as condições na combinação de 2x2, mas também calculamos separadamente as condições de anáfora e catáfora e as condições de pleno e nulo.

6.3.2 Tempo total de leitura ou tempo de reação

Esta é a medida que representa o tempo total que o estímulo foi exibido para o participante até a sua decisão de clicar no mouse. A medida vai do tempo 0, correspondente ao *onset*, até o tempo F (final), correspondente ao último instante de exibição da imagem.

6.3.3 Tempo total de fixação nas regiões de interesse

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

6.3.4 Duração média de fixação nos segmentos do período complexo

A duração média de fixação é calculada fazendo-se a soma de todas as fixações e dividindo-a pelo número de fixações. Foram feitos dois cálculos de duração média da fixação. Em primeiro lugar, a medida foi calculada para as regiões dos competidores. Em segundo lugar, foi calculada a medida para as demais regiões do período, com o objetivo de observar se havia algum efeito tardio de processamento, possível de ser inferido a partir da medida de duração média da fixação.

6.4 Resultados

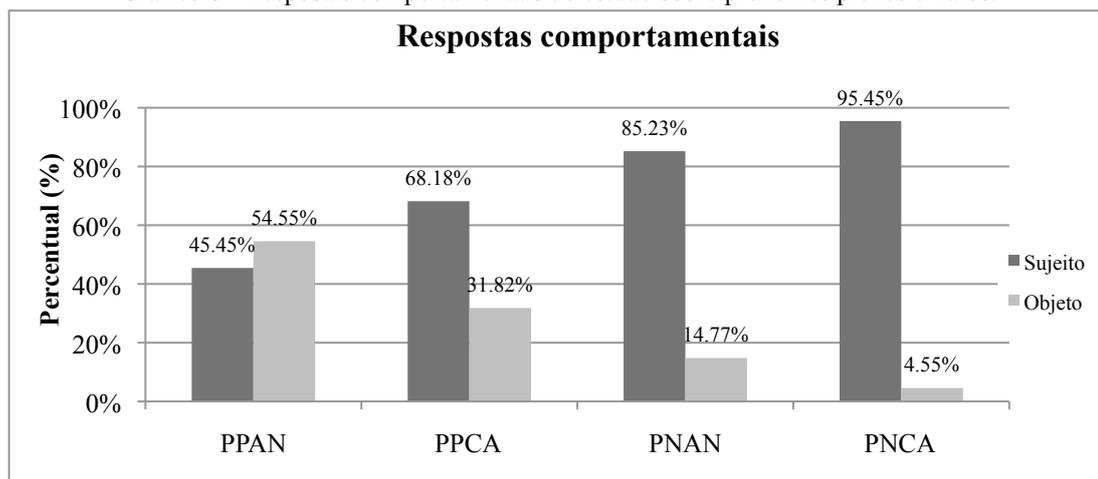
6.4.1 Performance comportamental

A correferência dos pronomes nulos e plenos foi estabelecida com o sujeito em posição canônica em todas as condições, como nos mostra o Gráfico 8, com exceção da condição de pronome pleno em posição anafórica (PPAN). Os dados são significativos para as condições de pronome nulo e pleno na posição catafórica e para as condições de pronome nulo na posição anafórica. Na condição PPAN, verifica-se que os valores encontrados giram em torno dos 50%, um resultado curioso que, isoladamente, não é muito indicativo, pois está no nível da aleatoriedade, mas que merece atenção porque sugere que o papel do pronome “ele” na anáfora não está claro para o falante do PB. Os testes estatísticos revelaram que há uma interação significativa entre a condição de pronome e a condição de posição ($F[1,42]=11,318$, $p=0,002$), assim como foi percebida significância quando rodamos a interação entre o tipo de competidor escolhido e o tipo de pronome ($F[1,42]=50,871$, $p=0,00$), resultado que pode ser observado no Gráfico 9. A observação das interações entre o tipo de competidor e a posição do elemento correferente se mostraram igualmente significativas ($F[1,42]=11,285$, $p=0,002$).

Os resultados comportamentais encontrados nos surpreenderam porque não estavam dentro das predições que havíamos feito, com especialização funcional do pronome nulo e pleno. Interessante é notar que, em posição catafórica, os participantes fizeram a correferência entre o pronome pleno e a posição de objeto em 30% dos casos. Este índice é bem mais alto do que os valores encontrados para o pronome nulo, que seleciona em 90,34% das vezes o sujeito como antecedente. O resultado do pronome nulo não surpreende e

confirma as predições feitas a partir da Hipótese da Posição do Antecedente. Já o papel do pronome pleno, embora observemos um percentual de 54% de correferência com o antecedente em posição de objeto, o maior entre todas as condições, não corresponde às predições feitas, mas se alinha à flexibilidade do pronome pleno observada por Carminati (2002).

Gráfico 8 - Respostas comportamentais do estudo sobre pronomes plenos e nulos.



O que parece bastante claro é que sob a condição de nulo os participantes claramente identificam o sujeito como antecedente, mas que o papel do pronome ainda parece que precisa se submeter a muitos outros testes para ser esclarecido. Os resultados sustentam a interpretação de que o pronome pleno não está apto para resolver a ambiguidade imposta nas orações por competidores do mesmo gênero, do mesmo número e nas posições sintáticas estudadas. É importante salientar que, se observadas as preferências entre todas as condições, o sujeito é preferido numa proporção de 3:1 em relação ao objeto, como pode-se observar no Gráfico 11, abaixo.

Gráfico 9 - Contraste entre pronome pleno e nulo

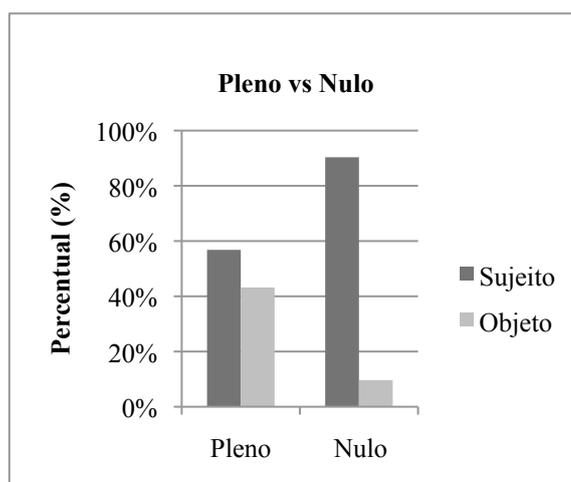


Gráfico 10 - Contraste entre anáfora e catáfora

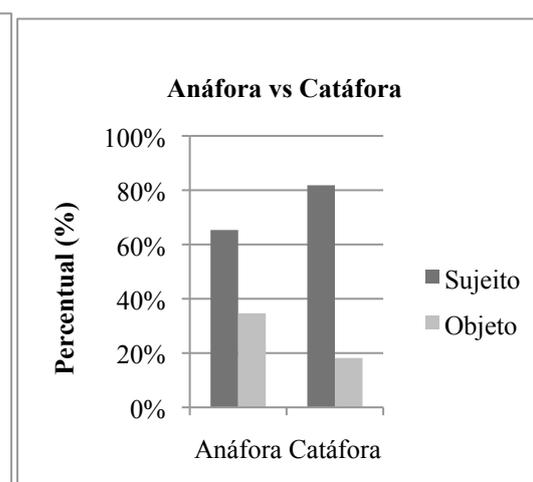
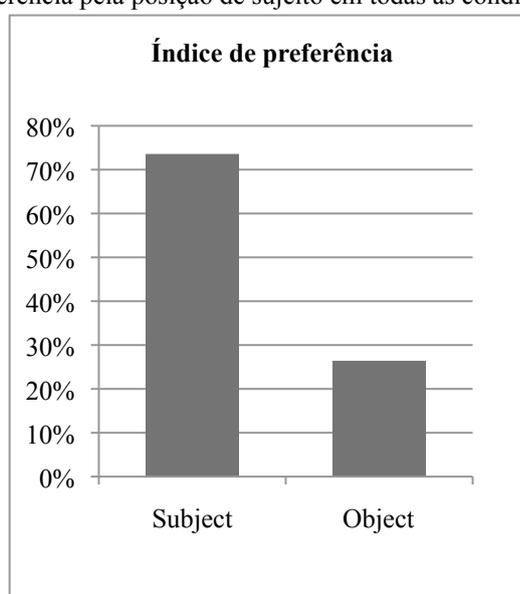


Gráfico 11 - Preferência pela posição de sujeito em todas as condições experimentais



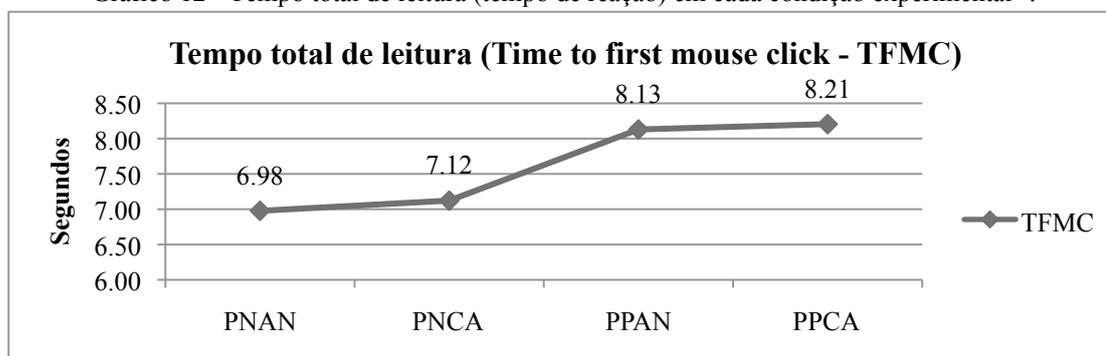
Como dito, o controle durante a produção dos períodos complexos, observando-se os papéis temáticos dos argumentos dos verbos, assim como o tamanho e a frequência dos competidores procurou afastar qualquer tendência para a correferência com a posição de sujeito. No entanto, o paralelismo das orações principal e subordinada, pode desempenhar um papel fundamental na atribuição da correferência, como discutiremos na seção 6.6 deste capítulo.

6.4.2 Movimentação ocular

Passemos agora a examinar os resultados encontrados no estudo da movimentação ocular dos participantes. Como apresentado na seção 6.3, iremos reportar os dados de três diferentes medidas das movimentações oculares.

A primeira medida, que apresentamos no Gráfico 12, é relativa ao tempo de reação, ou seja, ao tempo total de leitura dos períodos complexos. O Gráfico 12 apresenta um quadro geral que revela tempos de reação maiores para a leitura das condições com o pronome pleno em relação ao tempo de leitura das frases com pronome nulo. Comparando-se a posição do pronome pleno, percebe-se que há mais custo para a catáfora do que para a anáfora, assim como ocorre na condição do pronome nulo, já que registramos tempos maiores para a catáfora do que para a anáfora. Não é um resultado estranho pois a catáfora parece ser uma estratégia mais raramente utilizada, mesmo na língua escrita. Apesar de haver uma tendência, detectável pelo crescimento da curva da esquerda para a direita, a análise da variância não demonstra haver significância na interação das quatro condições ($F[3,19]=1,805$, $p=0,180$), assim como na interação de pares nas condições anáfora ($F[1,21]=2,837$, $p=0,107$) e catáfora ($F[1,21]=2,974$, $p=0,099$). Apesar de os resultados não serem significativos, alcançam um percentual que, para estudos do processamento da linguagem, poderíamos entender como marginalmente significativos, sobretudo se examinamos os pares isoladamente, pois a análise aponta uma variação em torno dos 10%. É provável que um aumento da amostra leve a índices de significância menores que 5%.

Gráfico 12 - Tempo total de leitura (tempo de reação) em cada condição experimental³⁵.



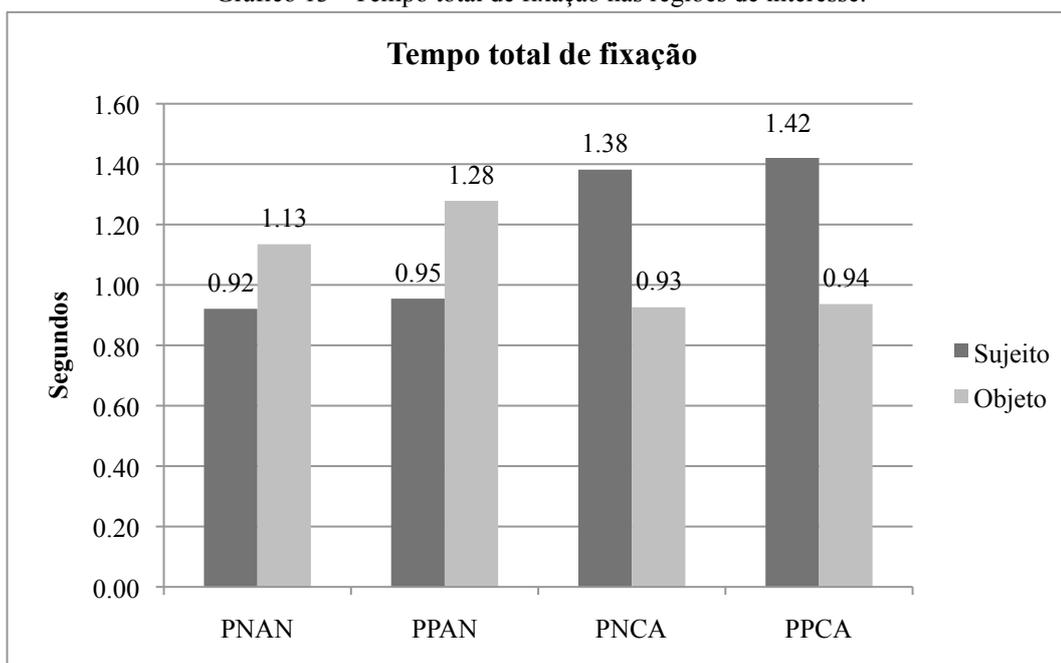
³⁵ TMFC é uma sigla em inglês correspondente a “time to first mouse click”. Este tempo é calculado subtraindo-se o tempo final (t_f), momento que o participante clica no mouse, do tempo inicial (t_i), relativo ao início da exibição do estímulo.

Quando examinamos o tempo total de fixação (GRÁFICO 13), encontramos efeitos significativos na interação entre a posição da expressão referencial (anáfora e catáfora) e as regiões dos competidores (sujeito e objeto), tanto na comparação das variáveis independentes da condição de pronome nulo ($F[1,20]=10,344$, $p=0,004$)³⁶, quanto da condição de pronome pleno ($F[1,21]=21,396$, $p=0,000$). O que se constata com esse resultado é que, na condição de pronome pleno anafórico, os participantes fixaram por mais tempo a região do objeto, o que talvez possa servir de indício para a explicação de maior atribuição de correferência ao objeto nesta mesma condição (PPAN). No entanto, apenas este dado não pode ser tomado como indicativo sem se considerar a análise conjunta dos demais tempos totais de fixação. Em linhas gerais, pode-se dizer que os participantes fixaram por mais tempo as regiões do sujeito na condição de catáfora, sendo o inverso constatado para a condição de anáfora.

O resultado acima descrito pode ser analisado de muitas maneiras. Em primeiro lugar, se constatamos mais tempo nas regiões do sujeito para a condição de catáfora, isto pode querer dizer que a catáfora impõe dificuldade na atribuição do antecedente tanto para um pronome nulo (1.38s) quanto para um pronome pleno (1.42s). Após a leitura das frases, como já ficou claro, os participantes precisavam optar entre o sujeito e o objeto. O que observamos na decisão posterior, *off-line*, é que o sujeito foi majoritariamente preferido (< 80%) nesta condição de catáfora (ver GRÁFICO 10). Na condição de anáfora, percebemos que o tempo total fixado na região do objeto foi relativamente maior do que na região do sujeito, embora os dados comportamentais não atestem para a condição anafórica a mesma correspondência que para a catáfora. Esta correspondência nem poderia ser atestada, uma vez que o nulo estabelece correferência com o sujeito. O que, sim, se verifica, é que o índice de correferência com o objeto é maior, considerada a média, para a condição de anáfora (34.66%) do que de catáfora (18.18%), na razão de 1.9:1 (ver GRÁFICO 10).

³⁶ Para a análise destes resultados, precisamos excluir os dados de um participante porque não foram encontrados registros de suas fixações nas regiões dos competidores. Portanto, a análise foi feita considerando-se 21 participantes, no lugar de 22 como descrito na seção 6.2.1.

Gráfico 13 - Tempo total de fixação nas regiões de interesse.

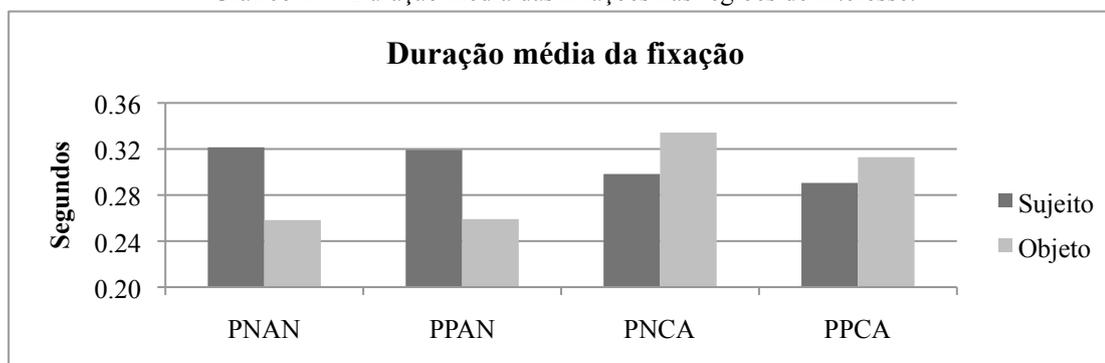


O terceiro dado de movimentação ocular que iremos analisar são os dados relativos à duração média de fixação. Rodamos a análise da variância (ANOVA) para observar a interação entre cada condição experimental e as regiões de interesse. Para a condição PNAN x Competidores, encontramos resultados significativos ($F[1,19]=13,529$, $p=0,002$)³⁷; para a condição PNCA x Competidores, não obtivemos resultados significativos ($p=0,242$); para a condição PPAN x Competidores, encontramos significância ($F[1,19]=14,176$, $p=0,001$); e para a condição PPCA x Competidores, não obtivemos resultados significativos ($p=0,336$). O quadro geral de resultados apresentado no Gráfico 14 nos leva a uma análise bastante difícil de ser elaborada, já que os resultados nos mostram proporções inversas, intuitivamente difíceis de serem compreendidas. É de se chamar atenção para o fato de que o tempo médio de fixação na região do objeto é menor (0.261s e 0.262s) do que na região do sujeito (0.331s e 0.325s), no entanto, o tempo total de fixação na região do objeto (1.13s e 1.28s) é maior do que na região do sujeito (0.92s e 0.95s). Os números interpretados nos podem sugerir que os participantes inspecionaram mais vezes a região do objeto, ou seja, fizeram mais visitas à região do objeto, mas se detiveram menos tempo inspecionando a região, ou seja, as visitas eram muitas mas breves. Já a região correspondente ao sujeito da oração foi por menos vezes visitada, mas as visitas foram mais longas do que as visitas à região do objeto. Isto explica porque, apesar de haver mais tempo total dispensado na

³⁷ Para este cálculo foram excluídos os dados de dois participantes, cujas médias estavam fora da curva da normalidade (gaussiana).

região do objeto, a correferência foi estabelecida com o sujeito, que está na região em que as visitas permaneceram por mais tempo, ainda que tenham sido mais escassas.

