

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA COGNITIVA

FRANCIS RICARDO DOS REIS JUSTI

**O efeito de vizinhança ortográfica no português brasileiro: uma comparação de
múltiplas tarefas**

RECIFE

2009

FRANCIS RICARDO DOS REIS JUSTI

**O efeito de vizinhança ortográfica no português brasileiro: uma comparação de
múltiplas tarefas**

Tese apresentada ao Programa de Pós-Graduação em Psicologia Cognitiva da Universidade Federal de Pernambuco como requisito parcial para a obtenção do título de Doutor em Psicologia.
Área de concentração: Psicologia Cognitiva.

Orientador: Prof. Dr. Antonio Roazzi

RECIFE

2009

Justi, Francis Ricardo dos Reis

O efeito de vizinhança ortográfica no português brasileiro : uma comparação de múltiplas tarefas / Francis Ricardo dos Reis Justi. -- Recife: O Autor, 2009.

193 folhas : il., fig., tab.

Tese (doutorado) – Universidade Federal de Pernambuco. CFCH. Psicologia, 2009.

Inclui: bibliografia e apêndices.

1. Psicologia cognitiva. 2. Semântica. 3. Língua portuguesa – Brasil - Lexicologia. 4. Vizinhança ortográfica. I. Título.

**159.9
150**

**CDU (2. ed.)
CDD (22. ed.)**

**UFPE
BCFCH2009/36**

FOLHA DE APROVAÇÃO

Francis Ricardo dos Reis Justi

O Efeito de Vizinhança Ortográfica no Português Brasileiro: uma comparação de múltiplas tarefas.

Tese apresentada ao Programa de Pós-Graduação em Psicologia Cognitiva da Universidade Federal de Pernambuco para obtenção do título de Doutor.

Área de Concentração: Psicologia Cognitiva

Aprovado em: 07 de outubro de 2009

Banca Examinadora

Prof. Dr. Antonio Roazzi
Instituição: U.F.PE

Assinatura: 

Profa. Dra. Angela Maria Vieira Pinheiro
Instituição: U.F.MG

Assinatura: 

Prof. Dr. Jorge Artur Peçanha de Miranda Coelho
Instituição: U.F.AL

Assinatura: 

Profa. Dra. Alina Galvão Spinillo
Instituição: U.F.PE

Assinatura: 

Profa. Dra. M^a da Graça Bompastor Borges Dias
Instituição: U.F.PE

Assinatura: 

*Dedico esse trabalho a você, Cláudia, meu amor...
Por estar presente nos momentos mais difíceis e por
tornar possíveis os momentos mais doces e
importantes da minha vida. Sem você, nada valeria
a pena.*

AGRADECIMENTOS

Gostaria de agradecer aos meus pais, minha irmã e meus avós por todo o apoio, incentivo e suporte durante todos esses anos.

Sou muito grato aos professores Saulo, Marcos e Márcia que durante os meus anos de graduação despertaram em mim o gosto pela pesquisa científica em Psicologia e à professora Ângela Pinheiro que me apresentou a área de reconhecimento visual de palavras durante o meu curso de mestrado. Em especial, gostaria de agradecer ao professor Antonio Roazzi, meu orientador nessa tese de doutorado, sobretudo pela sua confiança no meu trabalho, sua disponibilidade e apoio.

Agradeço ao Programa de Pós-Graduação em Psicologia Cognitiva da Universidade Federal de Pernambuco e a todos os professores e colegas que, de algum modo, contribuíram para minha formação durante esse curso de doutorado. Agradeço, em especial, aos membros da banca examinadora, por sua sensibilidade e presteza em permitirem que essa defesa pudesse estar ocorrendo hoje.

Gostaria também de agradecer aos alunos dos cursos de Psicologia da Universidade Federal de Pernambuco e da Universidade Federal de Alagoas por terem participado dessa pesquisa. Por fim, gostaria de agradecer ao CNPq pelo financiamento desse estudo por meio de uma bolsa de doutorado. Graças a instituições como o CNPq e à pessoas como os participantes desse estudo, a ciência pode acontecer.

RESUMO

Justi, F. R. R. (2009) *O efeito de vizinhança ortográfica no português brasileiro: uma comparação de múltiplas tarefas*. 2009. 193 f. Tese (Doutorado) – Centro de Filosofia e Ciências Humanas, Universidade Federal de Pernambuco, Recife.

O presente estudo investigou o efeito de vizinhança ortográfica no reconhecimento visual de palavras em adultos falantes do português do Brasil. Uma análise de corpus indicou que N (*neighborhood density*) e NF (*neighborhood frequency*) apresentam uma correlação ínfima com TLN (*transposed letter neighbors*), sendo que, ao mesmo tempo, afetam uma proporção muito maior de palavras do português brasileiro. Desse modo, uma das primeiras conclusões desse trabalho foi a de que, pelo menos no caso do português brasileiro, N e NF são medidas mais relevantes de vizinhança ortográfica do que TLN. Além da análise de corpus, foram desenvolvidos quatro experimentos para investigar os efeitos de N e NF. Efeitos inibidores de NF foram observados nos experimentos que utilizaram ou uma tarefa de identificação perceptual ou uma tarefa de leitura em voz alta do tipo *Go/NoGo* (na qual as pseudopalavras tinham muitos vizinhos ortográficos). Já efeitos facilitadores de N foram observados nos experimentos que utilizaram ou uma tarefa de identificação perceptual ou uma tarefa de leitura em voz alta (onde as palavras e pseudopalavras deviam ser lidas em uma mesma sessão). No experimento que utilizou uma tarefa de categorização semântica os efeitos de N e NF foram nulos. Os resultados desse estudo foram então combinados com os resultados de outros estudos realizados com falantes do português brasileiro e uma meta-análise foi realizada. A meta-análise indicou que o tipo de tarefa empregada nos estudos é um moderador dos efeitos de N e NF. Sendo que, de uma forma geral, em tarefas onde o processo de identificação lexical foi enfatizado o efeito de NF foi inibidor e o de N nulo e, em tarefas onde algum tipo de processamento estratégico predominou, o efeito de NF foi nulo e o de N foi facilitador. Os resultados desse estudo corroboram a hipótese de que os efeitos de N e NF podem ser modulados pela demanda das tarefas utilizadas para se investigar o acesso lexical e, além disso, colocam em dificuldades modelos de reconhecimento visual de palavras que não têm mecanismos específicos que os permitam prever quais tarefas enfatizam processos estratégicos e quais enfatizam o processo de identificação lexical.

Palavras-chave: vizinhança ortográfica; reconhecimento visual de palavras; acesso lexical; meta-análise; análise de corpus.

ABSTRACT

Justi, F. R. R. (2009) *The orthographic neighborhood effect on visual word recognition processes of Brazilian Portuguese speakers: a cross-task comparison*. 2009. 193 f. Thesis (Doctoral) – Centro de Filosofia e Ciências Humanas, Universidade Federal de Pernambuco, Recife.

This study investigated the effects of orthographic neighborhood on visual word recognition processes of Brazilian Portuguese speakers. A linguistic corpus analysis indicated that N (*neighborhood density*) and NF (*neighborhood frequency*) are poorly correlated with TLN (*transposed letter neighbors*), and since, both N and NF are present in a much higher proportion of words these measures seem to be more relevant measures of orthographic similarity than TLN. Therefore this study carried out four experiments to investigate N and NF effects. NF showed inhibitory effects on the experiment with a Perceptual Identification Task, and on the experiment with a *Go/NoGo* Reading Aloud Task (with high N pseudowords). On the other hand, N showed facilitative effects on the experiment with a Perceptual Identification Task, and on the experiment with a Reading Aloud Task. Neither N nor NF showed any effects on the experiment with a Semantic Categorization Task. A meta-analysis was carried out combining the results of this study and other studies developed with Brazilian Portuguese speakers. A main result of the meta-analysis was that type of task is a moderator of N and NF effects. Overall, the results of the meta-analysis showed that in tasks which emphasize lexical identification NF has an inhibitory effect and N a null effect, and in tasks which emphasize some sort of strategy N has a facilitative effect and NF a null effect. These findings corroborate quite well the hypothesis that N and NF effects can vary according to task demands. Therefore visual word recognition models shall address this question to fully explain orthographic neighborhood effects on visual word recognition.

Keywords: orthographic neighborhood; visual word recognition; lexical access; meta-analysis; corpus linguistics.

RESUMO ESTENDIDO

Justi, F. R. R. (2009) *O efeito de vizinhança ortográfica no português brasileiro: uma comparação de múltiplas tarefas*. 2009. 193 f. Tese (Doutorado) – Centro de Filosofia e Ciências Humanas, Universidade Federal de Pernambuco, Recife.

Nas pesquisas sobre o reconhecimento visual de palavras, o efeito que palavras similares em termos ortográficos desencadeiam no processo de reconhecimento de um estímulo-alvo é conhecido por *efeito de vizinhança ortográfica*. No estudo desse tópico, duas variáveis têm sido intensamente investigadas: “N” (refere-se ao número de palavras que pode ser gerado ao se mudar uma letra da palavra-alvo, enquanto a posição das demais letras é preservada – p.ex.: a palavra ‘alma’ teria cinco vizinhos ortográficos – ‘arma’, ‘asma’, ‘alga’, ‘alça’ e ‘alta’) e “NF” (refere-se ao caso da palavra-alvo ter ou não vizinhos ortográficos de maior frequência de ocorrência – p.ex., a palavra ‘anão’ tem um vizinho ortográfico de maior frequência que é a palavra ‘ação’). O presente estudo investigou o efeito de vizinhança ortográfica no reconhecimento visual de palavras em adultos falantes do português do Brasil. Inicialmente, uma análise de corpus lingüístico foi desenvolvida visando elucidar qual seria a melhor medida de vizinhança ortográfica para os propósitos desse estudo. Essa análise avaliou a correlação entre medidas tradicionais de vizinhança ortográfica (N e NF) e o número de vizinhos ortográficos formados pela transposição de duas letras (TLN – *transposed letter neighbors*), bem como se diferenças na distribuição estatística das variáveis de vizinhança ortográfica poderiam explicar as diferenças encontradas nos efeitos dessas variáveis na língua portuguesa e inglesa. Para tanto, foram geradas estatísticas de vizinhança ortográfica para 8465 palavras do português. Não foram observadas grandes diferenças entre as palavras do português e as da língua inglesa no que tange às estatísticas de vizinhança ortográfica, tornando difícil explicar as discordâncias encontradas nos efeitos de N e NF com base na idéia de estruturas de vizinhança ortográfica diferentes entre as línguas. Além disso, a correlação entre as medidas tradicionais de vizinhança ortográfica e a medida TLN foram ínfimas, o que indica que os efeitos de N e NF, sejam quais forem, dificilmente podem ser explicados por sua relação com TLN. Como os resultados da análise desenvolvida evidenciaram que N e NF afetam uma proporção muito maior de palavras do que TLN e, além disso, são medidas contempladas pelos principais modelos de reconhecimento visual de palavras da atualidade, concluiu-se que N e NF são medidas mais relevantes do conceito de vizinhança ortográfica. Levando-se em consideração N e NF, quatro experimentos foram desenvolvidos visando investigar o efeito dessas variáveis em tarefas que enfatizavam o processo de identificação lexical e em tarefas que enfatizavam alguma forma de processamento estratégico. Trinta e dois alunos do curso de Psicologia da UFAL participaram dos quatro experimentos desenvolvidos. Os estímulos experimentais consistiram em 64 palavras dissílabas de baixa frequência de ocorrência que se subdividiam em quatro condições experimentais, de acordo com uma manipulação fatorial de “N” (um – ou – quatro ou mais vizinhos ortográficos) e “NF” (sem – ou – com vizinhos ortográficos de maior frequência). A classe gramatical das palavras foi mantida constante (todas as palavras eram substantivos), juntamente com o número de vizinhos transpostos (apenas um por condição experimental). Essas 64 palavras foram então subdivididas em quatro listas (pareadas pela natureza do fonema inicial e pela frequência de ocorrência média) que foram cruzadas entre as quatro condições experimentais. Por fim, a ordem de apresentação das listas foi contrabalanceada entre as quatro condições experimentais, de forma que cada participante foi exposto a apenas uma lista por condição experimental em cada experimento, permitindo assim que participasse dos quatro experimentos sem ser exposto a uma mesma palavra mais de uma vez. O primeiro

experimento realizado foi uma tarefa de leitura em voz alta do tipo *Go/NoGo* que visava enfatizar o processo de identificação lexical. Nessa tarefa os participantes foram expostos a palavras e pseudopalavras, sendo requerido desses a leitura em voz alta apenas das palavras. Nesse estudo observou-se uma interação estatisticamente significativa de N e NF, sendo que a presença de um vizinho ortográfico de maior frequência de ocorrência retardou o reconhecimento das palavras alvo (apresentou um efeito inibidor) apenas quando essas tinham quatro ou mais vizinhos ortográficos. O segundo experimento desenvolvido visou encorajar um processo de leitura estratégico que enfatizasse o processo de conversão grafema-fonema. Nessa tarefa os participantes também foram expostos a palavras e pseudopalavras, mas dessa vez, deviam responder a ambos os tipos de estímulos. Nesse estudo, os efeitos de N e NF apresentaram um efeito facilitador, ou seja, um aumento no número de vizinhos ortográficos e/ou a presença de um vizinho ortográfico de maior frequência tendeu a diminuir o tempo gasto no reconhecimento de uma palavra. Porém os efeitos de N, NF, ou sua interação, não foram estatisticamente significantes. O terceiro experimento desenvolvido foi uma tarefa de identificação perceptual. Nessa tarefa, as palavras alvo eram apresentadas brevemente (por 60 milissegundos), sendo requerido dos participantes que esses dissessem em voz alta, em um prazo máximo de 3 segundos, qual tinha sido a palavra apresentada. Nesse experimento, um aumento no número de vizinhos ortográficos (N) resultou em uma diminuição na porcentagem de erros de identificação cometidos, enquanto a presença de um vizinho ortográfico de maior frequência de ocorrência (NF) resultou em um aumento na porcentagem de erros. Os fatores N e NF não interagiram nesse experimento. Por fim uma tarefa de categorização semântica foi desenvolvida. Nessa tarefa os participantes deviam pressionar um botão se a palavra alvo representasse o nome de uma cor ou pressionar outro botão se a palavra não representasse o nome de uma cor. As palavras correspondentes à manipulação fatorial de N e NF correspondiam às respostas negativas (não representavam nomes de cores) e não se observou quaisquer efeitos de N ou NF, sendo, apenas, as palavras que representavam cores, reconhecidas mais rapidamente do que as que não representavam. Para melhor avaliar o padrão de resultados desses quatro experimentos, os resultados desse estudo foram combinados com os resultados de outros estudos realizados com falantes do português brasileiro e uma meta-análise foi realizada. Ao se tentar combinar os níveis de significância e a magnitude do efeito de N e NF em todos os estudos (sem considerar o tipo de tarefa utilizada), testes X^2 indicaram que os efeitos dessas variáveis eram heterogêneos. De uma forma geral, foi possível observar dois tipos de agrupamentos de tarefas que não produziram resultados heterogêneos: tarefas que enfatizavam o processo de identificação lexical (tarefas de identificação perceptual; tarefas de decisão lexical e de leitura em voz alta do tipo *Go/NoGo* nas quais as pseudopalavras utilizadas tinham muitos vizinhos ortográficos); e, tarefas que enfatizavam o processamento estratégico (tarefas de decisão lexical onde as pseudopalavras tinham poucos vizinhos ortográficos e tarefas de leitura em voz alta onde os participantes deviam ler palavras e pseudopalavras em uma mesma sessão). Nas tarefas onde a ênfase era na identificação lexical, NF apresentou um efeito inibidor ($r = 0,28$ e $p < 0,01$). Já o efeito de N, só não foi heterogêneo quando a tarefa de identificação perceptual foi excluída do agrupamento. Nessa situação, N apresentou um efeito inibidor ($r = 0,16$ e $p < 0,01$), porém esse foi mais fraco que o efeito de NF. Já nas tarefas que envolviam processamento estratégico, o efeito de NF foi nulo ($p > 0,39$) e o efeito de N foi facilitador ($r = -0,17$; $p < 0,01$). De uma forma geral, pode-se dizer que os resultados dessa tese corroboram a hipótese de que os efeitos de N e NF podem ser modulados pela demanda das tarefas utilizadas para se investigar o acesso lexical e, além disso, colocam em dificuldades modelos de reconhecimento visual de palavras que não têm mecanismos específicos que os permitam prever quais tarefas enfatizam processos estratégicos e quais enfatizam o processo de identificação lexical.

LISTA DE FIGURAS

Figura 1 – Interação entre N (eixo x) e NF (linhas de cor azul e verde)

120

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas para a o índice de familiaridade médio (IFM), frequência de ocorrência por milhão de palavras (FpM) e frequência logarítmica (FLog) de 400 palavras	77
Tabela 2 – Correlações entre diferentes medidas de frequência de ocorrência para uma amostra de 400 palavras	77
Tabela 3 – Exemplo de classificação das línguas europeias de acordo com a profundidade ortográfica (Tabela adaptada de Seymour, 2005)	80
Tabela 4 – Estatísticas descritivas para uma amostra de 8465 palavras de quatro a seis letras retiradas do <i>corpus</i> NILC (2005)	83
Tabela 5 – Correlações entre medidas de vizinhança ortográfica, frequência e número de letras para uma amostra de 8465 palavras do português do Brasil	84
Tabela 6 – Estatísticas descritivas de medidas de vizinhança ortográfica e de frequência de ocorrência para palavras do português do Brasil e do Inglês (adaptado de Andrews, 1997)	86
Tabela 7 – Correlações entre medidas de vizinhança ortográfica e frequência, para palavras de 4 letras, 5 letras (itálico) e 6 letras (negrito) do português brasileiro e do inglês (entre parênteses – adaptado de Andrews, 1997)	88
Tabela 8 – Média e desvio-padrão (entre parênteses) da frequência de ocorrência e das medidas de vizinhança ortográfica dos estímulos experimentais	95
Tabela 9 – Frequência de ocorrência das listas por condição experimental	96
Tabela 10 - Esquema de contrabalanceamento de listas para um design onde ‘N’ e ‘NF’ são fatores com medidas repetidas e ‘D’ e ‘E’ fatores entre sujeitos	109
Tabela 11 – Média e desvio-padrão do TR na tarefa de leitura em voz alta do tipo <i>Go/NoGo</i>	119
Tabela 12 – Média e desvio-padrão do TR na tarefa de leitura em voz alta	132
Tabela 13 – Média e desvio-padrão da porcentagem de erros na tarefa de identificação perceptual	143
Tabela 14 – Média e desvio-padrão do TR na tarefa de categorização semântica	155
Tabela 15 – Magnitude do efeito (geral) e nível de significância de N e NF por tipo de tarefa	164

SUMÁRIO

1 INTRODUÇÃO	14
2 MODELOS DE RECONHECIMENTO VISUAL DE PALAVRAS E O EFEITO DE VIZINHANÇA ORTOGRÁFICA	19
2.1 Modelos de Busca e Verificação	19
2.2 Modelos de Ativação Interativa e Competição	21
2.3 Modelos de Processamento Paralelo e Distribuído	25
2.4 O Modelo SERIOL	27
2.5 Rediscutindo a Definição de Vizinhança Ortográfica	32
3 REVISÃO DA LITERATURA	35
3.1 Tarefa de Decisão Lexical	36
3.2 Tarefa de Identificação Perceptual	47
3.3 Tarefa de Leitura em Voz Alta	53
3.4 Estudos Sobre o Movimento dos Olhos na Leitura	56
3.5 Tarefa de Categorização Semântica	59
3.6 Interpretações dos Resultados das Pesquisas e <i>Rationale</i> do Presente Estudo	64
3.6.1 Objetivos	70
3.6.2 Justificativas	71
4 ANÁLISE DE CORPUS	72
4.1 Estudo de Familiaridade	72
4.1.1 Método	74
4.1.1.1 Participantes	74
4.1.1.2 Material	74
4.1.1.3 Procedimentos	75
4.1.2 Resultados	76
4.1.3 Discussão	78
4.2 A Estrutura de Vizinhança Ortográfica das Palavras do Português Brasileiro	78
4.2.1 Método	81
4.2.1.1 Material	82
4.2.1.2 Procedimentos	82
4.2.2 Resultados	83
4.2.3 Discussão	89

5 MÉTODO	92
5.1 Participantes	92
5.2 Material	93
5.3 Procedimentos	97
5.4 Análise Estatística e <i>Design</i> Experimental	99
5.4.1 O Tempo de Reação como uma Variável Dependente	100
5.4.2 A Linguagem como um Efeito Fixo: Problemas e Soluções	101
5.4.2.1 O que são um Efeito Fixo e um Efeito Aleatório?	102
5.4.2.2 Qual o Modelo Estatístico Adequado para Analisar os Dados desses Experimentos?	104
5.4.2.3 Lidando com a Falácia da Linguagem como um Efeito Fixo	106
6 TAREFA DE LEITURA EM VOZ ALTA DO TIPO <i>GO/NOGO</i>	110
6.1 Processos Presentes na Tarefa de Leitura em Voz Alta do tipo <i>Go/NoGo</i>	111
6.2 Método	115
6.2.1 Participantes	115
6.2.2 Material	115
6.2.3 Procedimentos	117
6.3 Resultados	118
6.4 Discussão	122
7 TAREFA DE LEITURA EM VOZ ALTA	125
7.1 Tarefa de Leitura em Voz Alta com Palavras e Pseudopalavras Misturadas	126
7.2 Método	128
7.2.1 Participantes	128
7.2.2 Material	128
7.2.3 Procedimentos	130
7.3 Resultados	131
7.4 Discussão	133
8 TAREFA DE IDENTIFICAÇÃO PERCEPTUAL	136
8.1 Processos Presentes na Tarefa de Identificação Perceptual	137
8.2 Método	139
8.2.1 Participantes	139
8.2.2 Material	140
8.2.3 Procedimentos	141
8.3 Resultados	142

8.4 Discussão	144
9 TAREFA DE CATEGORIZAÇÃO SEMÂNTICA	147
9.1 Processos Presentes na Tarefa de Categorização Semântica	148
9.2 Método	151
9.2.1 Participantes	151
9.2.2 Material	151
9.2.3 Procedimentos	153
9.3 Resultados	155
9.4 Discussão	156
10 META-ANÁLISE E DISCUSSÃO GERAL	159
10.1 Meta-Análise	159
10.1.1 Critérios de Inclusão	160
10.1.2 Padronização dos Resultados	160
10.1.3 Busca por Moderadores	161
10.1.4 Resultados	163
10.1.5 Discussão	167
10.2 Discussão Geral	169
11 CONCLUSÃO	176
REFERÊNCIAS	179
GLOSSÁRIO	192
APÊNDICE A – PALAVRAS CORRESPONDENTES À MANIPULAÇÃO FATORIAL DE N E NF	194
APÊNDICE B – PSEUDOPALAVRAS E PALAVRAS UTILIZADAS COMO ESTÍMULOS COMPLEMENTARES	195

1 INTRODUÇÃO

Pode-se dizer que, pelo menos, dois processos estão envolvidos na leitura: o acesso lexical e a compreensão. O acesso lexical se refere ao acesso à representação mental das palavras que são reconhecidas em um texto e a compreensão ao conjunto de processos inter-relacionados por meio dos quais o leitor constrói uma representação do significado do texto (Perfetti, 1985).

O foco desse trabalho é sobre o processo de acesso lexical e sendo assim, mais uma distinção torna-se relevante. De acordo com Monsell, Doyle e Haggard (1989), pode-se distinguir *transcodificação lexical* de *identificação lexical*. A transcodificação lexical refere-se à geração de um código no sistema semântico ou fonológico a partir de um *input*. Uma etapa desse processo de transcodificação seria a seleção da forma no léxico mental que melhor corresponde ao estímulo alvo e esse processo de seleção é denominado de *identificação lexical*. Nesse sentido, uma das principais tarefas que o nosso sistema cognitivo desempenha durante o acesso lexical, é a seleção da entrada lexical correta diante das diversas entradas que podem ser ortograficamente similares ao estímulo alvo. Dentro da tradição de pesquisas sobre o reconhecimento visual de palavras, o efeito que as palavras similares em termos ortográficos desencadeiam no processo de reconhecimento de um estímulo alvo, é conhecido por *efeito de vizinhança ortográfica* e tem sido questão bastante estudada na literatura sobre o tópico (Andrews, 1989, 1992, 1997; Carreiras, Perea & Grainger, 1997; Coltheart, Davelaar, Jonasson & Besner, 1977; Grainger, O'Regan, Jacobs & Segui, 1989; Huntsman & Lima, 1996, 2002; Laxon, Masterson, Pool & Keating, 1992; Perea, Carreiras & Grainger, 2004; Perea & Pollatsek, 1998; Perea & Rosa, 2000, Sears, Hino & Lupker, 1995; Snodgrass & Minzer, 1993; entre outros).

A principal tentativa de se definir as palavras que seriam similares em termos ortográficos a uma palavra alvo consiste em considerar como vizinhos ortográficos da palavra alvo, qualquer palavra que possa ser criada ao se mudar uma letra dessa, enquanto se preservam as posições das demais letras. Desse modo, a palavra ‘alma’ teria cinco vizinhos ortográficos (‘arma’, ‘asma’, ‘alga’, ‘alça’ e ‘alta’). Essa pode ser considerada a definição mais tradicional de uma palavra vizinha (ou similar em sua forma ortográfica) e é atribuída a Coltheart e cols. (1977). O número de palavras que pode ser criado utilizando esse processo proposto por Coltheart e cols. é chamado de medida N (de *neighbourhood*, em inglês) e se refere ao número de vizinhos ortográficos da palavra alvo. Além dessa medida, nos estudos sobre o efeito de vizinhança ortográfica, também se têm utilizado a medida NF (de *neighbourhood frequency*, em inglês) que se refere à frequência de ocorrência dos vizinhos da palavra alvo, ou seja, se ela possui vizinhos de frequência maior do que a dela ou não (Grainger & cols., 1989). No caso de NF, pode-se dizer que a palavra ‘anão’ tem um vizinho ortográfico de maior frequência que é a palavra ‘ação’. A operacionalização dessas medidas (N e NF) tem o intuito de permitir aos pesquisadores estabelecer se o efeito de vizinhança ortográfica seria facilitador ou inibidor. Assim sendo, se o efeito de vizinhança ortográfica for facilitador, então o reconhecimento de palavras com um grande número de vizinhos ortográficos ou com, pelo menos, um vizinho de maior frequência será realizado com mais precisão e rapidez; se o efeito for inibidor, o contrário se verificará, ou seja, as palavras nessa condição serão reconhecidas mais lentamente e com menos precisão.

Atualmente as pesquisas sobre o efeito de vizinhança ortográfica têm se revestido de grande importância teórica (Andrews, 1997). Uma das razões para isso é que muitos modelos de reconhecimento visual de palavras postulam a existência de um “dicionário mental” onde o leitor armazena as representações lexicais das formas ortográficas das palavras que são lidas (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Grainger & Jacobs, 1996; McClelland &

Rumelhart, 1981; Murray & Forster, 2004; Paap & Johansen, 1994). Os estudos sobre o efeito de vizinhança ortográfica são importantes porque podem ajudar a revelar a dinâmica do processo de acesso a essas representações lexicais (Perea & Rosa, 2000). Por exemplo, esse processo implicaria em uma espécie de “competição” entre as unidades lexicais que são similares ao estímulo alvo, como acontece nos modelos de McClelland e Rumelhart (1981), Grainger e Jacobs (1996) e Coltheart e cols. (2001)? Ou esse processo se assemelharia a uma busca serial em um conjunto de unidades lexicais similares ao estímulo alvo, como proposto por Murray e Forster (2004) e Paap e Johansen (1994)? Ou ainda, de forma mais radical, deve-se abandonar a idéia de um léxico ortográfico composto por representações discretas de palavras, em prol da idéia de que as representações dessas palavras se assemelhariam a padrões de ativação distribuídos em uma rede de unidades mais simples, como nos modelos de Seidenberg e McClelland (1989) e de Plaut, McClelland, Seidenberg e Patterson (1996)?

Os estudos sobre o efeito de vizinhança ortográfica ajudariam a responder tais questões, porque certos efeitos seriam mais naturalmente acomodados por um modelo de reconhecimento visual de palavras do que por outro. Por exemplo, um efeito inibidor seria mais naturalmente acomodado pelos modelos de McClelland e Rumelhart (1981), Grainger e Jacobs (1996) e Coltheart e cols. (2001), enquanto um efeito facilitador seria mais natural nos modelos de Seidenberg e McClelland (1989) e Plaut e cols. (1996), conforme demonstrado nas simulações computacionais de Grainger e Jacobs (1996) e de Sears, Hino e Lupker (1999).