Gráfico 14 - Duração média das fixações nas regiões de interesse.



Nos Gráfico 15 e 16, apresentamos a duração média das fixações em toda a extensão do período complexo, com base na leitura de parte dos dados. Este gráfico foi elaborado com a intenção de se observar custos de processamento em outras regiões do período, que servissem como índice de processamento tardio ou antecipatório e também para comparar a leitura nas condições de pronome nulo e pleno.

Gráfico 15 - Duração média da fixação em cada segmento para a condição de Anáfora

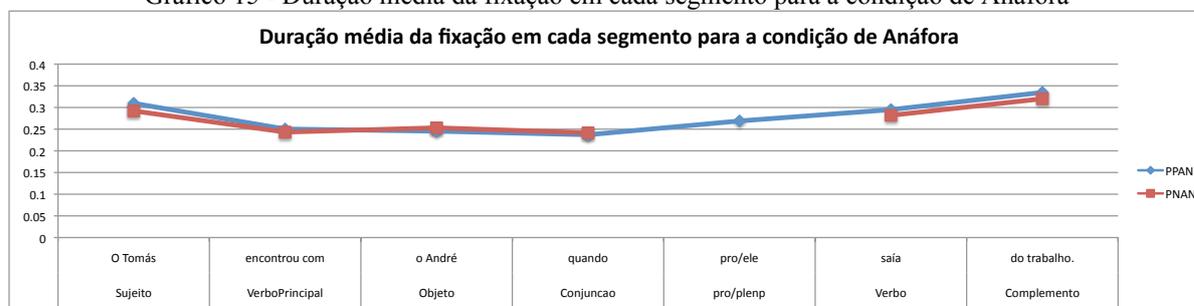
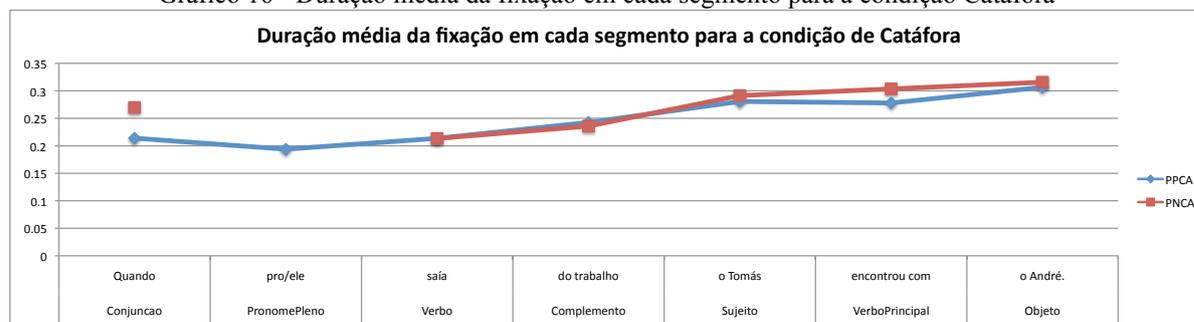


Gráfico 16 - Duração média da fixação em cada segmento para a condição Catáfora



Observamos que não há praticamente diferenças nos tempos médios de duração entre pronomes nulos e plenos quando se observa separadamente a condição de anáfora e catáfora. Estes dados, no entanto, não nos permitem afirmar que o processador se comporta igualmente em ambas as condições porque são apenas o tempo médio de fixação. Interessante seria a realização de testes ou a análise dos dados que separassem os tempos de primeira e segunda leituras, com o objetivo de detectar custos do processador durante a primeira leitura dos estímulos.

6.5 Sondagem sobre produção de períodos complexos com subordinada temporal

Em face dos resultados do estudo do processamento de frases, acima descrito, que sugerem falta de especialização para o pronome pleno em português e forte interferência da catáfora e anáfora no processamento do sistema pronominal do PB, decidimos ainda realizar um pequeno experimento, na verdade, mais propriamente uma pequena sondagem com o objetivo de identificar preliminarmente as decisões tomadas pelos falantes do Português Brasileiro ao produzir um período complexo em sua língua primeira.

A sondagem consistia na realização de uma tarefa de produção, que solicitava ao participante unir duas informações apresentadas em frases separadas, em apenas um período, que deveria ser complexo por subordinação e fazer uso de uma conjunção temporal.

A tarefa era propositadamente superespecificada, no sentido de limitar o participante a ter de escolher entre apenas quatro tipos diferentes de respostas potenciais. Estes quatro tipos potenciais estariam finamente ajustados ao conjunto experimental utilizado na tarefa de compreensão reportada neste capítulo.

Para realizar nosso objetivo, aplicamos o teste a 46 pessoas diferentes, 23 do sexo masculino e 23 do sexo feminino, falantes nativos do Português Brasileiro, com nível de instrução elevado, já que a maioria era formada por jornalistas, alunos ou professores de programas de pós-graduação de universidades brasileiras (pertencentes a variadas áreas de estudos).

Distribuímos as tarefas de produção em dois grupos. Cada grupo, composto por metade dos participantes, recebeu a tarefa de um só tipo, condição A ou B. Na condição A, o participante era obrigado a produzir uma correferência com o sujeito da oração principal e na condição B, a restrição lhe era imposta para produzir a correferência com o objeto da oração principal. Os participantes receberam as instruções por email, abaixo resumidas (ver

APÊNDICE C). Nas instruções estavam descritas todas as restrições para a produção do período. Os participantes enviaram suas respostas por email.

Instruções

Gostaria que você formasse UMA ÚNICA FRASE, obedecendo às seguintes restrições:

Os nomes próprios só podem ser mencionados uma única vez.

É preciso utilizar a conjunção enquanto.

É preciso utilizar a frase 1 sem modificações.

Frase 1 - O André conversou com o Tomás.

Frase 2 – [O André fazia café.]_{Condição A}/ [O Tomás fazia café.]_{Condição B}

Nossas predições sobre as respostas esperadas para as condições A e B nos diziam que os participantes poderiam optar por quatro variações. Em uma escala de preferência, acreditávamos que a estrutura catafórica, exemplificada em (29) e (30) seria a menos preferida.

- (29) O André conversou com o Tomás, enquanto fazia café.
- (30) O André conversou com o Tomás, enquanto ele fazia café.
- (31) Enquanto fazia café, o André conversou com o Tomás.
- (32) Enquanto ele fazia café, o André conversou com o Tomás.

Encontramos os resultados sumarizados na Tabela 15. Nas respostas para a condição A, a grande maioria dos participantes preferiu fazer uso da categoria vazia [*pro*] em 91% dos casos. A categoria vazia *pro* foi colocada em posição catafórica em 50% dos casos e em posição anafórica nos demais 41%. Apenas 9% dos participantes usaram um pronome ou item lexical para estabelecer correferência com o sujeito. Os termos usados na condição A foram “aquele” e “o primeiro”.

Tabela 15 - Preferências no estabelecimento da correferência em tarefa de produção

Condição A: Estabelecer correferência com o sujeito

	Preferências		
	Total	Anáfora	Catáfora
<i>pro</i>	90.91%	40.91%	50.00%
Pronome	9.09%	4.55%	4.55%

Condição B: Estabelecer correferência com o objeto

	Preferências		
	Total	Anáfora	Catáfora
Pronome	95.45%	95.45%	-
<i>pro</i>	4.55%	-	4.55%

Nas respostas para a condição B, 95% dos participantes fizeram uso de um pronome como expressão anafórica apropriada para correferir com o objeto da oração principal. Comparando-se estes resultados com a condição A, diríamos que o nulo está em distribuição complementar com um pronome (pleno ou outro), já que o nulo se especializa para correferir com o sujeito e o pronome com o objeto. No entanto, o pronome escolhido para estabelecer a correferência com a posição menos saliente da frase, o objeto, não foi o pronome “ele”, como havíamos em princípio hipotetizado. Em nossos dados, apenas 13% dos participantes fizeram uso do pronome pessoal “ele”. A maioria (82.45%) dos participantes produziram a frase fazendo uso do pronome demonstrativo “este” como expressão anafórica correferente ao objeto da oração principal. O pronome demonstrativo “este”, que possui traços semânticos e sintáticos de proximidade, de recência, se torna, no contexto escrito, o candidato ideal para desfazer a ambiguidade de casos como os testados por esta sondagem. O que resta a verificar é qual pronome, e se há um pronome, seria escolhido para estabelecer este tipo de correferência na produção oral em tarefas semelhantes a que propusemos nesta sondagem. Como dito, Duarte (1995) e outros pesquisadores, em estudos de *corpora* sobre o Português Brasileiro, já vêm observando a tendência para o preenchimento da posição de sujeito nulo com o uso do pronome pessoal “ele”. Esta diferença entre produção, escrita e oral, e compreensão, auditiva e leitora, pode ser uma chave que nos ajude a entender a distribuição do pronome “ele” e sua função na língua portuguesa, em todas as suas variantes, já que os estudos de compreensão leitora, isoladamente, não têm ainda aportado contribuições definitivas para esta questão.

A sondagem revelou escolhas em perfeita distribuição complementar, com margens de erro aceitáveis, da ordem de 5% e 9%. Mas o aporte mais importante desta sondagem está na força de seus resultados, que nos motiva a continuar investigando a aparente robustez da escolha do tipo de expressão referencial para cada condição testada em tarefas de produção. Acreditamos que subseqüentes testes de produção escrita e oral podem nos auxiliar a compreender melhor o sistema pronominal do Português Brasileiro.

6.6 Discussão

Os estudos, reportados neste capítulo, procuraram observar, em períodos complexos por subordinação, efeitos de posição (catáfora e anáfora) e de tipo de expressão anafórica (pronome nulo e pleno), em termos de custos de processamento e em termos de preferências no estabelecimento da correferência.

Assim como dito na seção 6.1, a hipótese de especialização das funções de um sistema pronominal composto por uma categoria vazia e um pronome pleno está ainda a ser confirmada para o PB. O que podemos dizer, a partir da análise de nossos dados, é que nosso trabalho de compreensão corrobora achados de pesquisas anteriores (LUEGI, 2012, MORGADO, 2011; MELO; MAIA; 2005; CORREA, 1998), que atribuem à expressão referencial, na forma de pronome nulo (a categoria vazia), a preferência no estabelecimento da correferência com o antecedente mais saliente, em posição de sujeito da oração.

Na análise dos resultados preliminares de nossa sondagem de produção, no entanto, sugerimos que há uma aparente distribuição complementar entre uma categoria vazia *pro* e um outro tipo de expressão anafórica pronominal, quer seja o pronome pessoal “ele”, quer seja qualquer outra forma pronominal, como o demonstrativo “este”. Pode ser que a categoria vazia *pro* seja uma categoria não marcada, enquanto um pronome lexicalizado se distribua em diferentes contextos e talvez de modo diferente nas quatro modalidades da linguagem.

A análise dos dados comportamentais também nos indica que é possível distinguir no sistema pronominal do PB uma escala de acessibilidade para os pronomes nulos e plenos. O pleno tenderia a se ligar a antecedentes menos acessíveis, enquanto o pronome nulo se ligaria àqueles mais salientes e acessíveis.

Nossa afirmação se baseia no fato de que os participantes claramente preferiram estabelecer correferência entre o pronome nulo e a posição do sujeito em qualquer das condições experimentais, com subordinadas em posição anafórica ou catafórica. Também constatamos que efeitos de recência e de paralelismo entre o pronome nulo e seu antecedente podem ser um fator importante a se considerar quando se pretende explicar a correferência do pronome nulo com o sujeito tanto em posição anafórica quanto catafórica.

Na resposta ao estudo comportamental, decidimos por não oferecer ao leitor a possibilidade de escolher um referente outro para o pronome pleno ou nulo, porque

observamos índices muito baixos ou nulos para estas opções nos resultados de Sorace e Filiaci (2006) e Fonseca e Guerreiro (2012), comentados na seção 6.1.

Diante dessa escolha forçada, por assim dizer, a análise dos nossos resultados comportamentais, obtidos no experimento *off-line* de identificação do antecedente do pronome nulo e pleno, nos sugere que a categoria vazia estudada seleciona de fato seu antecedente, em uma relação cuja força nos causa dúvida se é uma questão de preferência ou se é uma restrição imposta pelo sistema de regras do falante do português do Brasil. Uma vez constatada a regularidade, é de se supor que não se deveria falar de mudança de parâmetro de sujeito nulo em PB, já que a resposta comportamental é por demais robusta e clara, indicando a coerência que faz com que a categoria vazia preencha suas características semânticas e sintáticas com o antecedente em posição mais proeminente na oração principal ou, simplesmente, na posição estrutural de sujeito da oração.

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

A análise das medidas de duração média de fixação ocular sugeriram que os competidores escolhidos como correferentes são fixados por mais tempo ainda que por menos vezes. Observamos ainda que a posição produz um efeito nos custos de processamento, que são inversamente proporcionais para sujeito e objeto, com índices mais altos para o sujeito em anáfora e mais altos para o objeto em catáfora.

Os tempos totais de fixação comparados à duração média das fixações também sugerem que os participantes realizam menos visitas às regiões em que estabelecem correferência, porque a ambiguidade seria desfeita com mais antecedência do que nos casos em que realizam mais visitas à região dos competidores. No entanto, apenas um exame acurado em torno da comparação dos tempos de primeira e segunda leitura³⁸ nos poderia dar segurança para concluir a respeito do comportamento do *parser* nesta tarefa.

Nossa investigação observou ainda que os participantes reagiram à ambiguidade, fixando por mais tempo, durante seu processo de reanálise, as áreas da oração que competiam pelo estabelecimento da correferência. Mesmo a observação continuada dos papéis temáticos dos competidores pelos participantes não os levaram a escolher, na tarefa de compreensão,

³⁸ As medidas de primeira e segunda leituras precisam ser feitas por um software que não dispúnhamos na fase de captura e análise dos dados de movimentação ocular.

respostas coerentes que nos facultasse afirmar que pronome nulo e pleno exercem funções complementares em PB.

7 CONSIDERAÇÕES FINAIS

Esta tese procurou examinar dois tipos de preferência durante a resolução da correferência em Português Brasileiro. No primeiro estudo, examinamos preferências de ordem semântica, para procurar um efeito em relação à super ou sub especificação da expressão anafórica, quando retoma um antecedente acessível. No segundo estudo, observamos as preferências sintáticas, à procura de um efeito que uma ambiguidade poderia provocar. A ambiguidade examinada pode ser traduzida na competição entre dois antecedentes potenciais, que, segundo nossa hipótese, se refletiria em custos de processamento diferenciados na resolução anafórica entre o pronome pleno e a categoria vazia *pro*.

Os estudos abordaram dois tipos de estrutura, a estrutura coordenada, no exame das preferências semânticas, e a estrutura subordinada, no exame das preferências sintáticas. Retomaremos primeiramente a análise do primeiro estudo, discutindo os resultados em função das nossas hipóteses e em seguida discutiremos o segundo estudo, também confrontando nossas hipóteses com os dados coletados.

No estudo das preferências semânticas, que não deixava de estar correlacionado ao exame de preferências sintáticas, como pode ser observado no Capítulo 5, comparamos a correferência com antecedentes em duas posições sintáticas distintas, a de sujeito e de objeto. A primeira hipótese que formulamos dizia que encontraríamos custos menores para a correferência de expressões na posição sintática de sujeito.

O tempo total de fixação nas regiões das expressões anafóricas na condição de sujeito (2.5s) foram, de fato, menores do que os encontrados na posição de objeto (2.8 s). O número médio de fixações, uma medida relativamente correlata ao tempo total de fixação, também apresentou a mesma tendência, com um menor número de fixações para a condição de sujeito. No entanto, não pudemos fazer uma análise estatística entre as duas condições porque, como explicamos no Capítulo 5, realizamos o experimento em dois grupos distintos.

Consideramos, portanto, oportuno dizer que, em estudos posteriores que procurem os efeitos da posição estrutural na frase durante a resolução de expressões correferenciais com descrições definidas, seria a posição estrutural um relevante fator a (re)analisar, como já o fizeram Chambers e Smyth (1998) e Leitão (2005).

Dito de outro modo, se comparamos os custos entre a retomada de sujeito e a de objeto, observamos um aumento do custo de processamento para a condição de objeto,

contrariamente ao que projetamos em nossa segunda hipótese para este trabalho, que dizia não haver custos adicionais de processamento quando o objeto direto da primeira oração torna-se o centro catafórico da segunda oração coordenada.