Dada a sua importância teórica, diversos estudos sobre o efeito de vizinhança ortográfica têm sido realizados com falantes de diferentes línguas, porém os resultados têm sido um tanto conflitantes (os trabalhos de Andrews, 1997 e de Perea & Rosa, 2000 trazem uma revisão dos estudos e ilustram essa situação de conflito nos resultados). Uma explicação propõe que diferenças entre as línguas nas quais os estudos foram conduzidos poderiam

modular os efeitos de N e NF (Andrews, 1997), enquanto outra propõe que o efeito dessas variáveis poderia ser modulado de acordo com as demandas das diferentes tarefas utilizadas para investigar o processo de acesso lexical (Perea & Rosa, 2000). De forma mais específica, essa última hipótese propõe que em tarefas onde a ênfase é no processo de acesso lexical os efeitos de N e NF seriam inibidores, já em tarefas onde o uso de estratégias fosse enfatizado, o efeito dessas variáveis poderia ser nulo ou até mesmo facilitador.

Tendo essas considerações em mente o objetivo mais geral do presente trabalho é investigar o efeito de vizinhança ortográfica no reconhecimento visual de palavras em adultos falantes do português do Brasil. Outro objetivo é observar se os efeitos de N e de NF podem, de fato, variar de acordo com a ênfase das tarefas utilizadas na pesquisa em processos de acesso lexical ou em algum tipo de processamento estratégico. Para tal, essa tese está organizada em 11 capítulos, incluindo o presente. O segundo capítulo é dedicado a uma discussão dos atuais modelos de reconhecimento visual de palavras, tendo como foco suas predições quanto ao efeito de vizinhança ortográfica. Esse capítulo é finalizado com uma discussão do próprio conceito de vizinhança ortográfica e com um conjunto de justificativas para as medidas de vizinhança ortográfica adotadas nesse estudo. O terceiro capítulo apresenta uma revisão da literatura sobre os efeitos de vizinhança ortográfica, juntamente com as principais hipóteses propostas para explicar esses efeitos. Esse capítulo apresenta também, pormenorizadamente, os objetivos gerais e específicos dessa tese, bem como, suas justificativas. O quarto capítulo, ‘Análise de Corpus’, relata diversas análises realizadas visando auxiliar na escolha dos estímulos experimentais e também serve de suporte empírico para as discussões acerca do conceito de vizinhança ortográfica realizadas no segundo capítulo. Além disso, esse capítulo traça uma comparação entre a estrutura da vizinhança ortográfica das palavras do português e a estrutura da vizinhança ortográfica das palavras da língua inglesa. O quinto capítulo versa sobre os procedimentos comuns aos experimentos

desenvolvidos. Já os capítulos de número 6, 7, 8 e 9, descrevem os 4 experimentos realizados nessa tese e suas implicações. O décimo capítulo apresenta uma meta-análise dos resultados desses experimentos e de outros desenvolvidos com falantes do português brasileiro, avalia a adequação dos diferentes modelos de reconhecimento visual de palavras aos dados do português brasileiro e faz uma discussão geral dos achados dessa tese. Por fim, o capítulo de número 11, 'Conclusão', sumaria os principais achados dessa tese.

2 MODELOS DE RECONHECIMENTO VISUAL DE PALAVRAS E O EFEITO DE VIZINHANÇA ORTOGRÁFICA

De acordo com os objetivos do presente trabalho, parece profícuo organizar a apresentação dos modelos de reconhecimento visual de palavras em três classes: modelos de busca e verificação; modelos de ativação interativa e competição; modelos de processamento paralelo e distribuído (PPD); e, modelo SERIOL. Essa organização visa facilitar a exposição desses modelos e de suas predições no que diz respeito ao efeito de vizinhança ortográfica. Não se quer com isso, dizer que os modelos de uma mesma classe são idênticos. Não é também objetivo do presente trabalho comparar esses modelos ou seu desempenho em outros aspectos da literatura sobre o reconhecimento visual de palavras (para esse tipo de comparação ver, por exemplo, o trabalho de Roazzi, Justi & Justi, 2008).

2.1 Modelos de Busca e Verificação

Os principais modelos dessa classe são o Modelo *Bin* (MB – "*bin model*" em inglês – de Forster, 1976; 1992; Murray & Forster, 2004) e o Modelo de Verificação (MV – de Paap, Newsome, McDonald & Schvaneveldt, 1982; Paap & Johansen, 1994). No Modelo *Bin*, presume-se que o acesso lexical envolva dois estágios separados e que o sistema de acesso ortográfico seja organizado em conjuntos de *Bins*. O sistema de acesso ortográfico é responsável, no modelo, por armazenar a informação sobre a ortografia das palavras. Esse sistema se divide em subconjuntos de *Bins*. Cada *Bin* armazena um grupo de entradas lexicais

(“palavras”) que tenham propriedades ortográficas similares. Dentro de cada *Bin*, essas entradas lexicais são ordenadas em termos da frequência de ocorrência das palavras, sendo as mais frequentes, as primeiras da lista.

Nesse modelo, o acesso lexical acontece da seguinte maneira: diante de uma palavra alvo, o sistema localiza o *Bin* que poderia conter essa palavra e então, inicia-se uma comparação seqüencial entre esse *input* e as entradas lexicais contidas no *Bin*. Como o *Bin* é organizado de acordo com a frequência de ocorrência das palavras, as comparações sempre começam com a entrada lexical mais freqüente e terminam com a menos freqüente dentro desse. O primeiro estágio de acesso lexical termina quando a entrada lexical correspondente à palavra alvo é encontrada dentro do *Bin*. A partir daí, um segundo estágio começa, no qual a informação dessa entrada é utilizada para se conseguir uma descrição completa da palavra alvo (informação semântica, classe sintática, etc.).

No que diz respeito ao efeito de vizinhança ortográfica, apenas o primeiro estágio é relevante, pois é nele em que ocorre a identificação da palavra alvo com base em sua ortografia. É possível analisar, então, quais seriam as predições desse modelo para as variáveis N e NF. No que tange a variável NF, pode-se dizer que seu efeito, é parasitário ao efeito da variável frequência de ocorrência. Isso é, uma vez que a busca dentro de cada *Bin* é ordenada pela frequência, as palavras com vizinhos de maior frequência serão sempre comparadas mais tarde do que as palavras que não tenham vizinhos de maior frequência, pois as últimas serão sempre as primeiras do *Bin*. Nesse sentido o modelo *Bin* prevê um efeito inibidor da variável NF no reconhecimento visual de palavras. Já o número de vizinhos (N), não teria qualquer efeito dentro desse modelo, pois o que influencia o tempo de reconhecimento de uma palavra alvo é a frequência dessa em relação aos seus vizinhos. Não importa se essa palavra tem dois ou dez vizinhos, se ela tem uma frequência de ocorrência

maior que a dos seus vizinhos, sua entrada lexical será sempre comparada com o *input* primeiro.

Considerando-se as predições que dizem respeito ao efeito de vizinhança ortográfica, o Modelo de Verificação (Paap & cols., 1982; Paap & Johansen, 1994) faz predições idênticas às do Modelo *Bin*. No Modelo de Verificação, diante de uma palavra alvo, um conjunto de entradas lexicais é selecionado e então, um processo compara seqüencialmente, em ordem de freqüência, cada uma dessas entradas lexicais com a palavra alvo. Pode-se dizer que a diferença entre esses modelos é que no Modelo de Verificação, o conjunto de entradas lexicais a ser vasculhado e a rotina de comparação são computados *on-line* (no momento da busca), enquanto no Modelo *Bin*, as entradas lexicais já estão organizadas em *Bins* e ordenadas pela freqüência de ocorrência. No entanto, como já ressaltado, essas diferenças não fazem com que os modelos tenham predições diferentes sobre o efeito de vizinhança ortográfica.

2.2 Modelos de Ativação Interativa e Competição

Uma classe de modelos de reconhecimento visual de palavras muito influente atualmente é derivada do Modelo de Ativação Interativa e de Competição (AIC) de McClelland e Rumelhart (1981) e Rumelhart e McClelland (1982). Nessa classe, incluem-se o Modelo Múltiplo de Leitura (ML) de Grainger e Jacobs (1996) e o Modelo de Dupla Rota em Cascata (DRC) de Coltheart e cols. (2001). Como o modelo de Coltheart e cols. pode implementar as diferentes estratégias propostas pelo modelo de Grainger e Jacobs e ter uma aplicação mais geral que esse, apenas o modelo DRC será discutido.

O modelo de Coltheart e cols. (2001) baseia-se no uso de duas rotas para processar a leitura, isto é, a transformação de um *input* ortográfico em um *output* semântico e/ou fonológico. A primeira rota seria a rota lexical que pode ser subdividida em lexical semântica (ainda não implementada) e lexical não-semântica. A principal função dessa rota é permitir a leitura de palavras familiares e palavras irregulares (palavras contendo uma ou mais correspondência grafema-fonema irregular) que estejam armazenadas em nosso dicionário interno (léxico de *input* ortográfico). Essa parte do modelo é implementada tomando-se como referência a arquitetura do modelo AIC, podendo ser considerada uma generalização deste. De fato, pode-se dizer que a rota lexical do modelo DRC coincide com a arquitetura dos modelos AIC e ML.

O modelo DRC postula que diante de uma palavra alvo, há três níveis distintos de análise que são levados em consideração: nível de característica, nível da letra e nível da palavra. O nível da característica analisa os diferentes padrões visuais existentes no estímulo apresentado que, por meio de uma ligação unidirecional com o nível da letra, ativa as letras correspondentes. Esses dois níveis são comuns às duas rotas do modelo. Já o nível da palavra pertence apenas à rota lexical e funciona da seguinte maneira: ativada uma letra (nível da letra), todas as palavras que compartilham essa letra, na mesma posição que a do estímulo alvo, serão também ativadas (nível da palavra) no léxico de *input* ortográfico. Isso gera uma espécie de competição entre esse conjunto de palavras, que deve ser solucionada por meio do mecanismo de *inibição lateral*. Nesse mecanismo, cada palavra ativada tenta inibir as outras palavras do conjunto reduzindo-lhes o nível de ativação. A cada padrão visual encontrado no *input*, as palavras que o contém recebem ativação proveniente do nível da letra e inibem com intensidade proporcional ao seu nível de ativação as palavras competidoras. Esse processo se repete, em condições normais, até que a palavra correspondente “vença” suas competidoras e seja identificada no léxico de *input* ortográfico (Coltheart e cols., 2001). Paralelamente, as

palavras no léxico de input ortográfico repassam ativação às unidades do *léxico de output fonológico* (que armazena a representação fonológica das palavras) e essas unidades, por sua vez, repassam ativação ao *nível de fonema*, iniciando assim a ativação de fonemas para gerar a pronúncia da palavra alvo.

Nesse momento, ao se considerar o processo de identificação de uma palavra alvo no léxico de *input* ortográfico, pode-se perceber que o efeito de vizinhança ortográfica é inibidor. Afinal, as palavras que têm muitos vizinhos ou vizinhos de maior frequência recebem mais inibição lateral de seus competidores, o que aumenta o tempo necessário para seu reconhecimento. No entanto, talvez o número de vizinhos de maior frequência (NF) seja a variável mais importante nesse modelo, pois pode ocorrer que uma palavra mais frequente iniba, rapidamente, suas competidoras (N), não sofrendo assim, tanta inibição.

A outra rota do modelo é a rota fonológica. Sua principal função é possibilitar a leitura de palavras desconhecidas e de pseudopalavras (seqüências de letras, construídas com estruturas ortográficas possíveis na língua em questão, mas não associadas a nenhum significado) e isso é feito por meio do uso de regras de conversão grafema-fonema da seguinte forma: ao receber um *input* (uma seqüência de letras) proveniente do nível da letra, busca-se na memória uma regra de conversão que seja adequada à primeira letra desse *input* que é então, convertida em seu fonema correspondente. A seguir, a próxima letra do *input* torna-se disponível ao processo de conversão grafema-fonema, que usa suas regras para traduzir essas duas letras em um fonema (no caso dos grafemas, como por exemplo, ‘ch’) ou em uma seqüência de dois fonemas (no caso de ‘cr’, por exemplo). Desse modo, essa rota trabalha, letra a letra, da esquerda para a direita, sendo seu funcionamento mais lento do que o da rota lexical (Coltheart & cols., 2001).

No caso dessa rota, a ênfase não é na identificação inequívoca de uma palavra, mas sim, na geração dos fonemas necessários à pronúncia de um estímulo-alvo (palavra ou

pseudopalavra). Nesses casos, o efeito do número de vizinhos ortográficos tende a ser facilitador e pode ocorrer em duas partes do modelo, conforme demonstrado na simulação de Reynolds e Besner (2002). Em um primeiro momento, o efeito facilitador poderia ser proveniente da reverberação das ativações do léxico de *input* ortográfico ao nível da letra. Dessa forma, conjuntos de letras que ativam várias palavras também recebem ativação dessas palavras, facilitando sua identificação, isso é, o estímulo alvo e seus vizinhos ortográficos tendem a facilitar o reconhecimento das letras que compartilham, graças à reverberação da ativação. Já em um segundo momento, o efeito do número de vizinhos ortográficos seria proveniente da ativação que o léxico de *output* fonológico repassa para o nível de fonema, iniciando a ativação de fonemas que seriam comuns às palavras vizinhas. Nesse caso, palavras com muitos vizinhos ortográficos (léxico de *input* ortográfico), também tendem a ter muitos vizinhos fonológicos (léxico de *output* fonológico) e esses tenderão a auxiliar a ativação dos fonemas que têm em comum (nível de fonema). Como o funcionamento das duas rotas é paralelo e em cascata, a rota fonológica pode-se beneficiar da reverberação da ativação que facilita a identificação das letras e da pré-ativação de alguns fonemas ao gerar a pronúncia do estímulo-alvo, sem sofrer a influência negativa do processo de inibição lateral que ocorre na rota lexical. É interessante observar que, nessas situações, pode-se prever um efeito facilitador de N (Reynolds & Besner, 2002), no entanto, não necessariamente de NF, porque ter um vizinho ortográfico de maior frequência pode não contribuir tanto para a ativação de letras e/ou fonemas compartilhados quanto ter vários vizinhos ortográficos.

Em suma, essa classe de modelos parece prever que em condições nas quais a identificação inequívoca de uma palavra alvo é necessária (uso da rota lexical), o efeito de NF deve ser inibidor. Porém, quando o importante é apenas gerar os fonemas necessários à pronúncia de uma palavra alvo, o efeito de N deve ser facilitador. No entanto, é ainda importante ressaltar que existem situações em que mesmo a rota lexical pode apresentar um

efeito facilitador da vizinhança ortográfica. Essas condições envolvem a alteração de parâmetros dessa rota para simular uma estratégia de “adivinhação rápida”. As condições em que essa estratégia pode ocorrer e como os modelos dessa classe podem acomodá-la foram propostas inicialmente por Grainger e Jacobs (1996) em seu modelo ML e, posteriormente, incorporadas por Coltheart e cols. (2001) ao modelo DRC. Essas situações serão discutidas no capítulo ‘Revisão da Literatura’.

2.3 Modelos de Processamento Paralelo e Distribuído

Outra classe de modelos de reconhecimento de palavras encontrada na literatura é a dos modelos de Processamento Paralelo Distribuído (PPD). O primeiro modelo dessa classe foi proposto por Seidenberg e McClelland (1989) e, posteriormente, teve seu esquema representacional modificado por Plaut e cols. (1996).

Esses modelos sugerem que aprender a ler palavras envolve aprender a computar de um estímulo visual os seus códigos: ortográfico; fonológico; e, semântico (Seidenberg & McClelland, 1989; Plaut & cols., 1996). Para isso, eles contam com três camadas de unidades de processamento. Uma camada para unidades de *input*, uma camada de unidades de *output* e uma camada de unidades escondidas (camada intermediária). A camada de unidades escondidas teria a função de ajustar os pesos nas conexões entre as unidades por meio de uma regra de aprendizagem. A pronúncia de uma palavra envolveria o mapeamento de seu código ortográfico ao seu código fonológico e seria efetivada por meio de padrões de ativação distribuídos nas unidades constituintes da rede.

Ao contrário do modelo DRC, estes modelos não fazem uso de representações localizadas, isso é, uma unidade específica para representar cada palavra, como ocorre no léxico de *input* ortográfico do modelo DRC. Por serem modelos do desenvolvimento, pode-se dizer que suas “arquiteturas” se auto-organizam à medida que os modelos aprendem por meio do algoritmo de retropropagação (regra de aprendizagem do modelo). Outra diferença entre esses modelos e o modelo DRC, é que nos modelos PPD não existem regras explícitas para a conversão grafema-fonema. Nesses modelos, a leitura de pseudopalavras é efetuada pelo mesmo mecanismo responsável pela leitura de palavras reais; assim, muitas vezes, as pseudopalavras são lidas por analogia às palavras reais. Por fim, é importante considerar que recentemente Harm e Seidenberg (2004) desenvolveram uma versão do modelo PPD na qual um sistema semântico foi implementado. No entanto, infelizmente, os autores não desenvolveram nenhuma simulação ou predição teórica acerca de quais seriam os efeitos de N e NF dado esse sistema semântico, tornando difícil saber quais são as previsões do modelo para essas variáveis quando o acesso à semântica é uma parte primordial da tarefa. Assim sendo, serão discutidos aqui, os modelos PPD apenas em situações nas quais já existem predições quanto aos efeitos de N e NF.

No que tange a gerar a pronúncia de uma palavra, o efeito de vizinhança pode ser interpretado, dentro dessa classe de modelos, como facilitador. O efeito seria facilitador porque as unidades responsáveis por gerar a pronúncia de uma palavra alvo, durante o processo de treinamento, beneficiar-se-iam também do treino com as palavras que fossem ortograficamente similares à palavra alvo. Assim, durante o treinamento, palavras com muitos vizinhos fortaleceriam as conexões responsáveis por gerar a pronúncia dos grafemas compartilhados nessas palavras, e, durante a pronúncia, as conexões mais fortalecidas resultariam em mais precisão e rapidez no processamento das palavras que as compartilham (Seidenberg & McClelland, 1989).

De acordo com Seidenberg e McClelland (1989), em tarefas nas quais a demanda é distinguir uma palavra de uma pseudopalavra (como na tarefa de decisão lexical que será discutida no capítulo ‘revisão da literatura’) o efeito de vizinhança ortográfica também seria facilitador. Isso aconteceria porque para distinguir uma palavra de uma pseudopalavra nessas tarefas, o modelo traçaria uma comparação entre o *feedback* proveniente das unidades escondidas e o padrão de ativação das unidades ortográficas. Assim, o modelo classificaria como pseudopalavras, estímulos que gerassem maior discrepância entre esses dois padrões de ativação e classificaria como palavras, estímulos que gerassem pouca ou nenhuma discrepância. Do mesmo modo como na leitura, durante o treinamento, as unidades responsáveis por gerar o padrão de ativação de uma palavra alvo (unidades escondidas) se beneficiariam também do treino com palavras similares a essa palavra alvo. Desse modo, para palavras com muitos vizinhos, o *feedback* proveniente das unidades escondidas seria mais preciso e próximo do padrão do *input* nas unidades ortográficas. É importante ressaltar que um estudo de simulação computacional envolvendo os modelos de Seidenberg e McClelland (1989) e de Plaut e cols. (1996), desenvolvido por Sears, Hino e Lupker (1999), confirmou o efeito facilitador, tanto de N, quanto de NF nesses modelos.

2.4 O Modelo SERIOL

Como ressaltam Roazzi e cols. (2008), a leitura, para as pessoas alfabetizadas, é um processo tão automático que elas nem se dão conta de que reconhecem muito rapidamente uma mesma palavra *independente* DE **variações em suas características físicas**. Isso é um indício de que um dos primeiros feitos do nosso sistema cognitivo relacionado à leitura é a

capacidade de transformar um conjunto de padrões físicos em uma representação abstrata da ordem das letras (que independe de sua forma física). Todos os modelos discutidos até aqui assumem esse processo como dado e iniciam suas formulações a partir de um nível representacional abstrato. O modelo SERIOL desenvolvido por Whitey (2001; 2004; Whitney & Lavidor, 2005; Whitney & Cornelissen, 2008) tenta modelar de forma mais explícita como um conjunto de padrões projetados em nossas retinas pode ser transformado em uma representação abstrata da ordem das letras.

O modelo SERIOL propõe que o reconhecimento visual de uma palavra por meio do sistema lexical envolve cinco níveis de processamento: *bordas*; *características*; *letras*; *bigramas abertos*; e, *palavras*. Os dois primeiros níveis estão ligados diretamente à topografia da imagem na retina, enquanto os três últimos são abstratos (Whitney, 2001). O nível mais básico de processamento é o nível das bordas. Nesse nível, o grau de ativação dos detectores diminui à medida que a distância do ponto de fixação aumenta (devido à diminuição da concentração de cones). Isso é conhecido no modelo como *gradiente de acuidade*. Por exemplo, considerando-se que o sinal ‘+’ indica o ponto de fixação do indivíduo, para o estímulo ‘BO+LA’, os detectores correspondentes ao ‘O’ e ao ‘L’ teriam maior grau de ativação do que os detectores correspondentes ao ‘B’ e ao ‘A’.

O próximo nível de processamento no modelo é o nível das características que tem unidades responsáveis por detectar padrões que compõem letras (padrões sub-ortográficos, como por exemplo: ‘/’, ‘-’, ‘\’). Uma das principais funções desse nível de processamento é converter o gradiente de acuidade em um *gradiente de localização*. No gradiente de localização, o índice de ativação das unidades responsáveis por detectar os padrões correspondentes às letras decresce da esquerda para a direita de forma que as unidades responsáveis por detectar os padrões da letra mais à esquerda da palavra são as mais fortemente ativas e as responsáveis por detectar os padrões da letra mais à direita são as

menos ativas. O modelo SERIOL propõe que essa conversão de gradientes ocorre da seguinte maneira: a extração das características ocorre independentemente em cada hemisfério cerebral (as características correspondentes as letras à esquerda do ponto de fixação são extraídas no hemisfério direito e as características correspondentes as letras à direita do ponto de fixação são extraídas no hemisfério esquerdo); no hemisfério esquerdo, os gradientes de acuidade e de localização estão na mesma direção (ambos decrescem da esquerda para a direita), porém, no hemisfério direito, o gradiente de acuidade encontra-se em direção oposta ao de localização (o gradiente de acuidade aumenta da esquerda para a direita). Três mecanismos são responsáveis pela conversão do gradiente de acuidade no de localização: primeiro, as conexões excitatórias que vão do nível das bordas ao nível das características são mais fortes no hemisfério direito; segundo, no nível das características, os detectores à esquerda inibem os detectores à direita, reduzindo-lhes o nível de ativação (o que permite a inversão do gradiente de acuidade no hemisfério direito); por fim, quando a informação processada no hemisfério direito é transferida para o hemisfério esquerdo, os detectores de característica do hemisfério esquerdo são inibidos (isso juntamente com a maior ativação *bottom-up* dos detectores de característica do hemisfério direito, permite que as características processadas no hemisfério direito estejam mais ativas do que as processadas no esquerdo). Desse modo, o produto final desse processo é um gradiente de localização que decresce da letra mais à esquerda para a mais à direita.

O próximo nível de processamento no modelo é o nível das letras. Esse nível é responsável por computar uma representação abstrata da ordem das letras. Graças ao cômputo do gradiente de localização, a unidade (no nível das letras) que representa a letra na primeira posição da esquerda para a direita recebe o maior nível de ativação, enquanto a unidade representando a segunda letra recebe o segundo maior nível e assim por diante (Whitney, 2001). Nesse momento, o modelo SERIOL propõe que a representação abstrata da ordem das letras é computada graças a um ciclo de disparos em série que codificam temporalmente a

ordem das letras. Nesse ciclo de disparos, a unidade mais ativa dispara primeiro, depois a segunda mais ativa e assim por diante, sendo, dessa forma, o gradiente de localização convertido em um *gradiente posicional*. O gradiente posicional serve de *input* para as rotas lexical e sublexical do modelo, porém, até o presente momento, o modelo SERIOL apresenta uma descrição detalhada apenas da rota lexical (Whitney & Cornelissen, 2008).

A rota lexical inicia-se no nível de processamento conhecido como bigramas abertos (Grainger & Whitney, 2004). Nesse nível de processamento, a representação temporal gerada pelo gradiente posicional no nível das letras é convertida em uma representação atemporal e contextual da ordem das letras. Isso ocorre, porque as unidades nesse nível de processamento são especializadas em reconhecer pares de letras que disparam em uma ordem particular, mesmo que não sejam estritamente contíguas. Por exemplo, no nível das letras, o estímulo ‘BOLA’ implicaria no disparo seqüencial das unidades responsáveis por reconhecer as letras ‘B’, ‘O’, ‘L’ e ‘A’ nessa ordem, e, no nível dos bigramas abertos, essa ordem de disparo seria representada pelo seguinte conjunto de bigramas: ‘BO’, ‘BL’, ‘BA’, ‘OL’, ‘OA’, ‘LA’, além de ‘*B’ e ‘A*’ que representam o início e fim da seqüência. O nível de ativação de cada um desses bigramas depende diretamente da distância temporal entre o disparo de suas letras constituintes. Dessa forma, mantendo-se o exemplo de ‘BOLA’, os bigramas mais ativos seriam: ‘BO’, ‘OL’, ‘LA’ (Whitney & Cornelissen, 2008).

No nível das palavras, a ativação dos bigramas abertos serve de *input* para as unidades do léxico que representam as palavras que os contém. A ativação que essas unidades recebem é proporcional ao nível de ativação dos bigramas e ao peso das conexões que existem entre eles e as unidades que representam as palavras no léxico. É importante se considerar também que no nível das palavras, de forma análoga ao modelo IAC (McClelland & Rumelhart, 1981), o modelo SERIOL também propõe um mecanismo de *feedback* entre o nível das palavras e o nível das letras e um mecanismo de inibição lateral entre as palavras do léxico,

porém, no caso desse último mecanismo, o modelo propõe que a inibição é mais forte entre palavras mais similares do que entre palavras menos similares ortograficamente (Whitney & Cornelissen, 2008).

Tendo essas considerações em mente, é possível imaginar agora as predições do modelo SERIOL para os efeitos de vizinhança ortográfica. Whitney (2004; Whitney & Lavidor, 2005) tem proposto que o efeito de vizinhança ortográfica depende de uma interação entre o mecanismo de *feedback* que existe entre o nível das palavras e o nível das letras e o mecanismo de inibição lateral que existe no nível das palavras. Para Whitney e Lavidor, quando existem muitos vizinhos, mas nenhum deles é muito similar à palavra alvo, o mecanismo de *feedback* deve dominar o processo, pois a inibição lateral não é tão forte a ponto de cancelar os benefícios de uma ativação acelerada das letras e, desse modo, o resultado esperado é um efeito facilitador do número de vizinhos ortográficos. Já quando ocorre menos *feedback* (o número de vizinhos é pequeno) e a similaridade ortográfica entre o vizinho e a palavra alvo é maior (talvez pela presença de um vizinho de maior frequência), o mecanismo de inibição lateral domina o processo e o resultado predito é um efeito inibidor.

Uma vez que, para uma palavra alvo qualquer, um vizinho ortográfico de maior frequência de ocorrência pode ser considerado um competidor mais forte (por ser mais facilmente ativado no léxico) do que um vizinho ortográfico de menor frequência de ocorrência, pode-se considerar que o primeiro seja um vizinho mais similar à palavra alvo, afinal, é necessário que essa iniba mais fortemente a um vizinho mais ativo para ser reconhecida do que a um vizinho pouco ativo. Assim sendo, pode-se dizer que o modelo SERIOL prediz um efeito facilitador de N (quando não existem vizinhos ortográficos de maior frequência de ocorrência, pois nessa situação a similaridade ortográfica entre os vizinhos e a palavra alvo é menor e isso permite que o mecanismo de *feedback* domine o processo) e um efeito inibidor de NF (quando não existem muitos vizinhos ortográficos, pois,

nesse caso, a operação do mecanismo de *feedback* é reduzida, permitindo ao mecanismo de inibição lateral o domínio do processo).