Este resultado, ao nosso entender, corrobora a Teoria da Centralização, que estabelece que apenas um centro catafórico de um dado enunciado, será retomado pelo enunciado seguinte, transformando-se em um centro anafórico. Os centros anafóricos podem retomar tanto antecedentes na posição de sujeito quanto na posição de objeto, mas essa retomada se dará preferencialmente na posição de sujeito, já que retomar seria conferir ao antecedente estatuto de tópico discursivo.

A terceira hipótese que levantamos para o estudo das preferências semânticas em estruturas coordenadas dizia que há custos adicionais de processamento entre uma expressão anafórica que estabelece uma relação de hiponímia com seu antecedente quando comparado à relação de hiperonímia, hipótese que, se confirmada, corrobora os achados de Leitão (2005).

O teste de compreensão que avaliava o estabelecimento da correferência (seção 5.4.1) mostrou que os participantes a realizaram da forma esperada. No entanto, a correferência na relação de hiponímia apresentou índices mais baixos, chegando a uma diferença de 10% na condição de objeto. Esperávamos que a comparação desses valores não fosse significativa. Então, no que diz respeito a este dado, obtivemos o resultado esperado.

No entanto, queremos chamar atenção que, tanto na investigação desta tese quanto no estudo de Leitão (2005), os participantes estabeleceram correferência em um percentual menor na condição de hipônimo. Leitão (2005) encontrou valores na ordem dos 30%, ou seja, em torno de 1/3 dos participantes do estudo de Leitão (2005) não estabeleceram correferência entre a expressão anafórica em relação de hiponímia e um antecedente em posição de objeto da estrutura coordenada.

É provável que este fato indique, portanto, que a dificuldade de interpretação da correferência seja uma evidência de uma violação à hipótese da Teoria da Centralização sobre retomada do centro catafórico. Para não haver dúvidas, repetimos que a violação à qual nos referimos, não é a violação à regra 1 da Teoria da Centralização, mas à retomada de um centro catafórico por um centro anafórico em posição de objeto.

Os dados da movimentação ocular revelaram para esta terceira hipótese que a condição de hiponímia foi mais custosa do que a condição de hiperonímia, para a retomada em posição de sujeito e de objeto, confirmando os achados de Leitão (2005). Entretanto, como já foi dito no Capítulo 5, somente os dados do tempo total de leitura do período

complexo, isto é, o tempo total de exibição do slide, comprovam esta hipótese. As demais medidas, de duração média da fixação e de duração da primeira fixação, não revelaram diferenças significativas entre as condições testadas.

A duração média da fixação ocular aferida neste estudo revelou-se consistente em relação à duração média de fixação aferida durante a leitura de textos não manipulados. Isto nos leva a pelo menos duas conclusões: a primeira nos diz que nossos dados aferidos em estruturas manipuladas são consistentes com os aferidos em textos autênticos; e a segunda nos diz que esta consistência garante, em certa medida, a validade ecológica da nossa manipulação. No entanto, a hipótese nula precisaria ainda ser testada para sustentar a segunda conclusão.

A quarta hipótese levantada se referia ao paralelismo das estruturas testadas. Dissemos que o paralelismo seria um fator menos importante para a correferência do que a posição de sujeito. Como ficou claro, na explicação das demais hipóteses, o paralelismo das estruturas coordenadas testadas levou os participantes a estabelecerem a correferência em índices aceitáveis, maiores do que 87%. No entanto, os maiores custos encontrados para a posição de objeto e para a condição de hipônimo revelam uma variabilidade inter sujeito bastante grande para esta condição, evidenciada pelo desvio padrão da ordem de 50% (TABELA 4) nos tempos totais de leitura.

Acreditamos na força desta hipótese, mas queremos esclarecer que nosso estudo não pode ser conclusivo em relação a esta questão, porque as observações que fizemos são, de certo modo, indiretas, ou seja, nosso experimento não foi comparado com outras condições de não paralelismo que investigassem exaustivamente esta hipótese. Nossos resultados apenas apontam que os participantes estabeleceram correferência em estruturas paralelas, mas que a comparação entre a condição de sujeito e a de objeto revela um não paralelismo nos custos de processamento, fazendo-nos pensar que a posição sintática é mais forte na identificação de um antecedente do que o paralelismo das estruturas.

De uma maneira global, o estudo das preferências semânticas na retomada de antecedentes salientes pode ser compreendido com base na Teoria da Acessibilidade e na Hipótese da Carga Informacional.

Segundo a Teoria da Acessibilidade, uma expressão anafórica é marcada em função da alta acessibilidade do seu antecedente. Nas estruturas coordenadas que estudamos, é notória a alta acessibilidade do antecedente porque estamos tratando de retomada entre apenas duas orações coordenadas. No entanto, como seria previsível para estes casos, a

retomada não se deu por um pronome, mas por uma descrição definida. A descrição definida que retomava o antecedente adicionava informação nova nos casos de hiponímia (cachorro-labrador), enquanto nos casos de hiperonímia (cachorro-animal) apenas reativava os traços semânticos do nome já presentes na memória de trabalho dos participantes do teste.

A relação de hiperonímia pode ser interpretada pelo leitor como não preferida, uma vez que encontramos diferença significativa, com maiores custos para o hiperônimo (0.239s), na condição de sujeito, nos dados de movimentação ocular relativos à duração média da fixação. O hipônimo cuja média é inferior nesta medida (0.192s), apesar de adicionar informação nova, aportando mais traços semânticos, pode ter sido mais rapidamente interpretado pelo leitor porque se constitui uma expressão anafórica com justificativa funcional, segundo as predições da Hipótese da Carga Informacional.

No estudo das preferências sintáticas, em que observamos os custos dos pronomes nulo e pleno em posição catafórica ou anafórica, estava subjacente a idéia de que a forma da expressão anafórica (pronominal) possuía uma relação direta com a acessibilidade do referente. Como tratamos de um estudo que examinou períodos complexos por subordinação composto de duas orações, o antecedente satisfazia as condições para ser retomado na forma pronominal. Os antecedentes potenciais estavam sob a forma de nomes próprios em sintagmas nominais definidos. Segundo a hierarquia de acessibilidade de Ariel (1991), a retomada deste tipo de antecedente é preferencialmente realizada por uma expressão sem conteúdo semântico, ou seja, por um pronome. O pronome nestas estruturas também satisfazia o Princípio B da Teoria da Ligação, que diz que deve ser livre no seu domínio local. A posição de Spec IP do sujeito da oração principal das estruturas estudadas, segundo a Hipótese da Posição do Antecedente, seria preenchida pelo antecedente preferencialmente escolhido para estabelecer correferência com o pronome nulo.

A primeira hipótese levantada para o estudo das preferências sintáticas estava ligada à Hipótese da Posição do Antecedente, acima definida. Os dados comportamentais de nosso estudo corroboram fortemente esta hipótese para o Português Brasileiro. Nossos participantes realizaram em 90% dos casos a correferência com o antecedente em posição de Spec IP. Desse modo, corroboramos os achados e as conjecturas anteriores feitas para o Português Brasileiro e Europeu (CORREA, 1998; COSTA; FARIA; MATOS, 1998; FONSECA; GUERREIRO, 2012; LUEGI, 2012; MORGADO, 2011; MELO; MAIA, 2005, entre outros). Em relação aos dados de movimentação ocular, encontramos custos menores para as condições testadas de pronome nulo tanto em posição anafórica quanto catafórica. No

entanto, o exame detalhado das medidas de tempo total de fixação e duração média de fixação na região relativa ao sujeito da oração principal (posição de Spec IP) nos revela um dado em princípio contraditório.

Encontramos que cada fixação ocular do participante na região do sujeito foi mais longa, em torno de 0.32 s, mas a soma de todas as fixações revela que o olhar permaneceu menos tempo nesta região, em torno de 0.92 s. Dizemos, portanto, que os participantes realizaram em média 3 visitas à região do sujeito durante a leitura das frases. Já na região do objeto, cada fixação ocular foi mais curta, em torno de 0.26s. Acontece que os participantes passaram mais tempo nesta região, já que a soma do tempo de visitas foi de 1.13 s. Dividindo-se a soma pela média, dizemos que os participantes visitaram a região do objeto um pouco mais de 4 vezes.

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

A segunda hipótese levantada para o estudo da preferência sintática dizia que o paralelismo estrutural das orações seria um fator relativamente mais fraco do que a posição sintática de sujeito para o estabelecimento da correferência. Diante do estudo que conduzimos, não podemos assegurar qualquer conclusão sobre esta hipótese, mas podemos fazer uma observação. Na condição de correferência entre o pronome pleno e um dos dois competidores, em posição anafórica, obtivemos nas respostas comportamentais dados que revelam a não capacidade dos participantes de decidir se o pronome pleno era correferente com o antecedente sujeito ou objeto. Os resultados que obtivemos foram aleatórios. Obtivemos para a correferência do pronome pleno e do sujeito um índice de 45% e para a correspondência com o objeto, um índice de 55%. Nos casos em que os participantes estabeleceram correferência com o objeto não havia paralelismo, já que o pronome pleno se encontrava em posição de sujeito da oração subordinada.

Como o estabelecimento da correferência entre pronome pleno e o sujeito da oração principal não foi aleatório, pois quase 70% dos participantes atribuíram o pleno à

posição de sujeito, poderíamos dizer que nestas condições o paralelismo foi mais forte do que a hipótese de que o pleno preferiria a posição de objeto para estabelecer correferência. No entanto, o que o dado para a condição de catáfora pode nos revelar é que pode ter havido um efeito de recência, pois o pronome pleno foi correferenciado com o primeiro sintagma nominal potencialmente antecedente. O efeito de recência está ligado à capacidade de armazenamento de informação da memória de trabalho, o que nos leva a indagar se estaríamos falando de paralelismo ou de restrições de ordem não linguística, mas cognitivas.

A terceira hipótese levantada para o estudo dos períodos complexos por subordinação dizia que poderia haver custos adicionais de processamento para a correferência entre pronome pleno e objeto da oração principal. Hipótese esta, estreitamente relacionada à nossa quarta hipótese, baseada em Carminati (2002), de que a categoria vazia *pro* e o pronome pleno do Português Brasileiro estariam em distribuição complementar, desempenhando, funções anafóricas especializadas.

Como os dados comportamentais não revelaram uma especialização de funções entre o pleno e o nulo, não obtivemos medidas de movimentação oculares que nos permitissem uma análise da nossa terceira hipótese. A quarta hipótese de distribuição complementar não foi confirmada, o que corrobora os achados de Carminati (2002) que apontam para um comportamento de certo modo flexível do pronome pleno, que parece depender do contexto para determinar com que posição estrutural da frase deve se ligar.

Os dados encontrados em nosso estudo estão alinhados aos dados encontrados em trabalhos de tese recentemente concluídas sobre o Português Europeu e o Português Brasileiro (LUEGI, 2012; MORGADO 2011). Os trabalhos de Luegi (2012) e Morgado (2011), que investigaram a resolução pronominal anafórica sob variáveis diferentes das nossas, encontraram ampla preferência na correferência entre o nulo e o sujeito da oração principal. No entanto, Luegi (2012), em seu trabalho que procurou verificar a interferência de outros fatores para a preferência pelo sujeito da oração, como a posição estrutural na frase, invertendo a posição canônica Sujeito – Verbo – Objeto, afirma que diante de “dois potenciais antecedentes para a expressão anafórica pronominal, a proeminência resulta da combinação de fatores como a função sintática e a posição estrutural de cada entidade”.

Como não pudemos, através do estudo de compreensão leitora, confirmar a hipótese de que haveria na língua uma especialização das funções do pronome nulo e pleno, e como ainda acreditamos que é preciso investigar este objeto sob diferentes métodos, realizamos uma sondagem para saber quais as escolhas que os falantes de Português

Brasileiro fariam em uma tarefa de produção de enunciados. A sondagem nos revelou que a correferência com o sujeito foi preferencialmente estabelecida com o pronome nulo, variando igualmente em posição catafórica ou anafórica em relação ao antecedente.

A tarefa de produção ainda nos revelou que a correferência com o objeto foi realizada preferencialmente fazendo-se uso do pronome demonstrativo “este”, que servia ao propósito de desfazer a ambiguidade sugerida pela tarefa. Sugerimos que estudos posteriores devam ser conduzidos para avaliar os resultados preliminarmente encontrados nesta sondagem e para tentar compreender melhor as razões que levam uma língua a possuir, em seu inventário, dois tipos de pronomes que aparentemente exercem as mesmas funções.

Finalmente, queremos esclarecer que os resultados que encontramos, nos experimentos aqui reportados, e que o modo como os interpretamos nos permitem confirmar a hipótese geral desta tese, que dizia que, em períodos complexos, o sintagma nominal que ocupa a função de sujeito da oração principal (ou primeira, no caso das coordenadas) é o antecedente preferencialmente retomado para estabelecer correferência com a primeira expressão referencial anafórica do período.

Para encerrar queremos ainda lembrar e deixar bem claro que, como conhecemos as limitações de nossos estudos e como compreendemos a imensa variabilidade de situações de comunicação permitidas pela linguagem humana, acreditamos que a preferência na resolução anafórica pelo estabelecimento da correferência entre antecedentes em posição de sujeito da oração e sintagmas nominais definidos, pronomes plenos ou nulos, ainda deva ser exaustivamente testada para que se confirme nossa hipótese.

REFERÊNCIAS

- ALBRECHT, J. E.; CLIFTON, C. K. Accessing singular antecedents in conjoined phrases. *Memory and Cognition*, n. 26, p. 599-610, 1998.
- ALLOPENNA, P. D., MAGNUSON, J. S., TANENHAUS, M. K. Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language*, v. 38, p. 419-439, 1998.
- ALMOR, A. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, v. 106, p. 748-765, 1999.
- ALONSO-OVALLE, L., CLIFTON, C., FRAZIER, L., SOLERA, S. F. Null vs. Overt Pronouns and The Topic-Focus Articulation in Spanish. *Journal of Italian Linguistics*, v. 14, n. 2, p. 151-169, 2002.
- ANDERSON, A., GARROD, A.C., SANFORD, A.J. The accessibility of pronominal antecedents as a function of episode shifts in narrative texts. *Quarterly Journal of Psychology*, v. 35A, p. 427-440, 1983.
- ARIEL, M. *Givenness marking*. 1985. Dissertation (Ph.D. in Linguistics) - Tel-Aviv University, Tel-Aviv, 1985.
- ARIEL, M. Referring and accessibility. *Journal of Linguistics*, v. 24, p. 65-87, 1988.
- ARIEL, M. *Accessing Noun-Phrase Antecedents*. London: Routledge, 1990.
- ARIEL, M. The function of accessibility in a theory of grammar. *Journal of Pragmatics*, v. 16, n. 5, p. 443-464, 1991.
- ARIEL, M. Referring expressions and the +/- coreference distinction. In: GUNDEL, J.; FRETHEIM, T. (Eds.) *Referent and Referent accessibility*. Amsterdam: John Benjamins, 1996. p. 13-35.
- ARIEL, M. Accessibility theory: An overview. In: SANDERS, T.; SCHLIPEROORD, J.; SPOOREN, W. (Ed.) *Text representation: linguistics and psycholinguistics aspects*. Amsterdam: John Benjamins, 2001. p 29-87.
- ARNOLD, J. E., EISENBAND, J. G., BROWN-SCHMIDT, S., TRUESWELL, J. C. The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*. v. 76, n. 1, p. B13-26, 2000.
- APOTHÉLOZ, D.; CHANET, C. Défini et démonstratif dans les nominalisations. In: MULDER, W.; RYCK, L.T.; VETTERS, C. (Eds.) *Relations anaphoriques et (in)cohérence*. Amsterdam: Rodopi, 1997, p. 159-86.
- APOTHÉLOZ, D. Référent sans expression référentielle: gestion de la référence et opérations de reformulation dans des séquences métalinguistiques produites dans une tâche de rédaction conversationnelle. In: NÉMETH, E. (Ed.) *Pragmatics in 2000: selected papers from the 7th*

International Pragmatics Conference. Antwerp: International Pragmatics Association, 2001, p.30-38.

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

BATTIG, W. G.; MONTAGUE, W. E. Category norms for verbal items in 56 categories: a replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, v. 80, n. 3, p. 1-46, 1969.

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

CARREIRAS, M.; CLIFTON JR., C. (Ed.). *The on-line study of sentence comprehension: eye-tracker, ERPs and beyond*. New York: Psychology Press, 2004.

CAVALCANTE, M. M. Expressões referenciais: uma proposta classificatória. *Cadernos de Estudos Linguísticos*, Campinas, v. 44, p. 105-118, jan./jun., 2003.

CAVALCANTE, M. M. Processos de referenciação: uma revisão classificatória. Comunicação apresentada no XIX ENANPOLL. Alagoas: UFAL, 2004.

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

CHOMSKY, N. *Language and problems of knowledge: the Managua lectures*. Cambridge: MIT Press, 1988.

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

CHOMSKY, N. *Knowledge of language: its nature, origin and use*. New York: Praeger Publishers, 1986.

CIULLA, A. *A referenciação anafórica e dêitica – com atenção especial para os dêiticos discursivos*. 2002. Dissertação (Mestrado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, Fortaleza, 2002.

CIULLA, A. *Os processos de referência e suas funções discursivas - o universo literário dos contos*. 2008. Tese (Doutorado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, Fortaleza, 2008.

COOPER, R. M. The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, v. 6, p. 61-83, 1974.