2.5 Rediscutindo a Definição de Vizinhança Ortográfica

Pode-se notar que um fator crucial nessa discussão dos modelos e de suas predições quanto ao efeito de vizinhança ortográfica é o grau de similaridade entre os vizinhos e a palavra alvo. De uma forma geral, Grainger e Whitney (2004) têm argumentado que devido ao nível representacional dos bigramas abertos a definição de vizinhança ortográfica de Coltheart e cols. (1977) não corresponde ao grau máximo de similaridade ortográfica entre duas palavras no modelo SERIOL. Por exemplo, de acordo com o modelo SERIOL, a palavra ‘saco’ seria mais ortograficamente similar à palavra ‘asco’ do que à palavra ‘suco’, mesmo ‘asco’ não sendo um vizinho ortográfico de ‘saco’ de acordo com a definição de Coltheart e cols. (1977). Isso ocorre, porque ao se levar em consideração os bigramas abertos, a palavra ‘asco’ compartilha 75% desses bigramas com a palavra ‘saco’ (os bigramas: AC, AO, SC, SO, CO, O*) enquanto a palavra ‘suco’ compartilha 62,5% (os bigramas: SC, SO, CO, *S, O*). Dessa forma, o modelo SERIOL prevê uma definição de vizinhança ortográfica mais ampla que a de Coltheart e cols., pelo menos, uma que também considere vizinhos ortográficos formados pela transposição de duas letras (*transposed letter neighbors* – TLN), como no caso de ‘asco’ e ‘saco’ (Grainger & Whitney, 2004; Whitney & Lavidor, 2005). Existe, de fato, alguma evidência de que palavras que têm vizinhos ortográficos formados pela transposição de duas letras são mais difíceis de reconhecer do que palavras que não têm (Andrews, 1996). Dessa forma, torna-se necessário considerar se é ou não profícuo adotar as

definições mais tradicionais de vizinhança ortográfica indexadas por N (Coltheart & cols., 1977) e NF (Grainger & cols., 1989)?

A resposta a essa pergunta implica em considerações empíricas e teóricas. Do ponto de vista empírico, é necessário considerar até que ponto N, NF e TLN se correlacionam, pois caso exista uma correlação razoável entre essas variáveis é possível que os efeitos de N e NF sejam subprodutos de sua correlação com TLN. Pelo menos, no que tange ao português brasileiro, a correlação entre TLN e N ou TLN e NF é ínfima (ver a análise desenvolvida no capítulo ‘Análise de Corpus’ dessa tese para mais informações). Desse modo, é muito difícil explicar os efeitos de N ou NF com base em TLN e isso abre espaço para se considerar que N e NF possam ter efeitos genuínos e independentes de TLN, o que torna essas variáveis interessantes por si mesmas. Além disso, como apenas um número ínfimo de palavras do português do Brasil tem vizinhos formados pela transposição de duas letras, pelo menos de um ponto de vista prático, N e NF parecem ser variáveis mais importantes. Afinal, como a maioria das palavras tem vizinhos ortográficos e vizinhos de ortográficos de maior frequência, essas variáveis afetariam o processamento de um número muito maior de palavras do que TLN (aproximadamente, apenas 4% das palavras de 4 a 6 letras têm um vizinho formado pela transposição de duas letras, veja-se o capítulo ‘Análise de Corpus’ para maiores detalhes).

Por fim, considerando-se agora a mesma questão de um ponto de vista teórico, todos os modelos de reconhecimento visual de palavras discutidos nesse capítulo (inclusive o modelo SERIOL) trazem predições específicas para os efeitos de N e NF, enquanto apenas o SERIOL traz predições para TLN. Nesse sentido, é importante considerar ainda que, mesmo que conceitualmente as medidas N e NF se justem melhor aos pressupostos de modelos como os de ‘Busca e Verificação’ e de ‘Ativação Interativa e Competição’, simulações computacionais (Sears & cols., 1999), no caso dos modelos PPD, e análises precisas da dinâmica do modelo SERIOL (Whitney, 2004; Whitney & Lavidor, 2005), revelam que esses

modelos, de fato, trazem predições para o efeito dessas variáveis, o que não torna os modelos incomensuráveis (Kuhn, 1970) e, portanto, comparáveis. Sendo assim, considerando que nenhuma medida é perfeita e que um dos objetivos desse trabalho é discutir a capacidade dos modelos de reconhecimento visual de palavras de explicar o acesso lexical no português brasileiro, parece mais correto adotar uma medida de vizinhança ortográfica que seja comum a todos os modelos em questão, o que é o caso de N e NF. Dessa forma, embora N e NF possam ser considerados uma “boa primeira aproximação” do conceito de similaridade ortográfica (Grainger, 2008, p.5), para os propósitos desse estudo essas medidas ainda parecem ser a melhor aproximação possível, pois encontram justificativas empíricas, práticas e teóricas.

3 REVISÃO DA LITERATURA

De acordo com Andrews (1997), diversas tarefas têm sido utilizadas para se investigar os efeitos de N e de NF no reconhecimento visual de palavras e todas essas tarefas podem ser suscetíveis a algum tipo de crítica. Essa suscetibilidade a críticas deriva-se do fato de que as tarefas empregadas nos estudos, geralmente, envolvem outros processos que não apenas o de identificação lexical. Se o processo de identificação lexical for definido, como proposto por Monsell e cols. (1989), como a seleção da forma no léxico mental que melhor corresponde ao estímulo alvo, pode-se imaginar, por exemplo, que tarefas como a de leitura em voz alta incluirão processos adicionais como a computação e recuperação da pronúncia das palavras. Desse modo, certa precaução é necessária na interpretação dos resultados dos estudos, uma vez que não existem tarefas que podem ser consideradas medidas completamente “puras” do processo de identificação lexical. Tendo essa questão em mente, a revisão da literatura que se segue foi organizada de acordo com o tipo de tarefa empregada na investigação, diferentemente da forma mais tradicional que é a cronológica. Além disso, optou-se por restringir à discussão, aos estudos que avaliaram o efeito de vizinhança ortográfica, principalmente em palavras reais, já que essas estão, necessariamente, envolvidas nas atividades de leitura cotidianas. Outros estudos que não foram incluídos nessa revisão da literatura foram os estudos que utilizaram o paradigma de *priming* (p.ex., Moor & Brysbaert, 2000) e o paradigma de deslocar lateralmente a apresentação dos estímulos em relação ao ponto de fixação (p.ex., Ellis, 2004). Estudos com esses paradigmas não foram incluídos nessa revisão porque podem alterar a dinâmica do processo de acesso lexical “natural” (Andrew, 1997) e seriam muito difíceis de comparar com a maioria das tarefas tradicionalmente utilizadas para se investigar os efeitos de N e NF.

3.1 Tarefa de Decisão lexical

As tarefas de decisão lexical têm sido consideradas as tarefas mais adequadas para se investigar o acesso ao léxico (Rubenstein, Garfield & Millikan, 1970) e conseqüentemente, estão entre as tarefas mais utilizadas para se investigar o efeito de vizinhança ortográfica (Andrews, 1997). Nesse tipo de tarefa, os participantes têm que decidir se um conjunto de letras é uma palavra ou não. Para isso, a tarefa consiste na apresentação de um conjunto de letras (estímulo), geralmente no centro da tela de um computador. Uma vez que o estímulo tenha aparecido, o participante tem que apertar a tecla “sim”, caso este conjunto de letras seja uma palavra e a tecla “não”, caso este conjunto de letras seja uma pseudopalavra (seqüência de letras, construída com estrutura ortográfica possível na língua em questão, mas não associada a nenhum significado).

Grainger e cols. (1989) utilizaram a tarefa de decisão lexical para investigar o efeito de NF no reconhecimento visual de palavras de universitários falantes da língua francesa. Os pesquisadores utilizaram palavras de baixa freqüência de ocorrência e o número de vizinhos ortográficos de maior freqüência (NF) variava de acordo com três categorias: nenhum vizinho ortográfico de maior freqüência; um vizinho de maior freqüência; e, mais de um vizinho de maior freqüência. Além de utilizarem apenas palavras de baixa freqüência de ocorrência, Grainger e cols. também controlaram a freqüência subjetiva e a freqüência dos bigramas das palavras alvo. Os resultados indicaram um efeito inibidor de NF, porém esse efeito não se mostrou cumulativo: as palavras com muitos vizinhos de maior freqüência não apresentaram mais inibição do que as palavras com apenas um vizinho de maior freqüência, embora ambas as classes de palavras tenham sido reconhecidas mais lentamente do que as palavras sem vizinhos de maior freqüência. Em outro estudo utilizando a tarefa de decisão lexical, Grainger

(1990) encontrou o mesmo padrão de resultados em adultos falantes do holandês, mesmo controlando a frequência de ocorrência, a frequência dos bigramas, o fonema inicial, o número de vizinhos e a regularidade das correspondências grafema-fonema das palavras alvo do experimento.

Em um estudo posterior, também com universitários falantes do francês, Grainger e Segui (1990) replicaram o efeito inibidor de NF utilizando uma manipulação fatorial dessa variável (palavras sem vizinhos ortográficos de maior frequência e palavras com pelo menos um vizinho de maior frequência) e da frequência de ocorrência das palavras (palavras de baixa e de média frequência de ocorrência). Embora não tenham encontrado uma interação estatisticamente significativa entre NF e a frequência de ocorrência das palavras, o efeito inibidor de NF foi mais acentuado para as palavras de baixa frequência de ocorrência.

Em outro estudo, Grainger, O'Reagan, Jacobs e Segui (1992) investigaram se o efeito de NF era modulado pela posição na palavra alvo em que se geravam vizinhos de maior frequência de ocorrência. Por exemplo, a palavra 'anual' tem um vizinho ortográfico de maior frequência, 'atual', que se forma a partir da sua segunda letra; já a palavra 'atroz' tem um vizinho ortográfico de maior frequência, 'atriz', que se forma a partir da sua quarta letra. A pesquisa foi realizada com estudantes universitários falantes do francês e os pesquisadores encontraram que o efeito inibidor de NF era mais acentuado quando a palavra alvo tinha vizinhos ortográficos de maior frequência gerados a partir da sua quarta letra. Isso indica que a posição em que a palavra tem vizinhos de maior frequência pode influenciar no papel da variável NF.

Pugh, Rexer, Peter e Katz (1994) desenvolveram uma versão da tarefa de decisão lexical na qual, uma letra do meio da palavra alvo aparecia cem milésimos de segundo depois das outras letras dessa palavra. Nesse estudo foram avaliadas duas condições: em uma delas, a letra que demorava aparecer não apresentava ambigüidade (não se formavam, a partir da

mudança dessa letra, nenhum vizinho de frequência maior do que a da palavra alvo) e na outra condição, essa letra apresentava ambigüidade (podia-se formar um vizinho ortográfico de maior frequência nessa posição). Participaram desse estudo estudantes universitários falantes da língua inglesa e de forma condizente com um efeito inibidor de NF, os autores observaram que a condição em que a letra ambígua demorava a aparecer gerava mais erros e respostas mais demoradas nos participantes, em comparação com a condição em que a letra não ambígua demorava a aparecer. Em outro experimento, esses pesquisadores dividiram as pseudopalavras em duas condições: pseudopalavras com poucos vizinhos e pseudopalavras com muitos vizinhos. Observou-se que o efeito de NF só foi inibidor quando as pseudopalavras tinham muitos vizinhos ortográficos, do contrário, esse era nulo. Sendo assim, pode ser que o efeito de NF varie de acordo com a constituição das pseudopalavras do experimento também.

Um estudo que traz indícios sobre a importância do número de vizinhos ortográficos das pseudopalavras na tarefa de decisão lexical é o de Johnson e Pugh (1994). Esse estudo foi realizado com estudantes universitários falantes da língua inglesa e os pesquisadores desenvolveram diversos experimentos para investigar variações no efeito de N em função das características das pseudopalavras utilizadas, tendo especial relevância, nesse momento, seus dois primeiros experimentos. As palavras alvo foram as mesmas para ambos os experimentos e correspondiam a uma manipulação de N (palavras sem vizinhos ortográficos e palavras com mais de 6 vizinhos ortográficos) com o controle da frequência de ocorrência (todas as palavras eram de baixa frequência de ocorrência) e do número de letras dessas palavras. As pseudopalavras utilizadas foram diferentes em cada experimento: no primeiro, as pseudopalavras eram pronunciáveis e tinham um número de vizinhos ortográficos semelhante ao das palavras alvo; no segundo, as pseudopalavras não podiam ser pronunciadas e praticamente não tinham vizinhos ortográficos. No primeiro experimento, os estímulos

(palavras e pseudopalavras) foram apresentados em blocos, de acordo com o número de vizinhos. A ordem de apresentação desses blocos foi contrabalanceada entre os participantes (primeiro foram apresentados todos os estímulos com poucos vizinhos, depois, todos os estímulos com muitos vizinhos e vice-versa). Nesse experimento, os pesquisadores encontraram um efeito inibidor de N no tempo de reação dos participantes. Já no segundo experimento que contou com pseudopalavras que, além de não serem pronunciáveis, praticamente não tinham vizinhos ortográficos, o efeito de N foi facilitador.

Grainger e Jacobs (1996) utilizaram um raciocínio semelhante ao de Johnson e Pugh (1994) e desenvolveram duas variações da tarefa de decisão lexical para investigar os efeitos de N e NF em função das pseudopalavras utilizadas no estudo. O estudo foi realizado com universitários falantes da língua francesa. Os pesquisadores controlaram a frequência de ocorrência das palavras e a frequência dos bigramas, enquanto manipularam N e NF. No primeiro estudo, pseudopalavras com muitos vizinhos ortográficos foram utilizadas e encontrou-se um efeito nulo de N e um efeito inibidor de NF. No segundo experimento, pseudopalavras com poucos vizinhos ortográficos foram utilizadas e encontrou-se um efeito facilitador de N, no entanto, o efeito inibidor de NF permaneceu. É importante assinalar que, utilizando um raciocínio semelhante, Carreiras e cols. (1997), replicaram esse padrão de resultados investigando também estudantes universitários falantes da língua espanhola.

Pesquisando estudantes universitários falantes da língua inglesa, Huntsman e Lima (1996) manipularam NF (palavras com poucos vizinhos ortográficos de maior frequência de ocorrência e palavras com muitos vizinhos de maior frequência) e parearam as palavras de acordo com o número de vizinhos (N), a frequência de ocorrência, o número de letras e o número de sílabas. Os pesquisadores observaram um efeito inibidor de NF. É interessante observar que Huntsman e Lima desenvolveram análises *post hoc* e observaram que muito do efeito inibidor de NF se devia a comparação entre palavras sem vizinhos de maior frequência

de ocorrência e palavras com vizinhos de maior frequência, de forma semelhante aos resultados de Grainger e cols. (1989) e de Grainger (1990).

Perea e Pollatsek (1998) investigaram o efeito de NF utilizando uma tarefa de decisão lexical com estudantes universitários falantes da língua inglesa. Nesse estudo, NF foi manipulado com o requisito adicional de que nas palavras com vizinhos de maior frequência de ocorrência esses vizinhos ortográficos diferissem das palavras alvo em uma de suas letras mediais. Foram controlados o número de letras e a frequência de ocorrência das palavras alvo. Além disso, as pseudopalavras utilizadas eram pronunciáveis. Nesse estudo, as palavras com vizinhos ortográficos de maior frequência foram reconhecidas mais lentamente do que as palavras sem vizinhos ortográficos de maior frequência, indicando um efeito inibidor de NF. Um estudo desenvolvido por Sears, Campbell e Lupker (2006) utilizou os mesmos estímulos do trabalho de Perea e Pollatsek (1998) e replicou o efeito inibidor de NF em estudantes universitários falantes da língua inglesa, porém, quando novos estímulos foram utilizados em outra tarefa de decisão lexical, efeitos nulos de NF foram observados.

Até agora, pode-se dizer que o efeito de NF é razoavelmente consistente na tarefa de decisão lexical, porém alguns estudos realizados com falantes da língua inglesa têm apresentado resultados conflitantes com os dos estudos anteriores. Por exemplo, Andrews (1989) manipulou o número de vizinhos ortográficos e a frequência de ocorrência das palavras alvo em uma tarefa de decisão lexical. No primeiro experimento desenvolvido, pseudopalavras pronunciáveis foram empregadas, enquanto no segundo experimento, metade das pseudopalavras empregadas não era pronunciável. Estudantes universitários participaram dos experimentos e Andrews observou um efeito facilitador de N para as palavras de baixa frequência de ocorrência em ambos os experimentos, porém esse efeito foi mais acentuado no segundo experimento. Em um estudo posterior, também realizado com estudantes universitários, Andrews (1992) também efetuou uma manipulação fatorial de N e da

freqüência de ocorrência das palavras (palavras de baixa e de alta freqüência de ocorrência), porém, nesse estudo, a pesquisadora tomou o cuidado adicional de controlar a freqüência dos bigramas das palavras alvo. Mesmo assim, como no estudo anterior, um efeito facilitador de N que se restringiu às palavras de baixa freqüência de ocorrência foi observado. Mesmo não tendo manipulado NF, esses resultados podem ser considerados inconsistentes com os estudos anteriores que observaram um efeito inibidor de NF, pois N e NF se correlacionam positivamente na língua inglesa. Isto é, um grande número de vizinhos geralmente implica em, pelo menos, um vizinho ortográfico de freqüência maior que a da palavra alvo (Andrews, 1997).

Tentando esclarecer essa possível contradição nos estudos, Sears, Hino e Lupker (1995) desenvolveram uma série de experimentos baseando-se no paradigma da tarefa de decisão lexical. Todos os experimentos foram realizados com estudantes universitários falantes da língua inglesa. No primeiro experimento, as palavras alvo correspondiam à manipulação fatorial de N, NF e da freqüência de ocorrência. Nesse experimento, Sears e cols. (1995) observaram um efeito facilitador de N que não interagiu com NF e nem com a freqüência de ocorrências das palavras, enquanto o efeito de NF foi nulo. Em outro experimento Sears e cols. combinaram 65% das palavras do estudo de Andrews (1989) com outras novas palavras, de acordo com uma manipulação fatorial de N e da freqüência de ocorrência das palavras. Nesse estudo, tal como no de Andrews (1989), observou-se um efeito facilitador de N que se restringiu às palavras de baixa freqüência e um efeito nulo de NF. Na tentativa de se utilizar uma manipulação mais meticulosa de NF, Sears e cols. desenvolveram outro experimento utilizando as mesmas categorias utilizadas por Grainger e cols. (1989) na manipulação de NF: nenhum vizinho ortográfico de maior freqüência; um vizinho de maior freqüência; e, mais de um vizinho de maior freqüência. Além disso, para cada nível de NF, Sears e cols. manipularam também o número de vizinhos das palavras (tinham palavras com

muitos vizinhos ortográficos e palavras com poucos vizinhos ortográficos). Nesse experimento, todas as palavras eram de baixa frequência de ocorrência e observou-se um efeito facilitador de N e de NF, sendo que os dois fatores não interagiram. Por fim, os pesquisadores desenvolveram mais um experimento com a tarefa de decisão lexical. Nesse as palavras alvo também eram de baixa frequência de ocorrência e empregou-se uma manipulação fatorial de N (palavras com muitos e poucos vizinhos ortográficos) e NF (palavras sem vizinhos de maior frequência e palavras com vizinhos de maior frequência). As pseudopalavras desse experimento, ao contrário das utilizadas nos outros experimentos, tinham muitos vizinhos ortográficos. Mesmo nessa situação, Sears e cols. encontraram um efeito facilitador de N, no entanto, o efeito de NF foi nulo e não houve interação entre os dois fatores.

Forster e Shen (1996) também utilizaram a tarefa de decisão lexical para investigar os efeitos de N em estudantes universitários falantes da língua inglesa. Em seu primeiro experimento, Forster e Shen utilizaram uma manipulação bastante sofisticada de N. Sua manipulação experimental consistia em seis condições: palavras sem vizinhos ortográficos; palavras com um vizinho; palavras com dois vizinhos; palavras com três vizinhos; palavras com quatro vizinhos; e, palavras com cinco vizinhos. Nesse experimento, os pesquisadores observaram que o tempo de reação dos participantes tendia a decrescer à medida que o número de vizinhos ortográficos aumentava. Esse resultado também é inconsistente com o de estudos anteriores, pois nesse experimento, as palavras que tinham vizinhos ortográficos, tinham também, pelo menos, um vizinho de maior frequência. Sendo assim, Forster e Shen desenvolveram dois outros experimentos, dessa vez investigando os efeitos de N e de NF e nesses, encontraram, apenas, um efeito facilitador de N. O efeito de NF foi nulo e se manteve constante mesmo quando as pseudopalavras do experimento não tinham vizinhos ortográficos.

Em experimentos manipulando o número de vizinhos ortográficos e controlando NF, Huntsman e Lima (2002) observaram um efeito facilitador de N em estudantes universitários falantes do inglês. Em seus experimentos, as palavras alvo foram pareadas de acordo com a frequência de ocorrência, o número de letras e o número de sílabas. No primeiro experimento, todas as palavras tinham poucos vizinhos de maior frequência de ocorrência e no segundo experimento, as palavras alvo tinham mais vizinhos de maior frequência do que as palavras do primeiro experimento. Em ambos os experimentos as palavras com o maior número de vizinhos ortográficos foram reconhecidas mais rapidamente, indicando um efeito facilitador de N que foi independente do número de vizinhos de maior frequência das palavras.

Uma forma de se interpretar essa inconsistência nos resultados obtidos com a tarefa de decisão lexical tem sido proposta por Grainger e Jacobs (1996), Carreiras e cols. (1997) e Perea e Rosa (2000). Para esses pesquisadores, essa inconsistência nos resultados seria apenas aparente e os efeitos facilitadores de N seriam provenientes do uso de uma estratégia de “adivinhação”. De uma forma geral, os estudos que demonstraram um efeito facilitador de N, também apresentaram tempos de reação menores e uma maior porcentagem de erros, em comparação com os estudos que reportaram efeitos inibidores de NF (Grainger e Jacobs, 1996). Para esses pesquisadores, isso indicaria que nesses experimentos os participantes foram incentivados a dar respostas rápidas (ênfatizando o tempo de reação) e com isso, podem ter sido levados a adotar uma estratégia de “adivinhação rápida” (*fast guess strategy*). Nessa estratégia, os participantes emitiriam uma resposta com base na familiaridade dos estímulos sem, necessariamente, completar todos os processos envolvidos no acesso lexical. Desse modo, o desempenho na tarefa de decisão lexical, seria suscetível a influências de estratégias extra-lexicais, como já haviam argumentado anteriormente Balota e Chumbley (1984) e conseqüentemente, nem sempre representa uma medida “pura” do processo de identificação lexical.

No Modelo de Leitura Múltipla de Grainger e Jacobs (1996) e no Modelo de Dupla Rota em Cascata de Coltheart e cols. (2001), essa estratégia de “adivinhação rápida” pode ser interpretada da seguinte maneira: como o que interessa é responder o mais rápido possível, o sistema de acesso lexical não completa o processo de seleção da representação lexical (identificação da palavra alvo no léxico mental), mas sim, estabelece um limiar de ativação no sistema, a partir do qual, uma resposta positiva será emitida e, se em um determinado período de tempo este limiar não é atingido, o sistema emite uma resposta negativa. Assim sendo, um grande número de vizinhos ortográficos aumentaria a ativação do sistema lexical como um todo, fazendo com que a resposta positiva seja, rapidamente, emitida. Como as pseudopalavras, geralmente, têm um nível de ativação menor, o limiar não seria atingido a tempo e uma resposta negativa seria emitida, fazendo com que essa estratégia funcione razoavelmente bem, principalmente, quando as pseudopalavras têm poucos vizinhos ortográficos. No entanto, quando a precisão é enfatizada e as pseudopalavras do experimento têm um grande número de vizinhos, essa estratégia tende a falhar e, conseqüentemente, não vem a ser usada pelos participantes.

Considerando a proposta de Grainger e Jacobs (1996), Justi e Pinheiro (2006) desenvolveram dois experimentos com a finalidade de avaliar se os efeitos de N e NF podem ser modulados na tarefa de decisão lexical de acordo com a possibilidade de se usar ou não a estratégia de “adivinhação rápida”. Justi e Pinheiro investigaram o desempenho de estudantes de psicologia falantes do português do Brasil em duas versões da tarefa de decisão lexical. Na primeira delas, utilizaram uma tarefa de decisão lexical do tipo *Go/No-Go* que só requer que o participante responda às palavras. Nessa tarefa, foram utilizadas palavras e pseudopalavras que tinham um número de vizinhos e uma frequência de ocorrência dos vizinhos equiparáveis. Com esse controle, os participantes não poderiam usar a discrepância entre os índices de ativação lexical das palavras e pseudopalavras para estabelecer um limiar de resposta que

permitisse o uso da estratégia de “adivinhação rápida”. Nesse experimento, Justi e Pinheiro observaram um efeito inibidor de NF quando as palavras tinham poucos vizinhos e um efeito inibidor de N quando as palavras não tinham vizinhos de maior frequência. O segundo experimento de Justi e Pinheiro (2006) visava incentivar o uso da estratégia de “adivinhação rápida”. Para isso, os pesquisadores utilizaram uma tarefa de decisão lexical com tempo fixo, limitando o tempo de apresentação dos estímulos a 150 milésimos de segundo e empregaram pseudopalavras com um baixo índice de ativação lexical (as pseudopalavras desse experimento não tinham quaisquer vizinhos). Nesse experimento, ao se analisar a porcentagem de erros, observou-se um efeito facilitador de N (isso é, os estímulos com muitos vizinhos geraram uma menor porcentagem de erros) e uma tendência facilitadora de NF que não chegou a ser estatisticamente significativa. Os pesquisadores interpretaram essa inversão no padrão dos resultados entre os experimentos (da inibição a facilitação) como um indício de que os efeitos de N e NF realmente podem ser modulados pelo uso da estratégia de adivinhação rápida.

Em um estudo posterior, Justi e Pinheiro (2008) tentaram avaliar, de forma independente, o uso da estratégia de “adivinhação rápida”. Nesse outro estudo, Justi e Pinheiro, lançaram a hipótese auxiliar de que crianças seriam menos propensas a utilizar estratégias cognitivas complexas como a de “adivinhação rápida”. De acordo com os pesquisadores, isso aconteceria porque as crianças teriam pouca familiaridade com a tarefa de decisão lexical e, além disso, seus processos de reconhecimento visual de palavras ainda não estariam tão automatizados como os dos adultos, fazendo com que as crianças tivessem poucos recursos extras para empregar no uso de estratégias. Sendo assim, os pesquisadores hipotetizaram que as crianças deveriam apresentar um padrão de respostas semelhante aos dos adultos na tarefa de decisão lexical do tipo *Go/NoGo* e um padrão diferente na tarefa de decisão lexical com tempo fixo, posto que essa última incentivaria o uso da estratégia de

“adivinhação rápida”. Tendo essas considerações em mente, Justi e Pinheiro (2008) replicaram seu estudo anterior (Justi & Pinheiro, 2006) com uma amostra de crianças da quarta série do ensino fundamental e obtiveram os padrões esperados. Na tarefa de decisão lexical do tipo *Go/NoGo* efeitos inibidores de N e NF foram observados e na tarefa de decisão lexical com tempo fixo, diferentemente do estudo com adultos, efeitos facilitadores não foram observados, sendo observado um efeito inibidor de NF. Dessa forma, Justi e Pinheiro (2008) postularam que no português do Brasil o efeito de N e NF no processo de identificação lexical é genuinamente inibidor e que a tarefa de decisão lexical do tipo *Go/NoGo* avalia mais fidedignamente o acesso lexical genuíno, posto que nessa, o padrão de respostas de adultos (Justi & Pinheiro, 2006) e crianças (Justi & Pinheiro, 2008) foi semelhante.

No entanto, é importante ressaltar que Siakaluk, Sears e Lupker (2002) investigaram o efeito de N e NF em universitários falantes da língua inglesa em função do número de vizinhos ortográficos das pseudopalavras conforme sugerido por Grainger e Jacobs (1996) e não obtiveram efeitos inibidores de N ou NF, mesmo em situações em que as pseudopalavras tinham muitos vizinhos. Dessa forma, torna-se difícil se atribuir a inconsistência nos achados dos estudos, unicamente, a diferenças no uso ou não de estratégias por parte dos participantes. Pode ser, nesse caso, que diferenças entre as línguas, também estejam influenciando o resultado dos diferentes estudos (Andrews, 1997). Porém, seria interessante postergar essa discussão até que sejam considerados os achados sobre o efeito de vizinhança ortográfica obtidos em outras tarefas que não a de decisão lexical.