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

- COSTA, A.; LUEGI, P. Complexidade linguística e processamento referencial. *In: CONGRESSO INTERNACIONAL DA ABRALIN, VI, 2009, João Pessoa, Anais do VI Congresso Internacional da ABRALIN, João Pessoa: UFPB, 2009. p. 2127-2135. CD-ROM.*
- COSTA, A., FARIA, I. H., MATOS, G. Ambiguidade referencial na identificação do Sujeito em estruturas coordenadas. *In: ENCONTRO NACIONAL DA ASSOCIAÇÃO PORTUGUESA DE LINGUÍSTICA, XIII, 1997, Lisboa, Atas do XIII Encontro Nacional da Associação Portuguesa de Linguística. Lisboa: Colibri, 1998. p. 173-188.*
- CRAWLEY, R. A., STEVENSON, R. J., KLEINMAN, D. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, v. 19, p. 245-264, 1990.*
- DELEUZE, G. *Empirisme et subjectivité: essai sur la nature humaine selon Hume.* Paris: Press Universitaires de France, 1953.
- DERRIDA, J. *De la grammatologie.* Paris: Les Éditions de Minuit, 1967.
- DUARTE, M. E. L. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português brasileiro. *In: ROBERTS, I.; KATO, M. A. (Org.). Português Brasileiro: uma viagem diacrônica.* Campinas: Ed. Da Unicamp, 1993. p. 107-128.
- DUARTE, M. E. L. A perda do Princípio "Evite Pronome" no Português Brasileiro. 1995. Tese (Doutorado em Estudos da Linguagem) – Instituto de Estudos da Linguagem, Universidade de Campinas, Campinas, 1995.
- FOLEY, W. A.; VAN VALIN, R. D. *Functional Syntax and Universal Grammar.* Cambridge: Cambridge University Press, 1984.
- FONSECA, M. C. M.; GUERREIRO, E. Resolução de Correferência Pronominal no Português do Brasil. *Revista Linguística, v. 8, n. 2, p. 112-133, 2012.*
- FOUCAULT, M. *Les Mots et les Choses: une archéologie des sciences humaines.* Paris: Gallimard, 1966.
- GARNHAM, A. *Mental models and the interpretation of anaphora.* Hove: Psychology Press, 2001.
- GARROD, S.; SANFORD, A. Interpreting anaphoric relations: the integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior, v. 16, p. 77-90, 1977.*
- GARVEY, C.; CARAMAZZA, A.; YATES, J. Factors affecting assignment of pronoun antecedents. *Cognition, v. 3, p. 227-243, 1976.*
- GERNSBACHER, M. A. Mechanisms that improve referential access. *Cognition, v. 32, n. 2, p. 99-156, 1989.*
- GIVÓN, T. *Syntax and Semantics 12: discourse and syntax.* New York: Academic Press, 1979.

- GIVÓN, T. *Topic continuity in discourse: a quantitative cross-language study*. Typological Studies in Language 3. Amsterdam: John Benjamins, 1983.
- GORDON, P.; CAMBLIN, C.; SWAAB, T. On-line measures of coreferential processing. *In: CARREIRAS, M.; CLIFTON JR., C. The on-line study of sentence comprehension: eye-tracker, ERPs and beyond*. New York: Psychology Press, 2004.
- GORDON, P. C.; HENDRICK, R. Intuitive knowledge of linguistic co-reference. *Cognition*, v. 62, n. 3, p. 325-370, 1997.
- GORDON, P. C.; GROSZ, B. J.; GILLIOM, L. A. Pronouns, names, and the centering of attention. *Cognitive Science*, v. 17, n. 3, p. 311-347, 1993.
- GORDON, P. C., SCEARCE, K. A. Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory and Cognition*, v. 23, n. 3, p. 313-323, 1995.
- GRICE, H. P. Logic and conversation. *In: COLE, P.; MORGAN, J. (Eds.). Syntax and semantics 3: speech acts*. New York: Academic Press, 1975. p. 41-58.
- GRODZINSKY, Y.; REINHART, T. The Innateness of Binding and Coreference. *Linguistic Inquiry*, v. 24, n. 1, p. 69-101, 1993.
- GROSZ, B. J.; SIDNER, C. L. Attentions, intentions and the structure of discourse. *Computational Linguistics*, v. 12, p. 175-204, 1986.
- GROSZ, B. J., JOSHI, A.K., WEINSTEIN, S. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, v. 21, n. 2, p. 203-225, 1995.
- HALLIDAY, M. A. K.; HASAN, R. *Cohesion in English*. London: Longman, 1976.
- HAUSER, M. D.; CHOMSKY, N.; FITCH, W. T. The Faculty of Language: what is it, who has it, and how did it evolve? *Science*, v. 298, n. 5598, p. 1569-1579, 2002.
- HUANG, Y. *Anaphora in Chinese: toward a pragmatic analysis*. 1989. Dissertation (PhD in Linguistics) – Trinity College, University of Cambridge, Cambridge, 1989.
- JUST, M. C.; CARPENTER, P. A. A Theory of Reading: from eye fixations to comprehension. *Psychological Review*, v. 87, n. 4, p. 329-354, 1980.
- KAUP, B.; KELTER, S.; HABEL, C. Representing referents of plural expressions and resolving plural anaphors. *Language and Cognitive Processes*, v. 17, n. 4, p. 405-450, 2002.
- KINTSCH, W.; VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychological Review*, v. 85, n. 5, p. 363-394, 1978.
- KOCH, I. G. V. *Introdução à Linguística Textual: trajetória e grandes temas*. São Paulo: Martins Fontes, 2006.

- KOCH, I. G. V.; MARCUSCHI, L. A. Processos de referenciação na produção discursiva. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, v. 14, p.169-190, 1998.
- KOH, S.; CLIFTON, C. A. Resolution of the antecedent of a plural pronoun: ontological categories and predicate symmetry. *Journal of Memory and Language*, v. 46, p. 830-844, 2002.
- LEITÃO, M. M. Psicolinguística Experimental: focalizando o processamento da linguagem. In: MARTELLOTA, M. et al. (Orgs.) *Manual de Lingüística*. São Paulo: Contexto, 2008. p. 217-234.
- LEITÃO, M. M. Processamento co-referencial de nomes e pronomes em Português Brasileiro. *Revista Linguística*, v. 1, n. 2, p. 235-258, 2005.
- LEITÃO, M. M. *O processamento do objeto direto anafórico em Português Brasileiro*. 2005. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.
- LI, C. N.; THOMPSON, S. A. Third-person pronouns and zero-anaphora in Chinese discourse. In: GIVÓN, T. *Syntax and Semantics 12: discourse and syntax*. New York: Academic Press, 1979. p. 311-335.
- LUEGI, P.; COSTA, A.; FARIA, I. Mover para ler: o movimento dos olhos durante a leitura de textos. In: LOBO, M; COUTINHO, M. A. *Textos selecionados do XXII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Colibri, 2007. p. 431-445.
- LUEGI, P. *Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes*. 2012. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 2012.
- MAIA, M. The processing of Object Anaphora in Brazilian Portuguese. *Recherches Linguistiques de Vincennes*, v. 26, p.151-172, 1997.
- MAIA, M.; GARCIA, D. C.; OLIVEIRA, C. The processing of conceptual anaphors and fully specified pronouns in intra-sentential contexts in Brazilian Portuguese. *Revista Virtual de Estudos da Linguagem (REVEL)*, v. 10, n. 6, edição especial, p. 200 – 219, 2012.
- MARCUSCHI, L. A. O léxico: lista, rede ou cognição social? In: FOLTRAN, M. J (Org.). *Sentido e significação: em torno da obra de Rodolfo Ilari*. São Paulo: Contexto, 2004. p. 263-84.
- MELO, M. F. B.; MAIA, M. A. R. O processamento da correferência do sujeito pronominal em sentenças formadas por verbos de comunicação lingüística no português do Brasil. *Revista Linguística*, v. 1, n. 2, p. 177-205, 2005.
- MERIDOR, H. *An Experimental Investigation of the Antecedent Preferences of Hebrew Subject Pronouns*. 2006. Master of Science (Master on Cognitive Science and Natural Language Processing) – School of Informatics, Edinburg University, Edinburg, 2006.

MONDADA, L. e DUBOIS, D. Construction des objets de discours et catégorisation: une approche des processus de référenciation. *TRANEL: Travaux Neuchâtelois de Linguistique*, v. 23, p. 273-302, 1995.

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

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

MYERS, J. L.; O'BRIEN, E. J. Accessing the discourse representation during reading. *Discourse Processes*, v. 26, p. 131-157, 1998.

NAND, P. *Resolving co-reference anaphora using semantic constraints*. 2012. Dissertation (PhD in Computer and Mathematical Sciences) – School of Computer and Mathematical Sciences, Auckland University of Technology, Auckland, 2012.

O'BRIEN, E. J.; ALBRECHT, J. E. Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning Memory and Cognition*, v. 18, p. 777-784, 1992.

O'BRIEN, E. J.; ALBRECHT, J. E.; HAKALA, C. M.; RIZZELLA, M L. Activation and suppression of antecedents during reinstatement. *Journal of Experimental Psychology: Learning Memory and Cognition*, v. 21, p. 626-634, 1995.

O'BRIEN, E. J.; RANEY, G. E.; ALBRECHT, J. E.; RAYNER, K. Processes involved in the resolution of explicit anaphors. *Discourse Processes*, v. 23, p. 1–24, 1997.

OCH, F. J. Minimum error rate in training in statistical machine translation. In: ANNUAL MEETING ON ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 41, 2003, Sapporo, *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, Stroudsburg: Association for Computational Linguistics, v. 1, 2003. p. 160-167.

PIRON, C. *O Desafio das Línguas*. Campinas: Pontes, 2002.

QUARESMA, P.; PINHO, A. Análise de frequências da língua portuguesa. In: Conferência Ibero-Americana InterTIC, 2007, Porto, *Livro de Actas da Conferência Ibero-Americana InterTIC*, Porto: IASK, 2007. p. 267-272.

RAPOSO, E. P. *Teoria da Gramática: a Faculdade da Linguagem*. Lisboa: Caminho, 1992.

RAYNER, K. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, v. 124, n. 3, p. 372-422, 1998.

RAYNER, K.; KAMBE, G.; DUFFY, S. A. The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology*, v. 53A, p. 1061-1080, 2000.

RAYNER, K.; POLLATSEK, A. *The psychology of reading*. Englewood Cliffs: Prentice Hall, 1989.

REINHART, T. *Anaphora and semantic interpretation*. Chicago: Chicago University Press, 1983.

REULAND, E. Anaphoric dependencies: a window into the architecture of the language system. *Glott International*, v.7, p. 3-25, 2003.

REULAND, E. Primitives of Binding. *Linguistic Inquiry*, v. 32, n. 3, p. 439-492, 2001.

ROSCH, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, v. 104, p. 192-233, 1975.

SANFORD, A. J.; GARROD, S. C. *Understanding written language*. Chichester: Wiley, 1981.

SANFORD, A. J., LOCKHART, F. Description types and method of conjoining as factors influencing plural anaphora: a continuation study of focus. *Journal of Semantics*, v. 7, p. 365-378, 1990.

SANFORD, A.; STURT, P.; MOXEY, L.; MORROW, L.; EMMOTT, C. Production and Comprehension Measures in Assessing Plural Object Formation. In: CARREIRAS, M.; CLIFTON JR., C. *The on-line study of sentence comprehension: eye-tracker, ERPs and beyond*. New York: Psychology Press, 2004.

SMYTH, R. H. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, v. 23, p. 197-229, 1994.

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

STAUB, A.; RAYNEY, K. Eye movements and on-line comprehension processes. In: GASKELL, G. (Ed.). *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, 2007. p. 327-342.

TANENHAUS, M. K.; SPIVEY-KNOWLTON, M. J.; EBERHARD, K. M.; SEDIVY, J. C. Integration of visual and linguistic information in spoken language comprehension. *Science*, v. 268, n. 5217, p. 1632-1634, 1995.

TREIMAN, R. Reading. In: ARONOFF, M; REES-MILLER, J. (Eds.). *Blackwell Handbook of Linguistics*. Oxford: Blackwell, 2001. p. 664-672.

TRUESWELL, J. C.; TANENHAUS, M. K.; GARNSEY, S. M. Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, v. 33, p. 285-318, 1994.

VAN GOMPEL, R. P. G.; LIVERSEDGE, S. P.; PEARSON, J. Antecedent typicality effects in the processing of noun phrase anaphors. In: CARREIRAS, M.; CLIFTON JR., C. *The on-*

line study of sentence comprehension: eye-tracker, ERPs and beyond. New York: Psychology Press, 2004. p. 119-138.

VAN RULLEN, R.; THORPE, S. J. Is it a bird? Is it a plane? Ultra-rapid visual categorization of natural and artifactual objects. *Perception*, v. 30, p. 665-668, 2001.

WITTGENSTEIN, L. *Philosophical Investigations*. New York: Macmillan, 1953.

WITTGENSTEIN, L. *On certainty*. Oxford: Blackwell, 1969.

APÊNDICE A – ESTÍMULOS DO ESTUDO DE TEXTOS NÃO MANIPULADOS

Texto 1

Número de palavras: 569

Uma ecologia espiritual (Folha de São Paulo, 15/08/2010) por Marcelo Gleiser

O respeito à vida como verdade universal leva a um estado em que agimos como os guardiões dela

Está na hora de irmos em frente e deixar para trás o desgastado embate entre a ciência e a religião, que já não rende nada. É preciso encontrarmos um novo rumo, ir além da polarização linear que vem caracterizando as discussões do papel da fé e da razão na vida das pessoas por mais de cem anos. A ciência não se propõe a roubar Deus das pessoas, e nem toda prática religiosa é anticientífica. Existe uma outra dimensão a ser explorada, ortogonal a esse eixo em torno do qual giram os argumentos mais comuns.

Um caminho possível é explorar valores morais de caráter universal que desafiem a linearidade do cabo de guerra entre a ciência e a religião. Bem sei que, para muita gente, a proposta de encontrar valores morais universais representa já um beco sem saída. Relativistas culturais, por exemplo, argumentarão que esses valores universais não existem, que o que é certo para um pode ser errado para outro. Por exemplo, culturas nas quais a poligamia é aceita.

Para encontrar valores morais universais, precisamos ir mais fundo. Não podem ser valores que variem de cultura para cultura ou em épocas diferentes, como a ideia do casamento. Sugiro que o valor mais efetivo que podemos explorar vem da única certeza universal que temos: a morte.

A morte não é recebida com prazer em nenhuma cultura. Claro, alguns veem a morte como uma transição para uma nova vida, ou um mero aspecto de uma existência sem fim. Outros podem até vê-la como um ato heroico de martírio. Mas, tirando fundamentalistas radicais, ninguém em boa saúde física e mental escolhe morrer. Portanto, de todos os valores morais que podemos imaginar, proponho que o mais universal seja a preservação da vida.

Não me refiro apenas à vida humana. Quando percebemos o quanto nossas vidas dependem do planeta que habitamos, damos-nos conta de que precisamos agir para preservar todas as formas de vida. É óbvio que temos que garantir nossa existência, e que isso requer que consumamos alimentos. Mas esse consumo não precisa ser predatório. Pode ser planejado para que mantenha um equilíbrio saudável entre o que é produzido e o que é consumido.

Quanto mais saudável o planeta, mais saudável a economia. Isso pode não ser óbvio a curto prazo, mas em intervalos de décadas é. Este é o século em que finalmente iremos entender que precisamos estabelecer uma relação simbiótica com a Terra. Talvez essa seja a lição mais importante que a ciência moderna tem a ensinar.

O respeito à vida como moral universal leva a uma ecologia espiritual na qual nós, como espécie dominante do planeta, agimos como guardiões da vida. Com isso, a dimensão espiritual que nos é tão importante ganha expressão na devoção ao planeta e às suas formas de vida. Esse senso de conexão espiritual com a natureza é celebrado tanto na ciência quanto na religião. De Einstein a Santa Teresa de Ávila (grato a Frei Betto, por me chamar atenção para esta obra), o mundo é festejado como sacro. As palavras variam, mesmo a motivação pode variar; mas, em sua essência, a mensagem é a mesma. Acho difícil encontrar uma moral universal mais básica do que o respeito à vida e ao planeta que a abriga de forma tão generosa. Ao menos, é um começo.

Texto 2**Número de palavras: 527****O erro de Kepler (Folha de São Paulo, 01/08/2010) por Marcelo Gleiser**

Devemos julgar afirmações sobre "teorias de tudo" com enorme ceticismo; nosso conhecimento é limitado

Em 1596, com o furor de uma mente devota, o jovem Johannes Kepler, então com apenas 25 anos, publica seu primeiro livro, "Mysterium Cosmographicum" ou "O Mistério Cosmográfico". Nele, o astrônomo principiante propõe nada menos do que a solução para a estrutura do Cosmo, o que acreditava ser o plano divino da Criação.

Tudo se deu durante uma aula que ministrava para um punhado de estudantes desinteressados. Quando explicava as conjunções dos planetas Júpiter e Saturno, Kepler se perguntou se o fato de Saturno estar aproximadamente duas vezes mais longe do Sol do que Júpiter era sintoma de uma ordem mais profunda: talvez a estrutura cósmica seguisse as regras da geometria. Fosse esse o caso, a mente humana teria acesso aos segredos mais profundos da Criação e à mente de Deus.

E a língua em comum entre homem e Deus seria a matemática. Após várias tentativas frustradas, Kepler obteve a solução que tanto almejava. Na época, só eram conhecidos seis planetas, de Mercúrio a Saturno. Urano e Netuno, invisíveis aos olhos, só foram descobertos bem mais tarde. Kepler, numa visão genial, imaginou que o cosmo seria organizado a partir dos cinco sólidos platônicos, os cinco objetos mais simétricos que existem em três dimensões. Conhecemos bem dois deles, o cubo e a pirâmide (tetraedro). Kepler entendeu que, ao colocar um sólido dentro do outro, como aquelas bonecas russas, com esferas entre cada um deles, poderia acomodar apenas seis planetas: Sol no centro; esfera (Mercúrio); sólido; esfera (Vênus); sólido; esfera (Terra); sólido etc. Portanto, o número de planetas seria decorrente do número de sólidos perfeitos!

Kepler foi além. Como os sólidos obedecem às regras da geometria, seu arranjo determina também as distâncias entre si e, portanto, entre as esferas que os cercam. Experimentando com padrões diferentes, Kepler encontrou um que previa as distâncias entre os planetas com uma precisão de 5% -quando comparado com os dados astronômicos da época, um feito sensacional.

Para um homem que acreditava profundamente num Deus matemático, criador da ordem cósmica, nada mais natural do que uma solução geométrica. Kepler via seu arranjo como a expressão do sonho pitagórico de obter uma explicação geométrica para os mistérios do mundo. Para ele, essa era a teoria final.