3.2 Tarefa de Identificação Perceptual

Outra tarefa utilizada para se investigar o efeito de vizinhança ortográfica é a *tarefa de identificação perceptual*. O estudo de Grainger e Segui (1990) foi o primeiro a investigar o efeito de NF no reconhecimento visual de palavras por meio da tarefa de identificação perceptual. Esse estudo foi realizado com estudantes universitários falantes do Francês e teve como base o paradigma do desmascaramento progressivo (*progressive demasking task*): nesse paradigma, a palavra alvo aparecia por completo (p.ex.: ‘bola’) e rapidamente era substituída, ciclicamente, por uma máscara (p.ex.: ‘#####’); a cada ciclo, palavra/máscara, a exposição da palavra alvo era aumentada em 16 milissegundos e a da máscara diminuída na mesma quantidade. Os participantes tinham que pressionar uma tecla assim que identificassem a palavra alvo, após isso a palavra alvo desaparecia e os participantes deviam digitar qual era essa palavra. As palavras utilizadas como estímulo pelos experimentadores atenderam a uma manipulação fatorial de NF (nenhum vizinho de maior frequência e pelo menos um vizinho de maior frequência) e da frequência de ocorrência das palavras (palavras de baixa frequência de ocorrência e palavras de média frequência de ocorrência). Foram controlados a frequência de ocorrência dos bigramas e o número de vizinhos ortográficos (N) das palavras. Encontrou-se um efeito inibidor de NF e um efeito facilitador da frequência de ocorrência das palavras no tempo de reação, isso é, as palavras que tinham pelo menos um vizinho ortográfico de maior frequência foram reconhecidas mais lentamente do que as que não tinham, e, as palavras de média frequência de ocorrência foram reconhecidas mais rapidamente do que as de baixa frequência. Na análise da porcentagem de erros apenas o efeito inibidor de NF foi estatisticamente significativo.

Em outro estudo, Grainger e Jacobs (1996) tentaram replicar achados de Grainger e Segui (1990) na tarefa de identificação perceptual, utilizando dois paradigmas: o do desmascaramento progressivo; e, a versão padrão da tarefa de identificação perceptual (*standard perceptual identification task*). O estudo utilizando o paradigma do desmascaramento progressivo contou com a participação de trinta alunos de um curso de Psicologia, todos falantes da língua francesa. Os autores utilizaram os mesmos procedimentos de Grainger e Segui (1990), porém dessa vez, utilizaram apenas palavras de baixa frequência de ocorrência e observaram um efeito inibidor de NF e um efeito nulo de N. O estudo com a tarefa de identificação perceptual padrão (*standard perceptual identification task*) também foi realizado com trinta alunos de um curso de Psicologia, todos falantes da língua francesa e que não tinham participado do experimento anterior. Esse novo estudo teve como base o seguinte paradigma: aparecia no centro da tela uma máscara (#####) por 500ms, sendo imediatamente seguida pela palavra alvo que permanecia na tela por 40ms que, por fim, era substituída novamente por uma máscara (#####) que permanecia na tela até a emissão da resposta. Nesse estudo, não havia um tempo limite para a emissão da resposta e os participantes deviam digitar a palavra que tinham visto. As palavras utilizadas como estímulo pelos experimentadores atenderam a uma manipulação fatorial de NF (nenhum vizinho de maior frequência e pelo menos um vizinho de maior frequência) e da frequência de ocorrência das palavras (palavras de baixa frequência de ocorrência e palavras de alta frequência de ocorrência), sendo que o número de vizinhos ortográficos (N) foi mantido constante (média de 2,5 vizinhos ortográficos). As palavras de maior frequência de ocorrência foram reconhecidas de forma mais precisa de uma forma geral, enquanto o efeito inibidor de NF foi significativo apenas para as palavras de baixa frequência de ocorrência. Levando-se em consideração que no estudo anterior de Grainger e Segui (1990) não foram utilizadas palavras de alta frequência de ocorrência, os resultados dos estudos tornam-se bastante consistentes.

Em um estudo com universitários falantes do espanhol, Carreiras e cols. (1997) utilizaram a mesma metodologia empregada por Grainger e Segui (1990) em sua tarefa de identificação perceptual. As palavras-estímulo empregadas por Carreiras e cols. (1997) eram palavras de baixa frequência de ocorrência e atendiam a uma manipulação fatorial de N (palavras com poucos vizinhos ortográficos e palavras com muitos vizinhos ortográficos) e NF (palavras sem vizinhos ortográficos de maior frequência e palavras com pelo menos um vizinho ortográfico de maior frequência), enquanto a frequência das sílabas, o som inicial e o número de letras das palavras foram controlados. Os pesquisadores encontraram um efeito inibidor robusto de NF e um efeito inibidor fraco de N que não chegou a ser significativo estatisticamente.

Em um estudo utilizando outra variação da tarefa de identificação perceptual, Snodgrass e Mintzer (1993) desenvolveram cinco experimentos para investigar o efeito de N no reconhecimento visual de palavras em estudantes universitários falantes da língua inglesa. Têm especial relevância os resultados do segundo ao quinto experimento, pois esses contaram com as palavras-estímulo do estudo de Andrews (1989) que correspondem a uma manipulação fatorial de N e da frequência de ocorrência das palavras. Na tarefa de identificação perceptual empregada por Snodgrass e Mintzer, fragmentos da palavra alvo foram apresentados de maneira sucessiva e incremental, no centro da tela de um computador. Os participantes controlavam o aparecimento dos fragmentos da palavra, que iam sempre da versão mais fragmentada da palavra alvo (mais difícil de reconhecer) a menos fragmentada (eram oito níveis de fragmentação, sendo que o último nível correspondia à palavra não fragmentada). Os participantes eram orientados a digitar qual era a palavra alvo o quanto antes. Em seus dois primeiros experimentos, Snodgrass e Mintzer permitiram aos participantes que adotassem tentativas sucessivas de acertar a palavra alvo e obtiveram resultados inconsistentes: um efeito nulo de N, no primeiro experimento e, um efeito

facilitador, no segundo. No terceiro e quarto experimentos os autores permitiram aos participantes apenas uma tentativa de acertar o estímulo, nesse caso, observaram um efeito inibidor de N que se manteve consistente em ambos os experimentos. Por fim, Snodgrass e Mintzer desenvolveram um último experimento. Nesse experimento, os participantes não controlavam o aparecimento dos diferentes níveis de fragmentação. Os oito níveis de fragmentação apareciam em sucessão de 167 milissegundos cada e os participantes deviam pressionar uma tecla assim que tivessem identificado a palavra alvo (apenas uma tentativa de identificar a palavra era permitida). Consistente com os resultados do terceiro e quarto experimentos, os pesquisadores encontraram um efeito inibidor de N em seu quinto experimento.

Para Andrews (1997) um dos problemas dessa modalidade das tarefas de identificação perceptual é que os participantes podem utilizar diversas estratégias para tentar adivinhar a palavra correta à medida que as letras vão aparecendo e, talvez, isso não implique em acesso lexical genuíno. Para Andrews, a variação nos resultados obtidos, de acordo com o número de tentativas permitidas, seria um indício do uso dessas estratégias. Essa crítica é particularmente pertinente para os experimentos nos quais os participantes podiam controlar o aparecimento dos níveis de fragmentação das palavras, pois poderiam fazer hipóteses a respeito do estímulo antes de passar para o próximo nível de fragmentação. No entanto, o quinto experimento de Snodgrass e Mintzer (1993) é menos suscetível a essa crítica, porque os participantes não controlavam o aparecimento dos diferentes níveis de fragmentação. Nesse caso, a dinâmica temporal de aparecimento do estímulo alvo no quinto experimento do estudo de Snodgrass e Mintzer, pode ser considerada semelhante à de outros estudos. Além disso, nos estudos de Grainger e Segui (1990) e de Carreiras e cols. (1997), a palavra alvo aparecia por completo e era coberta, ciclicamente, por uma máscara, assim o processo perceptual não poderia ser afetado pela natureza incremental da tarefa como proposto por Andrews (1997).

Levando em conta essas críticas de Andrews (1997) e visando clarificar o papel de N e NF na tarefa de identificação perceptual Sears, Lupker e Hino (1999) desenvolveram uma série de quatro estudos com a versão padrão dessa tarefa (*standard perceptual identification task* – como a de Grainger & Jacobs, 1996). Sears e cols. (1999) dividiram a tarefa de identificação perceptual em duas versões: uma em que a visibilidade dos estímulos era maior e outra em que a visibilidade dos estímulos era menor. Em ambas as versões a palavra alvo era apresentada por apenas 24ms, sendo a principal diferença entre as versões o tempo de duração das máscaras (#####): na versão de maior visibilidade, as máscaras (prévia ao estímulo e posterior a esse) eram apresentadas por 500ms; enquanto na versão de menor visibilidade, as máscaras eram apresentadas por apenas 42ms.

Nos seus dois primeiros experimentos, Sears e cols. (1999) manipularam a frequência de ocorrência das palavras (palavras de baixa e alta frequência de ocorrência), o número de vizinhos ortográficos (palavras com poucos vizinhos ortográficos e palavras com muitos vizinhos ortográficos) e NF (palavras sem vizinhos ortográficos de maior frequência de ocorrência e palavras com pelo menos um vizinho ortográfico de maior frequência de ocorrência). Participaram dos experimentos 82 estudantes universitários falantes da língua inglesa, sendo que 36 participaram da versão na qual os estímulos tinham maior visibilidade e 46 da versão em que os estímulos tinham menor visibilidade. As palavras de maior frequência de ocorrência foram reconhecidas mais corretamente em ambas as versões dos experimentos. Já o efeito de vizinhança ortográfica apresentou um padrão diferente de acordo com a versão. Na versão onde a visibilidade dos estímulos era maior, N apresentou um efeito inibidor geral e NF apresentou um efeito inibidor que se restringiu às palavras de baixa frequência de ocorrência que tinham poucos vizinhos ortográficos. Já na condição em que a visibilidade dos estímulos era menor, tanto N quanto NF apresentaram um efeito facilitador. Na tentativa de compreender melhor o efeito de NF, Sears e cols. (1999) desenvolveram mais dois

experimentos com falantes da língua inglesa, porém, dessa vez, manipularam apenas NF, sendo que todas as palavras alvo tinham muitos vizinhos ortográficos e variavam em frequência de ocorrência (alta e baixa). Novamente duas condições de visibilidade foram empregadas. Na condição de maior visibilidade o efeito de NF foi nulo e na condição de menor visibilidade o efeito de NF foi facilitador. Na interpretação de seus dados, Sears e cols. (1999) argumentaram que na condição de maior visibilidade, os participantes teriam maior probabilidade de tentar “adivinhar” a palavra alvo e isso levaria a um efeito inibidor, porque quanto maior o número de vizinhos ortográficos de uma palavra, maior a probabilidade dos participantes “chutarem” um desses vizinhos ao invés da palavra alvo. Já na condição de menor visibilidade, os participantes “chutariam” menos, com isso, o processo de acesso lexical genuíno estaria mais presente e o efeito observado de N e NF deveria ser facilitador. O problema dessa interpretação de Sears e cols. (1999) é que ela tem como base a premissa um tanto contra-intuitiva de que os participantes vão tender a “chutar” mais vezes nas condições fáceis (condições de maior visibilidade) do que nas condições mais difíceis (condições de menor visibilidade). Ora, parece fazer mais sentido que as pessoas tentem “chutar” uma resposta nas situações em que elas têm menos clareza do que viram do que nas situações nas quais elas têm mais clareza do que podem ter visto.

Independente de qual interpretação possa fazer mais sentido, é importante se considerar que ambas as possibilidades são possíveis e que talvez o mais interessante seja tentar reduzir ao máximo a possibilidade de “chute” por parte dos participantes de forma a tentar garantir que o acesso lexical na tarefa em questão esteja ocorrendo de forma genuína. Na tarefa de Sears e cols. (1999), os participantes não tinham um limite de tempo para responder (digitar a palavra alvo) e isso certamente pode encorajar o uso de estratégias de adivinhação, independentemente, da condição de visibilidade dos estímulos. Nesse sentido, seria interessante considerar tarefas de identificação perceptual em que existe um limite de

tempo para a resposta ou em que o tempo de reação também é considerado, como no caso do quinto experimento de Snodgrass e Mintzer (1993) e dos estudos de Grainger e Segui (1990) e de Carreiras e cols. (1997).

3.3 Tarefa de Leitura em Voz Alta

A tarefa de leitura em voz alta também é uma tarefa que tem sido muito utilizada para a investigação dos efeitos de N e de NF. Nessa tarefa, apresenta-se aos participantes uma palavra alvo e esses têm que pronunciá-la o mais rapidamente possível. Diversos pesquisadores têm observado um efeito facilitador da vizinhança ortográfica nas tarefas de leitura em voz alta, mesmo pesquisando línguas diferentes (Andrews, 1997).

O estudo de Andrews (1989) investigou o efeito de N em tarefas de leitura em voz alta realizadas com estudantes universitários falantes da língua inglesa. Duas tarefas de leitura em voz alta foram empregadas, uma na qual as palavras deveriam ser pronunciadas o mais rápido possível e outra na qual a pronúncia das palavras só poderia ser emitida após o aparecimento de um sinal na tela do computador (*delayed naming task*). As palavras-estímulo do experimento correspondiam a uma manipulação fatorial de N (palavras com poucos vizinhos ortográficos e palavras com muitos) e da frequência de ocorrência das palavras (palavras de baixa frequência e palavras de alta frequência). No experimento em que a pronúncia das palavras era imediata, encontrou-se um efeito facilitador de N que foi mais acentuado para as palavras de baixa frequência de ocorrência (ou seja, as palavras com muitos vizinhos ortográficos foram pronunciadas mais rapidamente do que as com poucos vizinhos). Já no experimento em que a pronúncia deveria ser postergada, não se encontrou nenhum efeito de

N. Andrews argumentou que esse padrão de resultados indica que o efeito de N não se deve a processos meramente articulatórios (isso é, a seqüência de fonemas de palavras com poucos vizinhos seria mais difícil de articular), pois se esse fosse o caso, efeitos de N deveriam ser encontrados também na tarefa de leitura em voz alta na qual a pronúncia era postergada (afinal, a articulação da pronúncia das palavras também está envolvida nessa tarefa). Em um estudo posterior, Andrews (1992) replicou esses achados utilizando novamente as duas variações da tarefa de leitura em voz alta, porém, dessa vez, suas palavras-estímulo foram selecionadas de forma a controlar a freqüência de ocorrência dos bigramas.

Ainda no que diz respeito à língua inglesa, Laxon, Gallaguer e Masterson (2002) investigaram o efeito de N na leitura de crianças de 6 a 7 anos de idade. Nesse estudo, as palavras que tinham muitos vizinhos foram lidas com maior precisão do que as palavras com poucos vizinhos, no entanto esse efeito facilitador de N restringiu-se as palavras de três letras. Além disso, de uma forma geral, as palavras com muitos vizinhos apresentaram uma maior porcentagem de erros que foram classificados como lexicalizações (pronúncias incorretas que se assemelham a uma palavra comum). A existência desses erros foi interpretada como um indício de que as crianças faziam uso do conhecimento que tinham de outras palavras similares enquanto liam a palavra alvo.