Podemos aprender algo com Kepler. Soubesse ele da existência de outros planetas, Urano e Netuno, como teria reagido? Certamente, seu sonho de uma ordem geométrica para o Cosmo dependia do que se sabia na época. Seu erro foi ter dado ao estado do conhecimento empírico do mundo uma finalidade que não existe. Para Johannes Kepler, era inimaginável que o Cosmo pudesse se desviar de sua estrutura geométrica. No entanto, sabemos que nosso conhecimento do mundo é limitado, e será sempre.

Por isso, devemos julgar declarações sobre teorias de tudo ou teorias finais com enorme ceticismo. A história nos ensina que o progresso científico caminha de mãos dadas com nossa habilidade de medir a Natureza. Achar que a mente humana pode imaginar o mundo antes de medi-lo pode ocasionalmente dar certo. Mas, em geral, leva a mundos que existem apenas na imaginação.

Texto 3**Número de palavras: 579****Outra conversa sobre o tempo (Folha de São Paulo, 22/08/2010) por Marcelo Gleiser**

Extrapolando a expansão do Universo até seu início, existe um ponto em que não há como definir o tempo

Agora que o livro do Zuenir Ventura e do Luís Fernando Veríssimo, "Conversa sobre o tempo", está nas livrarias, não consegui resistir. Peguei emprestado o título para termos aqui um outro tipo de conversa sobre o mesmo tópico.

O tempo significa muitas coisas diferentes. E, por falar nisso, adianto que o próximo livro da série "Conversa sobre..." será uma conversa minha com o precioso Frei Betto, mediada por Waldemar Falcão. O tema será "Conversa sobre a fé". Mas isso é coisa para o final do ano.

Enquanto esse livro não chega, gostaria hoje de retomar um tema científico, sobre a origem do tempo. Uma das consequências mais diretas do Big Bang é que o Universo teve origem em um instante específico do passado. Isso se deu há cerca de 13,7 bilhões de anos. Uma das indicações mais óbvias disso é a expansão do Universo: o fato de as galáxias estarem se afastando umas das outras.

Portanto, passando o filme ao contrário, chegamos em um instante em que todas elas estão comprimidas em um único ponto. Esse é o momento da criação. E, portanto, o momento em que surge o tempo.

O problema é que essa descrição não funciona. Infelizmente, ao nos aproximarmos desse momento crítico, a teoria que usamos para descrever a expansão do espaço (a teoria da relatividade geral de Einstein) deixa de fazer sentido.

Chegamos à "singularidade", onde toda a matéria estaria comprimida em uma região de proporções não tão diferentes de um átomo. Com isso, a teoria de Einstein, que trata do espaço e do tempo como entidades contínuas e bem comportadas, precisa ser suplantada por conceitos da física quântica, que trata dos átomos e das partículas elementares da matéria. Aí a coisa fica feia.

Na teoria de Einstein, a gravidade é descrita como consequência da curvatura do espaço. A presença de uma massa, seja ela o Sol, você ou uma bola de tênis, deforma o espaço ao redor e afeta a passagem do tempo. Quanto mais matéria, maior a curvatura do espaço e mais lenta a passagem do tempo. Um relógio no Sol bate mais devagar do que na Terra. Os efeitos são bem pequenos.

Quanto tentamos "quantizar" a gravidade, temos de supor que, tal como no caso dos átomos, as mesmas estranhas regras se aplicam: no mundo do muito pequeno, tudo flutua, nada fica parado. Se você imaginar o espaço como uma membrana, feito o topo de um tambor, isso significa que ele vibrará de várias formas e o que ocorre aqui não é o que ocorre ali. O mesmo com o tempo. Ele não flui mais continuamente. Como a era quântica do Universo veio antes da era clássica (explicada pela teoria de Einstein), temos de supor que, nessa situação inicial onde tudo flutuava, o tempo usual não existia. Se extrapolarmos a expansão do Universo até o seu início, chegamos a um ponto em que não podemos definir o tempo de modo familiar.

Aliás, como disse já Santo Agostinho, o tempo e o espaço surgem com a criação. Na física, o tempo e o espaço einsteinianos, contínuos e bem comportados, surgem na transição da era quântica à era clássica. E como o Big Bang é precisamente o evento que marca a passagem do universo da era quântica para a era clássica, é ele também que marca o nascimento do tempo.

Texto 4**Número de palavras: 159****Café pode prevenir câncer de cabeça e pescoço (Veja, 16/07/2010)**

Pessoas que consomem café regularmente podem ser protegidas contra o câncer de cabeça e pescoço, segundo uma revisão publicada na revista *Cancer Epidemiology, Biomarkers & Prevention*. A partir de uma análise de nove estudos científicos, que envolveu mais de 5.139 pessoas com câncer de cabeça e pescoço e 9.028 sem a doença – concluiu-se que aqueles que bebem quatro xícaras de café por dia têm 39% menos chance de desenvolver a doença em comparação com aqueles que não bebiam. Chás ricos em cafeína não foram associados à prevenção desse tipo de câncer. Além disso, os pesquisadores não encontraram indícios em relação ao câncer de laringe. “Levando em conta o alto nível de consumo do café e a baixo índice de sobrevivência de pessoas com câncer de cabeça e pescoço, nossos resultados têm pontos importantes para a saúde pública”, disse a pesquisadora Mia Hashibe, do departamento de medicina preventiva da Universidade de Utah.

APÊNDICE B – ESTÍMULOS DO ESTUDO SOBRE AS PREFERÊNCIAS SEMÂNTICAS

Grupo 1

Código HISU - Hipônimo/Sujeito

HISU1 O felino fugiu da jaula no zoológico e depois o tigre violentamente atacou os visitantes estrangeiros.

HISU2 O réptil se aproximou da margem do rio e depois o jacaré instintivamente construiu o ninho com a vegetação.

HISU3 O cachorro brincou com a bola no gramado e depois o labrador alegremente tomou banho de mangueira.

HISU4 O macaco subiu na árvore mais próxima e depois o chimpanzé avidamente comeu os frutos maduros.

HISU5 O cavalo correu pela pista do hipódromo e depois o alazão espontaneamente voltou para a cocheira principal.

HISU6 O tubarão assustou os banhistas na praia e depois o cação repentinamente desapareceu no mar bravio.

HISU7 O médico examinou a jovem gestante e depois o obstetra imediatamente receitou alguns medicamentos holopáticos.

HISU8 O fruto caiu da árvore ainda verde e depois o abacate vagorosamente amadureceu no chão do sítio.

HISU9 O carro ultrapassou o sinal vermelho e depois o Meriva imprudentemente atropelou os pedestres na faixa.

HISU10O telefone vibrou sobre o tampo da mesa e depois o celular acidentalmente caiu no chão da sala.

Código SUSU - Superordenado/Sujeito

SUSU10 felino fugiu da jaula no zoológico e depois o bicho violentamente atacou os visitantes estrangeiros.

SUSU20 réptil se aproximou da margem do rio e depois o animal instintivamente construiu o ninho com a vegetação.

SUSU3O cachorro brincou com a bola no gramado e depois o animal alegremente tomou banho de mangueira.

SUSU4O macaco subiu na árvore mais próxima e depois o primata avidamente comeu os frutos maduros.

SUSU5O cavalo correu pela pista do hipódromo e depois o quadrúpede espontaneamente voltou para a cocheira principal.

SUSU6O tubarão assustou os banhistas na praia e depois o peixe repentinamente desapareceu no mar bravio.

SUSU7O médico examinou a jovem gestante e depois o profissional imediatamente receitou alguns medicamentos halopáticos.

SUSU8O fruto caiu da árvore ainda verde e depois o alimento vagorosamente amadureceu no chão do sítio.

SUSU9O carro ultrapassou o sinal vermelho e depois o veículo imprudentemente atropelou os pedestres na faixa.

SUSU10O telefone vibrou sobre o tampo da mesa e depois o aparelho acidentalmente caiu no chão da sala.

Grupo 2

Código HIOB - Hipônimo/Objeto

HIOB1- Os caçadores capturaram um felino na floresta mas depois libertaram o guepardo no parque.

HIOB2- As crianças provaram um legume na feira mas depois só comeram o pepino no jantar.

HIOB3- Os vizinhos adquiriram um carro na loja mas depois venderam o Meriva no feirão.

HIOB4- Os biólogos avistaram um réptil no rio mas depois assustaram o jacaré na margem.

HIOB5- Os amigos compraram um enlatado no mercado mas depois esqueceram o azeite no balcão.

HIOB6- Os meninos perderam um telefone na escola mas depois encontraram o celular no pátio.

HIOB7- Os empresários menosprezaram um operário na entrevista mas depois contrataram o pedreiro no dia seguinte.

HIOB8- Os garimpeiros acharam um metal no barranco mas depois jogaram o manganês no riacho.

HIOB9- Os irmãos ganharam um jogo no aniversário mas depois quebraram o dominó no natal.

HIOB10- Os ladrões seqüestraram um metalúrgico no banco mas depois soltaram o soldador na estrada.

Código SUOB - Superordenado/Objeto

SUOB1- Os caçadores capturaram um felino na floresta mas depois libertaram o bichano no parque.

SUOB2- As crianças provaram um legume na feira mas depois só comeram o vegetal no jantar.

SUOB3- Os vizinhos adquiriram um carro na loja mas depois venderam o veículo no feirão.

SUOB4- Os biólogos avistaram um réptil no rio mas depois assustaram o animal na margem.

SUOB5- Os amigos compraram um enlatado no mercado mas depois esqueceram o alimento no balcão.

SUOB6- Os meninos perderam um telefone na escola mas depois encontraram o aparelho no pátio.

SUOB7- Os empresários menosprezaram um operário na entrevista mas depois contrataram o trabalhador no dia seguinte.

SUOB8- Os garimpeiros acharam um metal no barranco mas depois jogaram o minério no riacho.

SUOB9- Os irmãos ganharam um jogo no aniversário mas depois quebraram o brinquedo no natal.

SUOB10- Os ladrões seqüestraram um metalúrgico no banco mas depois soltaram o profissional na estrada.

Condjunto de distratores

DISTR1- O núcleo de ecologistas do Rio Grande do Sul lançou uma nova proposta.

DISTR2- A sopa de amendoim é a novidade do verão no restaurante Bom Garfo.

DISTR3- O DEM e o partido popular que está no poder na Espanha se uniram.

DISTR4- Desfiles femininos em São Paulo apostam em visual vitoriano para o outono.

DISTR5- Uma assinante de TV está indignada com o tratamento que vem recebendo.

- DISTR6- Quem viu a estréia do Big Brother não imagina o sufoco passado pelos parentes.
- DISTR7- Quem vai do Rio ou de São Paulo para Angra desembolsa uma boa grana.
- DISTR8- A Secretaria do Meio Ambiente concluiu a licitação de vários quiosques.
- DISTR9- A música eletrônica tomou conta do Rio no último festival que aconteceu.
- DISTR10- O lançamento do livro de Jô Soares gerou uma fila enorme de pessoas.
- DISTR11- A possibilidade de clonagem humana tem levantado polêmica entre os médicos.
- DISTR12- O território disputado por palestinos e por israelenses já fez inúmeras mortes.
- DISTR13- A leitura de um bom livro pode render uma imaginação fértil por toda a vida.
- DISTR14- O tratamento para o câncer tem evoluído tanto que subiu o número de curas.
- DISTR15- As universidades brasileiras estão melhorando com o investimento do governo.
- DISTR16- Promover os autores novos que surgem por todo o Brasil é dever das editoras.
- DISTR17- Relatório da ONU responsabilizou a atividade humana pelo aquecimento global.
- DISTR18- Em visita ao Brasil, o papa Bento XVI condenou o uso de métodos contraceptivos.
- DISTR19- Parcerias internacionais ajudam o instituto Butantan em São Paulo.
- DISTR20- Ativistas defendiam o sacrifício do filhote de urso polar do zoológico de Berlim.
- DISTR21- A justiça italiana responsabilizou o diretor da Williams pela morte de Ayrton Senna.
- DISTR22- Atentados violentos em Bagdá expõem a falência do Estado iraquiano.
- DISTR23- Suspeitas de fraudes apagam o brilho do programa Luz para Todos.
- DISTR24- Gêmeos idênticos foram considerados pelo sistema de cotas como branco e negro.
- DISTR25- A cirurgia de transplante de retina consegue conter a degeneração macular.
- DISTR26- O crescimento econômico da China é sustentado pela indústria de falsificação.
- DISTR27- A imprensa tem cobrado explicações de autoridades sobre a crise dos aeroportos.
- DISTR28- De suas celas, líderes de facções criminosas continuam no comando do crime.
- DISTR29- Fidel Castro renunciou à presidência de Cuba após 50 anos no poder.
- DISTR30- Grupo de investidores pernambucanos aplica princípios empresariais na educação.
- DISTR31- A Colômbia mostrou-se determinada a acabar com o narcotráfico e o terrorismo.
- DISTR32- A confirmação da epidemia de dengue no Rio faz repelente sumir das prateleiras.
- DISTR33- Brasileiros com dívida em cheque especial pagarão juros ainda maior este ano.
- DISTR34- Escolas municipais antecipam o início das aulas para concluir mais cedo o semestre letivo.
- DISTR35- Governo Federal anuncia abertura de vagas para servidores técnico-administrativos.

DISTR36- Europeus procuram trabalho em países emergentes com o objetivo de fugir de crise econômica.

DISTR37- Os gastos dos brasileiros com saúde ultrapassaram os investimentos feitos pelo governo.

DISTR38- Empresas aéreas divulgaram que deve haver reajuste nos valores das passagens.

DISTR39- Naufrágio do navio italiano representa uma ameaça ecológica se houver vazamento de combustível.

DISTR40- Uso continuado de protetor solar evita o surgimento de manchas e diminui risco de câncer de pele.

APÊNDICE C – ESTÍMULOS DO ESTUDO SOBRE AS PREFERÊNCIAS SINTÁTICAS

Frases experimentais

Lista	Codigo	Frase
1	PPCA1	Enquanto ele fazia o café, o Tomás conversou com o Chico.
1	PPCA5	Quando ele cruzava o pátio, o Tiago acenou para o Fábio.
1	PPCA9	Enquanto ela estendia a roupa, a Ester brigou com a Maria.
1	PPCA13	Quando ela praticava yoga, a Joana conheceu a Marta.
1	PNCA4	Enquanto estava doente, o Diego passeou com o Artur.
1	PNCA8	Quando vivia em Recife, o Lúcio trabalhou com o Mário.
1	PNCA12	Enquanto escutava o rádio, a Paula chamou a Alice.
1	PNCA16	Quando saía do trabalho, a Leila encontrou com a Neide.
1	PPAN3	O Jorge telefonou para o César, enquanto ele viajava pelo país.
1	PPAN7	O Pedro jogou com o Lucas, quando ele estava de férias.
1	PPAN11	A Érica falou com a Luísa, enquanto ela cozinhava o arroz.
1	PPAN15	A Júlia morou com a Carla, quando ela estudava no Rio.
1	PNAN2	O Bruno saudou o Oscar, enquanto tocava uma música.
1	PNAN6	O André reconheceu o David, quando olhava pela janela.
1	PNAN10	A Sofia gritou com a Vanda, enquanto limpava a sala.
1	PNAN14	A Célia avistou a Lígia, quando caminhava pela rua.
2	PPCA2	Enquanto ele tocava uma música, o Bruno saudou o Oscar.
2	PPCA6	Quando ele olhava pela janela, o André reconheceu o David.
2	PPCA10	Enquanto ela limpava a sala, a Sofia gritou com a Vanda.
2	PPCA14	Quando ela caminhava pela rua, a Célia avistou a Lígia.
2	PNCA1	Enquanto fazia o café, o Tomás conversou com o Chico.
2	PNCA5	Quando cruzava o pátio, o Tiago acenou para o Fábio.
2	PNCA9	Enquanto estendia a roupa, a Ester brigou com a Maria.
2	PNCA13	Quando praticava yoga, a Joana conheceu a Marta.
2	PPAN4	O Diego passeou com o Artur, enquanto ele estava doente.
2	PPAN8	O Lúcio trabalhou com o Mário, quando ele vivia em Recife.
2	PPAN12	A Paula chamou a Alice, enquanto ela escutava o rádio.
2	PPAN16	A Leila encontrou com a Neide, quando ela saía do trabalho.

2	PNAN3	O Jorge telefonou para o César, enquanto viajava pelo país.
2	PNAN7	O Pedro jogou com o Lucas, quando estava de férias.
2	PNAN11	A Érica falou com a Luísa, enquanto cozinhava o arroz.
2	PNAN15	A Júlia morou com a Carla, quando estudava no Rio.
3	PPCA3	Enquanto ele viajava pelo país, o Jorge telefonou para o César.
3	PPCA7	Quando ele estava de férias, o Pedro jogou com o Lucas.
3	PPCA11	Enquanto ela cozinhava o arroz, a Érica falou com a Luísa.
3	PPCA15	Quando ela estudava no Rio, a Júlia morou com a Carla.
3	PNCA2	Enquanto tocava uma música, o Bruno saudou o Oscar.
3	PNCA6	Quando olhava pela janela, o André reconheceu o David.
3	PNCA10	Enquanto limpava a sala, a Sofia gritou com a Vanda.
3	PNCA14	Quando caminhava pela rua, a Célia avistou a Lígia.
3	PPAN1	O Tomás conversou com o Chico, enquanto ele fazia o café.
3	PPAN5	O Tiago acenou para o Fábio, quando ele cruzava o pátio.
3	PPAN9	A Ester brigou com a Maria, enquanto ela estendia a roupa.
3	PPAN13	A Joana conheceu a Marta, quando ela praticava yoga.
3	PNAN4	O Diego passeou com o Artur, enquanto estava doente.
3	PNAN8	O Lúcio trabalhou com o Mário, quando vivia em Recife.
3	PNAN12	A Paula chamou a Alice, enquanto escutava o rádio.
3	PNAN16	A Leila encontrou com a Neide, quando saía do trabalho.
4	PPCA4	Enquanto ele estava doente, o Diego passeou com o Artur.
4	PPCA8	Quando ele vivia em Recife, o Lúcio trabalhou com o Mário.
4	PPCA12	Enquanto ela escutava o rádio, a Paula chamou a Alice.
4	PPCA16	Quando ela saía do trabalho, a Leila encontrou com a Neide.
4	PNCA3	Enquanto viajava pelo país, o Jorge telefonou para o César.
4	PNCA7	Quando estava de férias, o Pedro jogou com o Lucas.
4	PNCA11	Enquanto cozinhava o arroz, a Érica falou com a Luísa.
4	PNCA15	Quando estudava no Rio, a Júlia morou com a Carla.
4	PPAN2	O Bruno saudou o Oscar, enquanto ele tocava uma música.
4	PPAN6	O André reconheceu o David, quando ele olhava pela janela.
4	PPAN10	A Sofia gritou com a Vanda, enquanto ela limpava a sala.
4	PPAN14	A Célia avistou a Lígia, quando ela caminhava pela rua.
4	PNAN1	O Tomás conversou com o Chico, enquanto fazia o café.

- 4 PNAN5 O Tiago acenou para o Fábio, quando cruzava o pátio.
 4 PNAN9 A Ester brigou com a Maria, enquanto estendia a roupa.
 4 PNAN13 A Joana conheceu a Marta, quando praticava yoga.

Frases distratoras

Código	Distratoras
DISTR1	Enquanto lavava a louça, a Diana pediu mais sabão ao Eduardo.
DISTR2	Quando ele caiu, o Ricardo socorreu o Dalton.
DISTR3	Quando ela pintou o cabelo, a Elaine cumprimentou a Tânia.
DISTR4	Quando ele mentiu para todos, o advogado não confiou mais no cliente.
DISTR5	A Flora aconselhou o filho porque ele se apaixonou de repente.
DISTR6	Como ela bateu a porta, o João brigou com a Joana.
DISTR7	Já que ele perdeu tudo, o Edson ajudou o colega.
DISTR8	Quando se separou, o pai apoiou a filha.
DISTR9	Quando ele acordou, o Alexandre estava em pé ao lado do Sandro.
DISTR10	Enquanto ela fazia a prova, a aluna fofocava com a colega.
DISTR11	Quando estava triste, Haroldo acordava Telma.
DISTR12	Quando ela foi embora, a Antonia sentiu falta da Júlia.
DISTR13	Quando ele se feriu, o Sávio socorreu o Marcos.
DISTR14	Enquanto estava só, o Nestor telefonava para a namorada.
DISTR15	Quando quebrou a perna, a Maíra carregava o Bruno.
DISTR16	Enquanto tocava piano, o Lucas elogiou a Maísa.
DISTR17	Enquanto ela dirigia, a Selma recebeu o presente da Meire.
DISTR18	Quando ela estava no hospital, a Luana levou flores para a Dilma.
DISTR19	Quando ela passava mal, a Lúcia socorreu a Deise.
DISTR20	Quando ele viu o escorpião, o Denis alertou o Paulo.
DISTR21	Desde que comprou um maiô novo, Ana vai à praia com Heitor.
DISTR22	Desde que virou chefe, Rodolfo não almoça mais com Liana.
DISTR23	Embora ela esteja de carro novo, Armanda leva a amiga ao trabalho.
DISTR24	Apesar de estar doente, o Pablo se divertiu com o amigo.
DISTR25	Apesar de ter sido aprovado, o aluno continuou a trabalhar para o engenheiro.
DISTR26	À medida que ele escrevia o livro, o autor entregava os capítulos ao editor.
DISTR27	Como ele se tornou presidente, o Joaquim não cumprimenta mais o secretário.

- DISTR28 Como ele queria comprar bombons, o menino deu dois reais ao amigo.
- DISTR29 Como ela era uma boa amiga, a Sandra pintava as unhas da Letícia.
- DISTR30 Porque ele cresceu rápido, Arnaldo doou seus sapatos para Estevão.
- DISTR31 Porque ele era um bom jogador, o técnico discutiu com o atacante.
- DISTR32 Como ele perdeu o telefone, José e Antonio não puderam se encontrar.
- DISTR33 Rômulo não queria ver seu pai, já que não passou de ano.
- DISTR34 Bruna pediu ajuda a Beatriz, porque estava sem dinheiro.
- DISTR35 Denis foi ao cinema com Emília, embora estivesse cheio de trabalho.
- DISTR36 Jefferson não recebeu sua prima, apesar de muita insistência.
- DISTR37 Mercedes detestava a Lia porque ela nunca comprava sua comida favorita.
- DISTR38 Osmar perdeu do Afonso, porque era um excelente tenista.
- DISTR39 Marisa levou Rodrigo para viajar, já que estava desempregado.
- DISTR40 Álvaro não devolveu os livros ao Carlos, porque os esqueceu em cima da mesa.
- DISTR41 Douglas criticava seu empregado, já que era lento no trabalho.
- DISTR42 Margarida deu um carro ao Fernando apesar de ser menor de idade.
- DISTR43 Sérgio dedicou seu livro ao Flávio, já que recebeu muitas contribuições relevantes.
- DISTR44 Helena ainda saiu para jantar com o Tarcísio apesar de estar farta.
- DISTR45 Ângela gostava de seu vizinho porque era uma boa pessoa.
- DISTR46 Nelson amava Lourdinha apesar de não ser correspondido.
- DISTR47 Guilherme contratou Ronaldo mesmo que ele estivesse muito gordo.
- DISTR48 Virgínia mandou lembranças para Rosa, apesar de ser sua concorrente.

APÊNDICE D – DADOS REFERENTES ÀS MEDIAS DO ESTUDO SOBRE AS PREFERÊNCIAS SEMÂNTICAS

Dados do estudo das preferências semânticas

Grupo 1 - Antecedente e retomada em posição de sujeito

Região da Anáfora

Condição Hiperônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
1	SUSU10	P01	0.233	0.175	2
1	SUSU10	P05	0.067	0.071	2
1	SUSU10	P09	0.259	0.171	2
1	SUSU10	P13	0.258	0.258	1
1	SUSU10	P20	0.308	0.308	1
1	SUSU7	P01	99.000	99.000	99
1	SUSU7	P05	0.158	0.121	4
1	SUSU7	P09	0.216	0.216	1
1	SUSU7	P13	0.308	0.308	1
1	SUSU7	P20	99.000	99.000	99
1	SUSU6	P01	0.250	0.169	3
1	SUSU6	P05	0.100	0.167	2
1	SUSU6	P09	0.117	0.094	3
1	SUSU6	P13	0.233	0.233	1
1	SUSU6	P20	0.067	0.229	2
1	SUSU8	P01	0.158	0.225	3
1	SUSU8	P05	99.000	99.000	99
1	SUSU8	P09	0.250	0.175	3
1	SUSU8	P13	0.225	0.225	1
1	SUSU8	P20	0.175	0.208	2
1	SUSU9	P01	0.391	0.244	4

1	SUSU9	P05	0.069	0.069	1
1	SUSU9	P09	99.000	99.000	99
1	SUSU9	P13	0.258	0.258	1
1	SUSU9	P20	0.275	0.275	1
2	SUSU2	P02	99.000	99.000	99
2	SUSU2	P06	99.000	99.000	99
2	SUSU2	P10	0.300	0.300	1
2	SUSU2	P18	0.300	0.300	1
2	SUSU2	P21	0.225	0.192	3
2	SUSU3	P02	0.333	0.294	3
2	SUSU3	P06	0.358	0.214	3
2	SUSU3	P10	0.167	0.139	3
2	SUSU3	P18	0.333	0.333	1
2	SUSU3	P21	0.237	0.306	4
2	SUSU1	P02	99.000	99.000	99
2	SUSU1	P06	0.250	0.221	2
2	SUSU1	P10	0.208	0.221	4
2	SUSU1	P18	0.333	0.333	1
2	SUSU1	P21	0.300	0.271	2
2	SUSU5	P02	0.291	0.230	5
2	SUSU5	P06	0.483	0.225	3
2	SUSU5	P10	0.283	0.212	2
2	SUSU5	P18	0.200	0.287	2
2	SUSU5	P21	0.258	0.258	1
2	SUSU4	P02	0.308	0.352	4
2	SUSU4	P06	99.000	99.000	99
2	SUSU4	P10	0.175	0.156	4
2	SUSU4	P18	0.225	0.204	2
2	SUSU4	P21	0.192	0.194	3

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 1 – Antecedente e retomada em posição de sujeito

Região da Anáfora

Condição Hipônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
1	HISU4	P01	0.158	0.182	5
1	HISU4	P05	0.275	0.303	3
1	HISU4	P09	0.067	0.114	3
1	HISU4	P13	0.233	0.233	1
1	HISU4	P20	0.183	0.167	2
1	HISU3	P01	0.333	0.333	1
1	HISU3	P05	0.075	0.153	5
1	HISU3	P09	0.125	0.125	1
1	HISU3	P13	0.192	0.171	2
1	HISU3	P20	0.133	0.158	3
1	HISU2	P01	0.216	0.241	4
1	HISU2	P05	0.208	0.161	3
1	HISU2	P09	0.083	0.122	5
1	HISU2	P13	0.192	0.171	2
1	HISU2	P20	0.458	0.458	1
1	HISU5	P01	0.333	0.286	3
1	HISU5	P05	0.108	0.114	3
1	HISU5	P09	0.233	0.192	2
1	HISU5	P13	0.142	0.133	2
1	HISU5	P20	99.000	99.000	99
1	HISU1	P01	0.266	0.200	3

1	HISU1	P05	0.241	0.212	2
1	HISU1	P09	0.108	0.108	1
1	HISU1	P13	0.225	0.225	1
1	HISU1	P20	0.150	0.150	1
2	HISU7	P02	0.258	0.242	3
2	HISU7	P06	0.167	0.229	4
2	HISU7	P10	99.000	99.000	99
2	HISU7	P18	0.308	0.308	1
2	HISU7	P21	0.125	0.125	1
2	HISU10	P02	1.191	0.536	3
2	HISU10	P06	0.533	0.306	5
2	HISU10	P10	99.000	99.000	99
2	HISU10	P18	0.225	0.157	2
2	HISU10	P21	0.233	0.147	3
2	HISU9	P02	0.192	0.289	6
2	HISU9	P06	99.000	99.000	99
2	HISU9	P10	99.000	99.000	99
2	HISU9	P18	0.358	0.358	1
2	HISU9	P21	0.300	0.148	6
2	HISU6	P02	0.283	0.175	8
2	HISU6	P06	0.391	0.391	1
2	HISU6	P10	0.175	0.120	5
2	HISU6	P18	0.250	0.183	2
2	HISU6	P21	0.117	0.104	2
2	HISU8	P02	0.100	0.205	7
2	HISU8	P06	0.675	0.433	2
2	HISU8	P10	99.000	99.000	99
2	HISU8	P18	0.092	0.171	2
2	HISU8	P21	0.125	0.172	5

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 2 - Antecedente e retomada em posição de objeto

Região da Anáfora

Condição Hiperônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
3	SUOB10	P03	0.175	0.179	3
3	SUOB10	P07	0.266	0.194	3
3	SUOB10	P19	0.175	0.168	5
3	SUOB10	P22	0.241	0.194	4
3	SUOB10	P25	0.258	0.405	3
3	SUOB10	P26	0.100	0.314	3
3	SUOB6	P03	0.241	0.142	4
3	SUOB6	P07	0.275	0.275	1
3	SUOB6	P19	0.200	0.172	3
3	SUOB6	P22	0.300	0.213	4
3	SUOB6	P25	0.275	0.275	1
3	SUOB6	P26	0.083	0.104	2
3	SUOB8	P03	0.233	0.205	7
3	SUOB8	P07	0.341	0.341	1
3	SUOB8	P19	0.150	0.187	2
3	SUOB8	P22	0.308	0.183	3
3	SUOB8	P25	0.375	0.375	1
3	SUOB8	P26	0.258	0.258	1
3	SUOB9	P03	0.233	0.187	2
3	SUOB9	P07	0.283	0.283	1
3	SUOB9	P19	0.200	0.175	4

3	SUOB9	P22	99.000	99.000	99
3	SUOB9	P25	0.266	0.266	1
3	SUOB9	P26	0.067	0.126	6
3	SUOB7	P03	0.200	0.250	2
3	SUOB7	P07	0.192	0.154	2
3	SUOB7	P19	0.158	0.198	5
3	SUOB7	P22	0.300	0.258	2
3	SUOB7	P25	0.516	0.312	2
3	SUOB7	P26	0.092	0.115	4
4	SUOB1	P04	0.167	0.300	2
4	SUOB1	P12	0.200	0.171	4
4	SUOB1	P16	0.266	0.302	3
4	SUOB1	P23	0.283	0.226	6
4	SUOB1	P24	0.192	0.194	8
4	SUOB4	P04	99.000	99.000	99
4	SUOB4	P12	0.183	0.121	4
4	SUOB4	P16	0.325	0.235	6
4	SUOB4	P23	0.183	0.183	1
4	SUOB4	P24	0.300	0.189	8
4	SUOB5	P04	0.158	0.158	1
4	SUOB5	P12	0.391	0.258	2
4	SUOB5	P16	0.458	0.308	3
4	SUOB5	P23	0.291	0.204	2
4	SUOB5	P24	0.308	0.308	1
4	SUOB2	P04	0.142	0.112	2
4	SUOB2	P12	0.108	0.163	5
4	SUOB2	P16	0.425	0.371	6
4	SUOB2	P23	0.158	0.175	5
4	SUOB2	P24	0.233	0.203	3

4	SUOB3	P04	0.200	0.136	3
4	SUOB3	P12	0.316	0.191	2
4	SUOB3	P16	0.283	0.230	3
4	SUOB3	P23	0.233	0.233	1
4	SUOB3	P24	0.075	0.108	5

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 2 – Antecedente e retomada em posição de objeto

Região da Anáfora

Condição Hipônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
3	HIOB1	P03	0.175	0.142	3
3	HIOB1	P07	0.400	0.400	1
3	HIOB1	P19	99.000	99.000	99
3	HIOB1	P22	0.508	0.199	10
3	HIOB1	P25	0.067	0.339	4
3	HIOB1	P26	0.533	0.412	2
3	HIOB4	P03	0.250	0.243	7
3	HIOB4	P07	0.300	0.300	1
3	HIOB4	P19	99.000	99.000	99
3	HIOB4	P22	99.000	99.000	99
3	HIOB4	P25	99.000	99.000	99
3	HIOB4	P26	0.075	0.081	3
3	HIOB5	P03	0.383	0.383	1
3	HIOB5	P07	0.167	0.233	2
3	HIOB5	P19	99.000	99.000	99
3	HIOB5	P22	0.316	0.183	3

3	HIOB5	P25	0.341	0.221	4
3	HIOB5	P26	0.133	0.267	2
3	HIOB2	P03	0.158	0.197	3
3	HIOB2	P07	0.225	0.225	1
3	HIOB2	P19	99.000	99.000	99
3	HIOB2	P22	0.175	0.187	4
3	HIOB2	P25	0.225	0.212	4
3	HIOB2	P26	0.150	0.158	2
3	HIOB3	P03	0.375	0.228	8
3	HIOB3	P07	0.183	0.183	1
3	HIOB3	P19	99.000	99.000	99
3	HIOB3	P22	99.000	99.000	99
3	HIOB3	P25	0.092	0.129	2
3	HIOB3	P26	0.450	0.450	1
4	HIOB10	P04	0.267	0.192	8
4	HIOB10	P12	0.466	0.237	8
4	HIOB10	P16	0.191	0.300	3
4	HIOB10	P23	0.250	0.167	2
4	HIOB10	P24	0.075	0.155	10
4	HIOB6	P04	0.100	0.129	2
4	HIOB6	P12	0.217	0.217	1
4	HIOB6	P16	0.291	0.241	2
4	HIOB6	P23	0.125	0.167	2
4	HIOB6	P24	0.083	0.129	2
4	HIOB9	P04	0.100	0.096	4
4	HIOB9	P12	0.225	0.171	2
4	HIOB9	P16	0.408	0.316	2
4	HIOB9	P23	0.233	0.205	3
4	HIOB9	P24	0.242	0.208	5

4	HIOB8	P04	0.100	0.119	6
4	HIOB8	P12	0.075	0.187	2
4	HIOB8	P16	0.241	0.371	5
4	HIOB8	P23	0.167	0.272	5
4	HIOB8	P24	0.167	0.158	4
4	HIOB7	P04	0.100	0.144	3
4	HIOB7	P12	0.200	0.167	2
4	HIOB7	P16	0.258	0.213	5
4	HIOB7	P23	0.192	0.171	2
4	HIOB7	P24	0.092	0.192	3

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 1 - Antecedente e retomada em posição de sujeito

Região do Spill-over

Condição Hiperônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
1	SUSU10	P01	0.192	0.253	3
1	SUSU10	P05	0.067	0.087	2
1	SUSU10	P09	0.175	0.183	2
1	SUSU10	P13	0.333	0.333	1
1	SUSU10	P20	0.275	0.208	4
1	SUSU7	P01	99.000	99.000	99
1	SUSU7	P05	0.233	0.155	3
1	SUSU7	P09	0.300	0.140	5
1	SUSU7	P13	0.333	0.333	3
1	SUSU7	P20	99.000	99.000	99
1	SUSU6	P01	0.333	0.333	1

1	SUSU6	P05	0.250	0.162	4
1	SUSU6	P09	0.092	0.285	4
1	SUSU6	P13	0.083	0.179	2
1	SUSU6	P20	0.408	0.346	2
1	SUSU8	P01	0.225	0.228	3
1	SUSU8	P05	0.125	0.106	5
1	SUSU8	P09	0.258	0.142	3
1	SUSU8	P13	0.283	0.283	1
1	SUSU8	P20	0.233	0.161	3
1	SUSU9	P01	0.266	0.210	4
1	SUSU9	P05	99.000	99.000	99
1	SUSU9	P09	0.150	0.171	2
1	SUSU9	P13	0.250	0.262	2
1	SUSU9	P20	0.375	0.261	3
2	SUSU2	P02	99.000	99.000	99
2	SUSU2	P06	99.000	99.000	99
2	SUSU2	P10	0.291	0.197	3
2	SUSU2	P18	0.200	0.200	2
2	SUSU2	P21	0.258	0.175	2
2	SUSU3	P02	0.416	0.255	6
2	SUSU3	P06	0.425	0.319	6
2	SUSU3	P10	0.142	0.158	4
2	SUSU3	P18	0.275	0.212	2
2	SUSU3	P21	0.196	0.182	4
2	SUSU1	P02	99.000	99.000	99
2	SUSU1	P06	0.325	0.272	3
2	SUSU1	P10	0.142	0.179	2
2	SUSU1	P18	0.316	0.233	2
2	SUSU1	P21	0.291	0.291	1

2	SUSU5	P02	0.300	0.244	5
2	SUSU5	P06	0.142	0.179	2
2	SUSU5	P10	0.100	0.104	4
2	SUSU5	P18	0.192	0.142	2
2	SUSU5	P21	0.466	0.466	1
2	SUSU4	P02	0.167	0.230	3
2	SUSU4	P06	99.000	99.000	99
2	SUSU4	P10	0.200	0.171	2
2	SUSU4	P18	0.216	0.175	2
2	SUSU4	P21	0.266	0.237	2

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 1 – Antecedente e retomada em posição de sujeito

Região do Spill-over

Condição Hipônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
1	HISU4	P01	0.425	0.270	5
1	HISU4	P05	0.300	0.182	5
1	HISU4	P09	0.092	0.100	3
1	HISU4	P13	0.183	0.169	3
1	HISU4	P20	0.416	0.333	3
1	HISU3	P01	0.475	0.291	2
1	HISU3	P05	0.158	0.219	3
1	HISU3	P09	0.083	0.127	5
1	HISU3	P13	0.341	0.217	3
1	HISU3	P20	0.133	0.175	7
1	HISU2	P01	0.142	0.160	5

1	HISU2	P05	0.200	0.147	3
1	HISU2	P09	0.225	0.150	4
1	HISU2	P13	0.416	0.416	1
1	HISU2	P20	0.391	0.333	2
1	HISU5	P01	0.433	0.269	3
1	HISU5	P05	0.175	0.150	2
1	HISU5	P09	0.808	0.808	1
1	HISU5	P13	0.358	0.358	1
1	HISU5	P20	99.000	99.000	99
1	HISU1	P01	0.175	0.164	3
1	HISU1	P05	0.083	0.122	3
1	HISU1	P09	0.133	0.133	1
1	HISU1	P13	0.466	0.286	5
1	HISU1	P20	0.208	0.219	3
2	HISU7	P02	0.466	0.283	2
2	HISU7	P06	0.216	0.211	7
2	HISU7	P10	99.000	99.000	99
2	HISU7	P18	0.308	0.287	2
2	HISU7	P21	0.283	0.283	1
2	HISU10	P02	0.325	0.175	2
2	HISU10	P06	0.949	0.545	2
2	HISU10	P10	99.000	99.000	99
2	HISU10	P18	0.200	0.193	5
2	HISU10	P21	0.250	0.231	4
2	HISU9	P02	0.358	0.228	3
2	HISU9	P06	99.000	99.000	99
2	HISU9	P10	99.000	99.000	99
2	HISU9	P18	0.142	0.125	2
2	HISU9	P21	0.183	0.262	4

2	HISU6	P02	0.158	0.279	4
2	HISU6	P06	0.233	0.250	2
2	HISU6	P10	0.300	0.174	4
2	HISU6	P18	0.175	0.144	3
2	HISU6	P21	0.258	0.194	3
2	HISU8	P02	0.325	0.215	5
2	HISU8	P06	0.192	0.308	2
2	HISU8	P10	99.000	99.000	99
2	HISU8	P18	0.183	0.161	3
2	HISU8	P21	0.300	0.212	2

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 2 - Antecedente e retomada em posição de objeto

Região do Spill-over

Condição Hiperônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
3	SUOB10	P03	0.275	0.171	2
3	SUOB10	P07	0.192	0.278	3
3	SUOB10	P19	0.092	0.158	4
3	SUOB10	P22	0.325	0.306	4
3	SUOB10	P25	0.250	0.225	4
3	SUOB10	P26	99.000	99.000	99
3	SUOB6	P03	0.316	0.242	3
3	SUOB6	P07	0.208	0.208	1
3	SUOB6	P19	0.400	0.252	2
3	SUOB6	P22	0.475	0.475	1
3	SUOB6	P25	0.391	0.391	1

3	SUOB6	P26	0.108	0.108	1
3	SUOB8	P03	0.250	0.202	4
3	SUOB8	P07	0.200	0.200	1
3	SUOB8	P19	0.142	0.252	5
3	SUOB8	P22	0.225	0.239	3
3	SUOB8	P25	0.300	0.341	2
3	SUOB8	P26	99.000	99.000	99
3	SUOB9	P03	0.325	0.270	3
3	SUOB9	P07	0.258	0.258	1
3	SUOB9	P19	0.200	0.319	3
3	SUOB9	P22	99.000	99.000	99
3	SUOB9	P25	0.167	0.333	2
3	SUOB9	P26	0.133	0.135	4
3	SUOB7	P03	0.142	0.112	2
3	SUOB7	P07	0.132	0.132	1
3	SUOB7	P19	0.150	0.145	5
3	SUOB7	P22	0.341	0.219	2
3	SUOB7	P25	0.308	0.308	1
3	SUOB7	P26	99.000	99.000	99
4	SUOB1	P04	0.466	0.466	1
4	SUOB1	P12	0.408	0.233	2
4	SUOB1	P16	0.325	0.325	1
4	SUOB1	P23	0.200	0.175	3
4	SUOB1	P24	0.250	0.220	8
4	SUOB4	P04	99.000	99.000	99
4	SUOB4	P12	0.383	0.383	1
4	SUOB4	P16	99.000	99.000	99
4	SUOB4	P23	0.241	0.241	1
4	SUOB4	P24	0.167	0.352	4

4	SUOB5	P04	0.258	0.217	2
4	SUOB5	P12	0.142	0.153	5
4	SUOB5	P16	99.000	99.000	99
4	SUOB5	P23	0.208	0.150	2
4	SUOB5	P24	0.075	0.104	2
4	SUOB2	P04	0.150	0.129	2
4	SUOB2	P12	0.075	0.182	8
4	SUOB2	P16	0.242	0.287	2
4	SUOB2	P23	0.142	0.254	4
4	SUOB2	P24	0.117	0.100	3
4	SUOB3	P04	0.092	0.092	1
4	SUOB3	P12	0.233	0.233	1
4	SUOB3	P16	0.500	0.500	1
4	SUOB3	P23	0.383	0.383	1
4	SUOB3	P24	0.075	0.111	4

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 2 – Antecedente e retomada em posição de objeto

Região do Spill-over

Condição Hipônimo

Lista	Item	Participantes	Primeira fixação	Duração média da fixação	Número de fixações
3	HIOB1	P03	0.300	0.325	2
3	HIOB1	P07	0.358	0.358	1
3	HIOB1	P19	99.000	99.000	99
3	HIOB1	P22	0.641	0.311	6
3	HIOB1	P25	0.491	0.321	2
3	HIOB1	P26	99.000	99.000	99

3	HIOB4	P03	0.033	0.033	1
3	HIOB4	P07	0.275	0.275	1
3	HIOB4	P19	99.000	99.000	99
3	HIOB4	P22	99.000	99.000	99
3	HIOB4	P25	99.000	99.000	99
3	HIOB4	P26	0.208	0.224	2
3	HIOB5	P03	0.316	0.405	2
3	HIOB5	P07	0.358	0.358	1
3	HIOB5	P19	99.000	99.000	99
3	HIOB5	P22	0.400	0.271	4
3	HIOB5	P25	0.200	0.291	4
3	HIOB5	P26	0.083	0.062	2
3	HIOB2	P03	0.167	0.122	3
3	HIOB2	P07	0.241	0.241	1
3	HIOB2	P19	99.000	99.000	99
3	HIOB2	P22	0.225	0.181	5
3	HIOB2	P25	0.108	0.210	4
3	HIOB2	P26	0.267	0.291	2
3	HIOB3	P03	0.117	0.280	5
3	HIOB3	P07	0.125	0.154	2
3	HIOB3	P19	99.000	99.000	99
3	HIOB3	P22	99.000	99.000	99
3	HIOB3	P25	0.458	0.458	1
3	HIOB3	P26	0.117	0.126	3
4	HIOB10	P04	0.125	0.185	4
4	HIOB10	P12	0.375	0.382	7
4	HIOB10	P16	0.191	0.452	4
4	HIOB10	P23	0.217	0.291	2
4	HIOB10	P24	0.167	0.333	6

4	HIOB6	P04	0.291	0.267	2
4	HIOB6	P12	0.308	0.218	2
4	HIOB6	P16	0.316	0.253	3
4	HIOB6	P23	0.316	0.316	1
4	HIOB6	P24	0.533	0.362	2
4	HIOB9	P04	0.108	0.108	1
4	HIOB9	P12	0.308	0.308	1
4	HIOB9	P16	0.225	0.154	2
4	HIOB9	P23	0.233	0.253	3
4	HIOB9	P24	0.266	0.328	6
4	HIOB8	P04	0.083	0.083	2
4	HIOB8	P12	0.150	0.171	2
4	HIOB8	P16	0.358	0.358	1
4	HIOB8	P23	0.175	0.346	3
4	HIOB8	P24	0.092	0.283	6
4	HIOB7	P04	99.000	99.000	99
4	HIOB7	P12	0.167	0.167	1
4	HIOB7	P16	0.133	0.133	1
4	HIOB7	P23	0.375	0.375	1
4	HIOB7	P24	0.158	0.222	8

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 1 - Antecedente e retomada em posição de sujeito

Lista	Condição	Participantes	Tempo total de leitura (Time to First Mouse Click)	
			Hipônimo	Hiperônimo
1	SUJEITO	P01	12.176	5.244
1	SUJEITO	P05	17.530	7.137
1	SUJEITO	P09	6.983	4.392

1	SUJEITO	P13	5.509	4.649
1	SUJEITO	P20	6.474	6.791
1	SUJEITO	P01	7.309	99.000
1	SUJEITO	P05	12.659	13.131
1	SUJEITO	P09	9.909	5.425
1	SUJEITO	P13	6.169	4.517
1	SUJEITO	P20	12.338	99.000
1	SUJEITO	P01	7.161	6.845
1	SUJEITO	P05	11.875	9.465
1	SUJEITO	P09	6.648	11.063
1	SUJEITO	P13	6.219	5.457
1	SUJEITO	P20	4.849	5.660
1	SUJEITO	P01	8.340	9.433
1	SUJEITO	P05	12.734	12.358
1	SUJEITO	P09	6.892	9.779
1	SUJEITO	P13	5.118	4.135
1	SUJEITO	P20	99.000	7.875
1	SUJEITO	P01	9.876	10.090
1	SUJEITO	P05	9.269	7.672
1	SUJEITO	P09	5.407	4.980
1	SUJEITO	P13	7.103	5.343
1	SUJEITO	P20	5.130	4.516
1	SUJEITO	P02	15.025	16.688
1	SUJEITO	P06	4.720	99.000
1	SUJEITO	P10	99.000	4.480
1	SUJEITO	P18	4.797	4.079
1	SUJEITO	P21	5.415	5.322
2	SUJEITO	P02	13.078	99.000
2	SUJEITO	P06	16.974	99.000

2	SUJEITO	P10	99.000	5.674
2	SUJEITO	P18	5.080	5.017
2	SUJEITO	P21	4.221	6.116
2	SUJEITO	P02	11.577	12.036
2	SUJEITO	P06	6.759	99.000
2	SUJEITO	P10	99.000	6.002
2	SUJEITO	P18	7.794	99.000
2	SUJEITO	P21	8.335	99.000
2	SUJEITO	P02	17.264	99.000
2	SUJEITO	P06	99.000	12.642
2	SUJEITO	P10	99.000	6.949
2	SUJEITO	P18	5.500	5.228
2	SUJEITO	P21	8.968	4.202
2	SUJEITO	P02	11.266	10.508
2	SUJEITO	P06	4.956	4.298
2	SUJEITO	P10	7.370	5.779
2	SUJEITO	P18	4.908	5.275
2	SUJEITO	P21	3.845	3.823

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 2 – Antecedente e retomada em posição de objeto

Lista	Condição	Participantes	Tempo total de leitura (Time to First Mouse Click)	
			Hipônimo	Hiperônimo
3	OBJETO	P03	3.571	3.802
3	OBJETO	P07	4.386	9.570
3	OBJETO	P19	99.000	12.222
3	OBJETO	P22	14.778	5.454
3	OBJETO	P25	10.648	24.395

3	OBJETO	P26	8.127	9.275
3	OBJETO	P03	8.587	5.330
3	OBJETO	P07	3.984	3.435
3	OBJETO	P19	99.000	5.777
3	OBJETO	P22	99.000	7.357
3	OBJETO	P25	99.000	5.053
3	OBJETO	P26	6.727	4.616
3	OBJETO	P03	3.159	11.022
3	OBJETO	P07	3.344	4.159
3	OBJETO	P19	99.000	6.448
3	OBJETO	P22	8.555	9.423
3	OBJETO	P25	14.909	6.421
3	OBJETO	P26	4.911	6.576
3	OBJETO	P03	7.106	3.574
3	OBJETO	P07	4.101	3.961
3	OBJETO	P19	99.000	6.200
3	OBJETO	P22	10.541	99.000
3	OBJETO	P25	15.228	11.572
3	OBJETO	P26	6.114	9.830
3	OBJETO	P03	17.901	2.963
3	OBJETO	P07	4.713	3.650
3	OBJETO	P19	99.000	7.020
3	OBJETO	P22	99.000	7.151
3	OBJETO	P25	6.719	4.989
3	OBJETO	P26	6.408	5.007
4	OBJETO	P04	8.984	6.661
4	OBJETO	P12	13.334	5.047
4	OBJETO	P16	8.810	6.327
4	OBJETO	P23	5.009	9.763

4	OBJETO	P24	28.177	17.879
4	OBJETO	P04	5.274	4.560
4	OBJETO	P12	4.738	5.489
4	OBJETO	P16	7.706	12.967
4	OBJETO	P23	4.393	5.524
4	OBJETO	P24	6.910	19.367
4	OBJETO	P04	6.357	3.893
4	OBJETO	P12	6.764	5.542
4	OBJETO	P16	9.265	6.522
4	OBJETO	P23	8.709	3.826
4	OBJETO	P24	13.250	6.589
4	OBJETO	P04	5.310	6.360
4	OBJETO	P12	4.858	11.284
4	OBJETO	P16	10.496	14.082
4	OBJETO	P23	9.048	7.786
4	OBJETO	P24	10.000	14.173
4	OBJETO	P04	7.512	4.945
4	OBJETO	P12	6.328	4.796
4	OBJETO	P16	6.583	6.620
4	OBJETO	P23	5.457	4.600
4	OBJETO	P24	8.954	10.492

Observação: O número 99.000 ou 99 representa ausência de dado para o item experimental

Dados do estudo das preferências semânticas

Grupo 1 - Antecedente e retomada em posição de sujeito

Lista Condição Participantes Tempo total de fixação nas regiões de interesse

	Lista	Condição	Participantes	Hiperônimo		Hipônimo		
				Anáfora	Spillover	Anáfora	Spillover	
1	SUJEITO	P01		3.090		3.115	3.664	4.031

1	SUJEITO	P05	1.027	1.818	2.923	2.673
1	SUJEITO	P09	1.366	2.973	1.566	2.474
1	SUJEITO	P13	1.283	2.498	1.408	3.365
1	SUJEITO	P20	1.674	3.689	1.949	4.564
2	SUJEITO	P02	4.689	5.109	6.896	3.789
2	SUJEITO	P06	3.231	4.605	5.014	4.456
2	SUJEITO	P10	2.648	2.340	2.790	2.452
2	SUJEITO	P18	1.949	1.924	1.688	2.707
2	SUJEITO	P21	4.403	2.856	2.523	3.265

Dados do estudo das preferências semânticas

Grupo 2 – Antecedente e retomada em posição de objeto

Lista	Condição	Participantes	Tempo total de fixação nas regiões de interesse			
			Hiperônimo		Hipônimo	
			Anáfora	Spillover	Anáfora	Spillover
3	OBJETO	P03	3.410	2.910	4.922	3.259
3	OBJETO	P07	1.791	1.631	1.574	1.541
3	OBJETO	P19	3.423	4.078	2.185	3.073
3	OBJETO	P22	3.686	3.094	4.472	5.496
3	OBJETO	P25	2.757	2.948	4.306	3.714
3	OBJETO	P26	2.624	0.650	2.365	1.532
4	OBJETO	P04	1.166	1.250	3.323	1.548
4	OBJETO	P12	2.065	3.300	3.160	3.924
4	OBJETO	P16	3.934	1.399	4.939	3.365
4	OBJETO	P23	2.182	2.465	2.982	3.071
4	OBJETO	P24	3.915	4.118	4.056	8.171

Dados do estudo das preferências semânticas

Grupo 1 - Antecedente e retomada em posição de sujeito

Lista	Condição	Participantes	Leitura correferente em percentual	
			Hiperônimo	Hipônimo
1	SUJEITO	P01	80%	100%
1	SUJEITO	P05	100%	100%
1	SUJEITO	P09	100%	100%
1	SUJEITO	P13	100%	100%
1	SUJEITO	P20	80%	80%
1	SUJEITO	P17	100%	80%
2	SUJEITO	P02	60%	100%
2	SUJEITO	P06	60%	80%
2	SUJEITO	P10	100%	20%
2	SUJEITO	P18	100%	100%
2	SUJEITO	P21	100%	100%
2	SUJEITO	P14	100%	100%

Dados do estudo das preferências semânticas**Grupo 2 - Antecedente e retomada em posição de objeto**

Lista	Condição	Participantes	Leitura correferente em percentual	
			Hiperônimo	Hipônimo
3	OBJETO	P03	100%	100%
3	OBJETO	P07	100%	100%
3	OBJETO	P19	100%	0%
3	OBJETO	P22	80%	60%
3	OBJETO	P25	100%	80%
3	OBJETO	P26	100%	100%
3	OBJETO	P11	80%	80%
3	OBJETO	P15	100%	100%
4	OBJETO	P04	100%	100%

4	OBJETO	P12	100%	100%
4	OBJETO	P16	100%	100%
4	OBJETO	P23	100%	100%
4	OBJETO	P24	100%	100%
4	OBJETO	P08	100%	100%

APÊNDICE E – DADOS REFERENTES ÀS MEDIAS DO ESTUDO SOBRE AS PREFERÊNCIAS SINTÁTICAS

Dados do estudo das preferências sintáticas

Performance Comportamental

Participantes	Condição Pronome Nulo				Condição Pronome Pleno			
	Correferência com o sujeito		Correferência com o objeto		Correferência com o sujeito		Correferência com o objeto	
	Anáfora (PNAN)	Catáfora (PNCA)	Anáfora (PNAN)	Catáfora (PNCA)	Anáfora (PPAN)	Catáfora (PPCA)	Anáfora (PPAN)	Catáfora (PPCA)
P02	25%	75%	75%	25%	50%	75%	50%	25%
P04	75%	100%	25%	0%	50%	25%	50%	75%
P05	75%	100%	25%	0%	75%	75%	25%	25%
P06	100%	100%	0%	0%	75%	25%	25%	75%
P07	75%	100%	25%	0%	75%	25%	25%	75%
P08	100%	100%	0%	0%	25%	100%	75%	0%
P09	100%	100%	0%	0%	0%	50%	100%	50%
P10	75%	100%	25%	0%	75%	75%	25%	25%
P12	100%	100%	0%	0%	25%	100%	75%	0%
P13	100%	75%	0%	25%	50%	25%	50%	75%
P14	100%	100%	0%	0%	75%	100%	25%	0%
P15	75%	100%	25%	0%	50%	100%	50%	0%
P16	75%	75%	25%	25%	50%	25%	50%	75%
P17	100%	100%	0%	0%	25%	75%	75%	25%
P18	50%	100%	50%	0%	0%	75%	100%	25%
P19	100%	100%	0%	0%	50%	25%	50%	75%
P20	100%	100%	0%	0%	50%	100%	50%	0%
P21	75%	100%	25%	0%	25%	75%	75%	25%
P22	100%	100%	0%	0%	75%	100%	25%	0%
P23	100%	100%	0%	0%	50%	100%	50%	0%

P24	75%	100%	25%	0%	25%	50%	75%	50%
P25	100%	75%	0%	25%	25%	100%	75%	0%

Dados do estudo das preferências sintáticas

Duração média da fixação nas regiões de interesse

Participantes	Condição Pronome Nulo				Condição Pronome Pleno			
	Região do Sujeito		Região do Objeto		Região do Sujeito		Região do Objeto	
	Anáfora (PNAN)	Catáfora (PNCA)	Anáfora (PNAN)	Catáfora (PNCA)	Anáfora (PPAN)	Catáfora (PPCA)	Anáfora (PPAN)	Catáfora (PPCA)
P02	0.428	0.354	0.348	0.235	0.448	0.285	0.226	0.238
P04	0.266	0.257	0.236	0.398	0.308	0.221	0.244	0.283
P05	0.362	0.389	0.233	0.402	0.279	0.322	0.293	0.402
P06	0.382	0.308	0.249	0.260	0.282	0.315	0.308	0.282
P07	0.408	0.269	0.231	0.195	0.362	0.345	0.222	0.245
P08	0.365	0.272	0.241	0.426	0.361	0.322	0.312	0.410
P09	0.302	0.333	0.368	0.349	0.284	0.327	0.329	0.280
P10	0.409	0.417	0.280	0.656	0.464	0.447	0.281	0.544
P12	0.320	0.248	0.198	0.163	0.275	0.318	0.243	0.205
P13	0.318	0.270	0.248	0.460	0.388	0.260	0.263	0.293
P14	0.130	0.410	0.266	0.445	0.280	0.237	0.180	0.305
P15	0.467	0.335	0.370	0.360	0.364	0.408	0.302	0.500
P16	0.373	0.330	0.225	0.350	0.388	0.370	0.280	0.420
P17	0.393	0.213	0.175	0.468	0.310	0.198	0.240	0.225
P18	0.276	0.275	0.182	0.240	0.277	0.300	0.243	0.235
P19	0.175	0.255	0.183	0.170	0.223	0.198	0.197	0.208
P20	0.300	0.285	0.358	0.275	0.258	0.343	0.303	0.215
P21	0.218	0.250	0.308	0.315	0.247	0.283	0.247	0.443
P22	0.248	0.318	0.246	0.263	0.277	0.280	0.250	0.343
P23	0.270	0.268	0.283	0.188	0.326	0.198	0.232	0.203

P24		0.343	0.330	0.265	0.413	0.385	0.268	0.238	0.393
P25	-		0.180	0.190	0.325	0.240	0.153	0.270	0.213

Dados do estudo das preferências sintáticas

Tempo total de fixação nas regiões de interesse

Participantes	Condição Pronome Nulo				Condição Pronome Pleno			
	Região do Sujeito		Região do Objeto		Região do Sujeito		Região do Objeto	
	Anáfora (PNAN)	Catáfora (PNCA)	Anáfora (PNAN)	Catáfora (PNCA)	Anáfora (PPAN)	Catáfora (PPCA)	Anáfora (PPAN)	Catáfora (PPCA)
P02	0.704	3.858	2.262	1.191	1.146	2.230	1.086	1.262
P04	1.584	2.708	1.924	0.789	1.714	1.512	3.688	1.655
P05	1.049	1.830	0.783	1.247	1.075	2.624	1.212	2.049
P06	0.550	1.266	1.065	0.687	0.554	2.106	2.107	0.500
P07	0.697	0.781	0.684	0.343	1.295	1.202	1.136	0.525
P08	0.927	1.308	1.323	2.169	1.409	1.256	2.425	0.894
P09	1.115	1.640	2.038	1.164	0.793	2.418	1.368	0.806
P10	0.804	0.947	0.851	1.085	0.727	1.000	1.077	0.902
P12	0.945	1.160	0.705	0.623	1.103	1.243	1.185	0.773
P13	1.355	1.558	1.640	1.370	1.700	2.200	1.635	0.867
P14	0.130	0.615	0.600	1.280	0.560	0.323	0.878	0.470
P15	1.928	1.803	1.783	1.480	2.050	3.550	1.273	2.445
P16	2.430	2.200	2.335	1.525	1.133	2.790	1.898	2.525
P17	0.650	0.778	0.550	0.658	0.985	0.690	0.513	0.270
P18	0.623	0.565	0.263	0.365	0.500	0.348	0.688	0.338
P19	0.983	1.368	0.503	0.395	1.170	1.293	0.603	1.120
P20	0.930	1.655	1.363	0.785	0.648	0.970	0.670	0.295
P21	0.478	0.430	0.503	0.450	0.538	0.388	0.670	0.773
P22	0.578	1.240	0.788	0.393	0.385	0.775	1.030	0.343
P23	0.457	1.083	1.460	0.595	0.338	0.793	1.113	0.520

P24		0.428	1.235	0.870	0.980	0.943	1.238	1.203	0.853
P25	-		0.375	0.675	0.800	0.240	0.308	0.670	0.423

Dados do estudo das preferências sintáticas

Tempo total de leitura (Time to first mouse click)

Participantes	Condição Pronome Nulo		Condição Pronome Pleno	
	Anáfora (PNAN)	Catáfora (PNCA)	Anáfora (PPAN)	Catáfora (PPCA)
P02	8.384	14.785	7.890	10.909
P04	13.618	10.163	19.502	10.008
P05	6.563	9.136	7.163	12.135
P06	7.240	7.742	17.635	13.000
P07	4.114	4.239	8.923	5.795
P08	8.126	9.100	10.533	7.587
P09	10.780	7.950	8.609	11.077
P10	6.494	6.407	6.284	8.741
P12	5.620	5.128	7.035	7.098
P13	9.670	8.433	10.635	9.225
P15	9.203	10.100	10.673	19.405
P16	15.418	11.923	8.883	17.193
P17	4.213	4.630	4.538	4.398
P18	3.714	3.265	4.837	3.225
P19	6.498	7.095	6.583	8.895
P20	8.010	9.133	5.755	5.653
P21	3.374	3.100	3.960	3.265
P22	4.540	5.715	5.833	4.923
P23	6.050	5.245	6.595	4.728
P24	5.268	7.345	7.935	6.960

Dados do estudo das preferências sintáticas**Duração média da fixação nos segmentos do período complexo por subordinação****Condição Pronome Nulo****Anáfora (PNAN)**

Participantes	Oração principal			Oração subordinada temporal			Complemento
	Sujeito	Verbo	Objeto	Conjunção	pro	Verbo	
P12	0.323	0.235	0.199	0.177	-	0.241	0.201
P13	0.317	0.271	0.246	0.204	-	0.348	0.220
P14	0.133	0.212	0.266	0.239	-	0.257	0.536
P15	0.429	0.335	0.381	0.257	-	0.264	0.333
P16	0.371	0.268	0.226	0.260	-	0.296	0.433
P17	0.393	0.185	0.175	0.231	-	0.287	0.310
P18	0.278	0.221	0.181	0.366	-	0.197	0.295
P19	0.174	0.206	0.183	0.197	-	0.244	0.391
P20	0.299	0.284	0.357	0.208	-	0.285	0.200
P21	0.217	0.257	0.304	0.231	-	0.356	0.405
P22	0.248	0.264	0.243	0.275	-	0.384	0.233
P23	0.272	0.244	0.267	0.243	-	0.239	0.181
P24	0.342	0.180	0.263	0.254	-	0.268	0.422

Condição Pronome Pleno**Anáfora (PPAN)**

Participantes	Oração principal			Oração subordinada temporal			Complemento
	Sujeito	Verbo	Objeto	Conjunção	Pronome Pleno	Verbo	
P12	0.275	0.198	0.242	0.185	0.215	0.252	0.207
P13	0.388	0.243	0.264	0.213	0.289	0.309	0.298

P14	0.279	0.200	0.180	0.220	0.217	0.233	0.795
P15	0.380	0.306	0.274	0.250	0.246	0.271	0.335
P16	0.387	0.228	0.281	0.225	0.215	0.295	0.321
P17	0.311	0.201	0.238	0.200	0.246	0.249	0.260
P18	0.275	0.262	0.244	0.220	0.213	0.207	0.163
P19	0.221	0.244	0.197	0.150	0.236	0.211	0.196
P20	0.259	0.344	0.300	0.251	0.291	0.621	0.431
P21	0.249	0.231	0.246	0.252	0.196	0.319	0.349
P22	0.275	0.281	0.250	0.296	0.548	0.258	0.293
P23	0.339	0.300	0.237	0.332	0.289	0.340	0.283
P24	0.387	0.216	0.238	0.289	0.297	0.270	0.420

Condição Pronome Nulo**Catáfora (PNCA)****Oração subordinada temporal****Oração principal**

Participantes	Conjunção	[pro]	Verbo	Complemento	Sujeito	Verbo	Objeto
P12	0.251	-	0.218	0.171	0.246	0.234	0.163
P13	0.297	-	0.207	0.157	0.270	0.330	0.459
P14	0.200	-	0.127	0.220	0.415	0.236	0.442
P15	0.340	-	0.250	0.261	0.336	0.319	0.358
P16	0.299	-	0.239	0.284	0.329	0.293	0.347
P17	0.317	-	0.231	0.206	0.215	0.329	0.467
P18	0.285	-	0.173	0.289	0.274	0.242	0.243
P19	0.221	-	0.292	0.235	0.256	0.227	0.171
P20	0.355	-	0.273	0.204	0.284	0.330	0.279
P21	0.201	-	0.131	0.297	0.251	0.325	0.318
P22	0.239	-	0.255	0.211	0.315	0.404	0.261

P23	0.224	-	0.216	0.304	0.266	0.303	0.187
P24	0.278	-	0.162	0.227	0.333	0.375	0.411

**Condição Pronome Pleno
Catáfora (PPCA)**

Oração subordinada temporal

Oração principal

Participantes	Pronome			Complemento	Oração principal		
	Conjunção	Pleno	Verbo		Sujeito	Verbo	Objeto
P12	0.189	0.169	0.187	0.207	0.316	0.303	0.205
P13	0.260	0.188	0.192	0.180	0.256	0.242	0.293
P14	0.138	0.144	0.164	0.179	0.236	0.215	0.304
P15	0.288	0.256	0.270	0.284	0.404	0.327	0.499
P16	0.241	0.210	0.262	0.305	0.371	0.317	0.418
P17	0.239	0.205	0.237	0.204	0.197	0.292	0.224
P18	0.189	0.167	0.199	0.211	0.299	0.327	0.236
P19	0.168	0.237	0.220	0.231	0.199	0.269	0.207
P20	0.254	0.186	0.206	0.275	0.343	0.292	0.219
P21	0.187	0.179	0.197	0.237	0.284	0.236	0.442
P22	0.231	0.196	0.223	0.245	0.281	0.253	0.344
P23	0.165	0.202	0.264	0.372	0.195	0.237	0.200
P24	0.236	0.182	0.157	0.223	0.268	0.305	0.394

ANEXOS

Aprovação do Comitê de Ética em Pesquisa da Universidade Federal do Ceará



Universidade Federal do Ceará
Comitê de Ética em Pesquisa

Of. Nº 111/11

Fortaleza, 24 de maio de 2011.

Protocolo COMEPE nº 056/11

Pesquisador responsável: Elisângela Nogueira Teixeira

Título do Projeto: “Rastreamento Ocular de Processos Cognitivos”

Levamos ao conhecimento de V.S^a. que o Comitê de Ética em Pesquisa da Universidade Federal do Ceará – COMEPE, dentro das normas que regulamentam a pesquisa em seres humanos, do Conselho Nacional de Saúde – Ministério da Saúde, Resolução nº 196 de 10 de outubro de 1996 e complementares, aprovou o protocolo e o TCLE do projeto supracitado na reunião do dia 28 de abril de 2011.

Outrossim, informamos, que o pesquisador deverá se comprometer a enviar o relatório final do referido projeto.

Atenciosamente,

A handwritten signature in black ink, appearing to read 'F. Frota Bezerra', written over a circular stamp.

Dr. Fernando A. Frota Bezerra
Coordenador do Comitê
de Ética em Pesquisa
COMEPE/UFC

Termo de Consentimento Livre e Esclarecido referente ao Protocolo COMEPE n.056/11

Você está sendo convidado(a) a participar do projeto de pesquisa intitulado *Rastreamento ocular de Processos Cognitivos* que tem por objetivo descrever a trajetória da movimentação ocular vista como reflexo de processos cognitivos complexos, como a busca visual e a leitura. Neste estudo, você será convidado a sentar diante de uma tela de computador para realizar tarefas de dois tipos: buscas visuais e leitura de textos. Esta tela está equipada com uma câmera que registra o movimento dos seus olhos.

Para que você participe deste estudo, sua visão deve ser normal ou corrigida pelo uso de lentes ou óculos. Não há qualquer risco para você. Este registro não se utiliza de nenhum procedimento invasivo e você não sentirá qualquer desconforto.

Trata-se de um estudo experimental que não trará qualquer benefício direto para o participante. Testamos a hipótese de que a trajetória ocular segue um comportamento semelhante a de outros sistemas complexos da natureza descritos pela teoria das redes complexas, que procura descrever de forma simplificada como ocorrem as interações dentro de um sistema. Neste âmbito, investigamos os papéis das estruturas linguísticas e visuais que tornam possível a compreensão de textos e a procura de objetos em meio a distratores visuais.

Somente no final do estudo, poderemos concluir que o comportamento ocular humano em diferentes tarefas cognitivas pode seguir padrões em função de especificidades das tarefas e dos objetivos de cada tarefa. Estes resultados podem contribuir para o esclarecimento de como as informações visuais e linguísticas são processadas mentalmente, por meio do tempo de execução das tarefas, que é uma medida indireta sobre o tempo de processamento da informação.

Você tem a liberdade de retirar seu consentimento a qualquer momento e deixar de participar do estudo, sem que isto acarrete qualquer prejuízo para você.

As informações obtidas serão analisadas em conjunto com outros participantes, não sendo divulgado a identificação de nenhum participante da pesquisa.

Você tem o direito de ser mantido atualizado sobre os resultados parciais das pesquisas, quando em estudos abertos, ou de resultados que sejam do conhecimento dos pesquisadores.

Não há despesas pessoais para o participante em qualquer fase do estudo, incluindo exames e consultas. Também não há compensação financeira relacionada a sua participação. Se existir qualquer despesa adicional, ela será absorvida pelo orçamento da pesquisa.

O pesquisador se compromete em utilizar os dados e o material coletado somente para esta pesquisa.

Em qualquer etapa do estudo, você terá acesso aos profissionais responsáveis pela pesquisa para esclarecimento de eventuais dúvidas. A principal investigadora é a doutoranda Elisângela Nogueira Teixeira, que pode ser encontrada no endereço: Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará, CEP 60020-181, Fortaleza, Ceará, no e-mail: teixeiraelis@gmail.com ou nos telefones: (85) 3366 7626 ou 7627.

Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) do HUWC – Rua Capitão Francisco Pedro 1290, Rodolfo Teófilo; fone: 3366-8589 – E-mail: cephuwc@huwc.ufc.br.

Caso você se sinta suficientemente informado a respeito das informações que leu ou que foram lidas para você sobre os propósitos do estudo, os procedimentos a serem realizados, seus desconfortos e riscos, as garantias de confidencialidade e de esclarecimentos permanentes e de que sua participação é voluntária, de que não há remuneração para participar do estudo e se você concordar em participar, solicitamos que assine no espaço abaixo.

Assinatura do participante ou representante legal Data ____ / ____ / ____

Assinatura da testemunha Data ____ / ____ / ____

Para casos de participantes menores de 18 anos, analfabetos, semi-analfabetos ou portadores de deficiência auditiva ou visual.

Assinatura do responsável pelo estudo Data ____ / ____ / ____