Em um estudo realizado com universitários falantes da língua francesa, Peereman e Content (1995) utilizaram, como Andrews (1989; 1992), duas versões da tarefa de leitura em voz alta: uma com pronúncia imediata e outra com a pronúncia postergada. O estudo contou com uma manipulação fatorial de N e da freqüência de ocorrência das palavras e observou-se um efeito facilitador de N que se restringiu às palavras de baixa freqüência de ocorrência na tarefa de leitura em voz alta em que a pronúncia era imediata e um efeito nulo na tarefa em que a pronúncia deveria ser postergada.

Em um estudo com estudantes universitários falantes do espanhol, Carreiras e cols. (1997) investigaram o efeito de N e de NF em uma tarefa de leitura em voz alta. As palavras utilizadas eram todas de baixa frequência de ocorrência e foram pareadas de acordo com a frequência silábica, o som inicial e o número de letras. Nesse estudo, os pesquisadores também encontraram um efeito facilitador de N, no entanto, esse efeito restringiu-se às palavras que tinham vizinhos de maior frequência de ocorrência. Além disso, observou-se uma tendência a um efeito facilitador de NF que se restringiu às palavras com muitos vizinhos ortográficos.

Por fim, no que tange ao efeito de NF na tarefa de leitura em voz alta, Grainger (1990) investigou o efeito dessa variável em adultos falantes do holandês. Seu estudo empregou uma manipulação fatorial de NF (palavras sem vizinhos ortográficos de maior frequência; palavras com um vizinho ortográfico de maior frequência; e, palavras com mais de um vizinho de maior frequência) e da frequência de ocorrência das palavras (palavras de baixa frequência de ocorrência e palavras de média frequência de ocorrência). Grainger observou que o efeito de NF interagiu com a frequência de ocorrência das palavras. Para as palavras de baixa frequência de ocorrência, um aumento no número de vizinhos de maior frequência fazia com que elas fossem lidas mais rapidamente (efeito facilitador de NF), enquanto para as palavras de média frequência, NF não teve efeito algum.

O problema da tarefa de leitura em voz alta é que o desempenho nessa tarefa nem sempre depende da identificação lexical do estímulo alvo. O simples fato dos participantes poderem pronunciar corretamente uma pseudopalavra serve como um indício de que eles podem pronunciar os estímulos do experimento (palavras ou pseudopalavras) sem acessar, necessariamente, sua representação lexical, afinal, uma pseudopalavra não tem representação lexical (Coltheart, 1978). Um indício a favor dessa hipótese é a observação de um efeito facilitador de N na pronúncia de pseudopalavras (Laxon & cols., 1992; 2002 – pesquisando

falantes do inglês). É interessante observar que, em um estudo realizado com falantes do italiano que investigou os efeitos de N e NF na leitura em voz alta de pseudopalavras, encontrou-se um efeito facilitador de N, mas não de NF (Arduino & Burani, 2004). Estudos como os de Andrews (1989; 1992) e de Peereman e Content (1995), trazem indícios de que o *locus* do efeito de vizinhança ortográfica não pode ser atribuído a processos meramente articulatórios, no entanto, isso não é o mesmo que dizer que esse efeito se deve, exclusivamente, ao processo de identificação lexical.

3.4 Estudos Sobre o Movimento dos Olhos na Leitura

Nesse momento, pode-se perguntar até que ponto os efeitos de vizinhança ortográfica estariam presentes também no processo natural de leitura silenciosa. Isto é, os efeitos de vizinhança ortográfica também seriam observados em experimentos com maior validade ecológica ou esses seriam subprodutos específicos do *design* experimental das tarefas tradicionais? Um dos métodos mais utilizados para se estudar a leitura silenciosa, que apresenta maior validade ecológica, é o estudo do movimento dos olhos na leitura (Eysenck & Keane, 2007).

Um dos primeiros estudos sobre o efeito de vizinhança ortográfica a utilizar informações a respeito do movimento dos olhos foi o de Grainger e cols. (1989). Nesse estudo, os leitores, estudantes universitários falantes da língua francesa, deveriam julgar se duas palavras de baixa frequência de ocorrência eram semanticamente relacionadas. A palavra alvo era vista primeiro e enquanto isso, a segunda palavra ficava encoberta por uma máscara (#####), assim que os participantes moviam os olhos para a segunda palavra, ela se tornava

disponível e a primeira palavra, passava a ser encoberta por uma máscara (#####). Os resultados do estudo de Grainger e cols. (1989) indicaram que os participantes passavam mais tempo observando as palavras alvo que tinham pelo menos um vizinho de maior frequência do que as palavras alvo que não tinham vizinhos de maior frequência, mesmo sendo controladas a frequência de ocorrência das palavras e a frequência dos bigramas.

O estudo do movimento dos olhos é bastante complexo e, talvez, os estudos utilizando esse paradigma possam ser mais bem compreendidos tendo como pano de fundo um modelo teórico a respeito dos processos envolvidos. Um dos modelos teóricos mais influentes sobre o movimento dos olhos durante a leitura é o *Modelo E-Z Reader* de Reichle, Pollatsek, Fisher e Rayner (1998) e Reichle, Rayner e Pollatsek (2003). Esse modelo propõe a existência de dois estágios no acesso lexical: um primeiro estágio, conhecido como cheque de familiaridade (*familiarity check*) e um segundo, no qual o acesso lexical é completado (*lexical completion stage*). Acredita-se que as informações sobre as palavras de um texto são extraídas durante as fixações (momentos em que nossos olhos se fixam em uma determinada palavra) e essas duram em média de 200 a 250 milissegundos (Eysenck & Keane, 2007; Rayner, Juhasz & Pollatsek, 2005). Nesse sentido, o primeiro estágio (*familiarity check*) é indexado principalmente pela duração da primeira fixação dos olhos na palavra e pela probabilidade dessa palavra ser pulada (*skipped*) durante a leitura; enquanto o segundo estágio (*lexical completion*) é indexado principalmente por três medidas: o tempo que o indivíduo passa fixando a próxima palavra (*spillover effect*); o número de regressões que ele faz à palavra alvo para fixá-la novamente; e, pelo tempo total que ele passa fixando a palavra alvo.

Tendo esses dois estágios em mente, Perea e Pollatsek (1998) investigaram o efeito de NF no movimento dos olhos durante a leitura em estudantes universitários falantes da língua inglesa. Para isso, eles construíram várias sentenças que eram neutras em seu significado (isto é, eram igualmente compatíveis com a palavra alvo e com o vizinho de maior frequência

dessa palavra) e incluíram nelas as palavras alvo do experimento de acordo com uma manipulação de NF. Além disso, as palavras alvo com e sem vizinhos ortográficos de maior frequência, foram pareadas de acordo com a sua frequência de ocorrência e com o seu número de letras. A partir disso, os pesquisadores mensuraram variáveis ligadas a ambos os estágios propostos pelo modelo. Observou-se que NF não teve qualquer efeito sobre as medidas relacionadas ao primeiro estágio (*familiarity check*) proposto pelo modelo. Já com relação às medidas relacionadas ao segundo estágio (*lexical completion stage*) proposto pelo modelo, o efeito de NF foi inibidor: os participantes passaram mais tempo fixando a próxima palavra; apresentaram um maior número de regressões à palavra alvo; e, um maior tempo total de fixação nessa, quando a palavra alvo tinha vizinhos de maior frequência de ocorrência.

Em outro estudo, também com universitários falantes da língua inglesa, Pollatsek, Perea e Binder (1999) investigaram o efeito de N em dois experimentos utilizando a mesma metodologia do estudo de Perea e Pollatsek (1998). Em seu primeiro experimento, os pesquisadores observaram um efeito inibidor de N no tempo gasto nas primeiras fixações na palavra alvo e no tempo total que os participantes passavam processando essa palavra. Como nesse primeiro experimento NF não foi controlado, os pesquisadores utilizaram uma análise de regressão para tentar separar os efeitos de N e de NF. Os pesquisadores encontraram um efeito inibidor de NF, mas observaram também, um efeito facilitador fraco de N que, ao contrário de NF, ocorreu no primeiro estágio (*familiarity check*). Tendo isso em mente, Pollatsek e cols. desenvolveram outro experimento no qual o número de vizinhos ortográficos de maior frequência das palavras foi mantido constante e N foi manipulado. Nesse experimento, Pollatsek e cols. observaram um efeito facilitador de N na probabilidade da palavra alvo ser pulada (*skipped*). No entanto, esse efeito facilitador foi contraposto por uma maior probabilidade de se regressir à palavra alvo e por um maior tempo total despendido no processamento da palavra alvo. Desse modo, o efeito de N parece ser facilitador no primeiro

estágio proposto pelo modelo (*familiarity check*) e inibidor no segundo (*lexical completion stage*).

Perea e Rosa (2000) têm interpretado esses achados como um indício de que a variável N parece afetar o primeiro estágio do acesso lexical, servindo, talvez, como uma espécie de índice de familiaridade, enquanto a variável NF, afetaria mais acentuadamente o estágio final, quando a identificação do estímulo deve ser inequívoca. No entanto, os resultados dos estudos investigando o movimento dos olhos durante a leitura também não são totalmente consistentes para se permitir uma conclusão tão direta como essa. Em um estudo recente, Sears e cols. (2006) desenvolveram dois experimentos que mensuraram o movimento dos olhos durante a leitura de sentenças. No primeiro, os mesmos estímulos empregados por Perea e Pollatsek (1998) foram utilizados e observou-se um efeito inibidor de NF no tempo total que os participantes passavam fixando a palavra alvo, porém esse efeito foi limitado às palavras de baixa frequência de ocorrência e aos participantes que faziam poucas regressões à palavra alvo. No segundo experimento, Sears e cols. utilizaram um conjunto novo de estímulos e não observaram qualquer efeito de NF.

3.5 Tarefa de Categorização Semântica

Forster e Shen (1996) decidiram investigar se os efeitos de vizinhança ortográfica também poderiam ser observados em uma tarefa de categorização semântica. Os pesquisadores desenvolveram uma tarefa na qual os participantes deveriam pressionar uma tecla caso uma palavra fosse o nome de um animal e pressionar outra tecla quando a palavra não fosse o nome de um animal. O experimento contou com 120 palavras que correspondiam

a nomes de animais e com 120 palavras que não representavam nomes de animais. As palavras alvo não representavam nomes de animais e, portanto, a única medida de interesse no estudo referia-se ao tempo despendido para se dar uma resposta negativa. As palavras alvo eram todas palavras de baixa frequência de ocorrência da língua inglesa e correspondiam a uma manipulação de N (palavras com um vizinho ortográfico; palavras com dois vizinhos; e, palavras com três ou quatro vizinhos ortográficos) e de NF (palavras sem vizinhos ortográficos de maior frequência e palavras com vizinhos de maior frequência). Participaram do estudo estudantes universitários falantes da língua inglesa. Os pesquisadores observaram uma interação estatisticamente significativa entre N e NF, que indicou que o efeito de NF foi inibidor quando as palavras tinham apenas um vizinho ortográfico e facilitador quando essas tinham muitos vizinhos ortográficos.

Carreiras e cols. (1997) utilizaram a mesma metodologia utilizada por Forster e Shen (1996) para investigar os efeitos de N e NF em estudantes universitários falantes da língua espanhola. A tarefa escolhida também foi a de categorização semântica (animal / não animal) e as respostas de interesse também eram as negativas. As palavras alvo correspondiam a uma manipulação fatorial de N (palavras com poucos vizinhos ortográficos e palavras com muitos) e de NF (palavras sem vizinhos ortográficos de maior frequência e palavras com vizinhos de maior frequência), sendo controlados a frequência de ocorrência (eram todas palavras de baixa frequência), o som inicial e o número de letras das palavras. Os pesquisadores observaram uma interação estatisticamente significativa entre N e NF, no entanto, o padrão dessa interação foi o inverso do encontrado por Forster e Shen (1996). Nesse estudo, NF apresentou um efeito inibidor para as palavras com muitos vizinhos ortográficos e um efeito facilitador para as palavras com poucos vizinhos ortográficos.

Outro estudo que também empregou o mesmo paradigma do estudo Forster e Shen (1996) na tarefa de categorização semântica foi o desenvolvido por Sears e cols. (1999).

Participaram desse estudo 35 estudantes universitários falantes da língua inglesa. As palavras alvo (respostas negativas) correspondiam a uma manipulação fatorial de N (palavras com poucos vizinhos ortográficos e palavras com muitos) e de NF (palavras sem vizinhos ortográficos de maior frequência e palavras com vizinhos de maior frequência), sendo todas as palavras de baixa frequência de ocorrência. No entanto, ao contrário do estudo de Forster e Shen (1996), nesse estudo Sears e cols. não encontram um efeito inibidor de NF (o efeito foi nulo) e, além disso, encontraram um efeito facilitador de N.

Usando uma metodologia diferente Bowers, Davis e Hanley (2005) também investigaram o efeito de vizinhança ortográfica na tarefa de categorização semântica. Bowers e cols. (2005) utilizaram em seu estudo palavras de baixa frequência de ocorrência sem quaisquer vizinhos ortográficos. Metade dessas palavras correspondia a entes naturais (p.ex.: pombo, banana) e a outra metade correspondia a artefatos (p.ex.: âncora, berço). Os pesquisadores decidiram investigar se o fato dos participantes aprenderem uma pseudopalavra que fosse ortograficamente vizinha das palavras alvo afetaria no reconhecimento das palavras alvo. Para isso, os pesquisadores geraram pseudopalavras que eram vizinhas das palavras alvo mudando uma de suas letras mediais. Isso ocorreu para ambos os tipos de palavras alvo (entes naturais e artefatos). Essas pseudopalavras foram divididas em duas listas que foram contrabalanceadas entre dois grupos de indivíduos. Metade dos participantes foi treinada em uma lista e a outra parte em outra lista. Dessa forma, cada palavra alvo, para metade dos participantes, tinha uma pseudopalavra ortograficamente vizinha e para a outra metade dos participantes, não tinha. Após o treinamento, que consistia em escrever e ler em voz alta as pseudopalavras, os participantes desempenharam a tarefa de categorização semântica que continha apenas as palavras alvo. A pesquisa foi desenvolvida com estudantes universitários falantes da língua inglesa e os pesquisadores observaram que as palavras alvo que tinham “ganhado” vizinhos ortográficos foram reconhecidas mais lentamente do que as que não

“ganharam” vizinhos ortográficos. Esse efeito inibidor aumentou linearmente em três dias de treinamento e testagem repetida. O estudo de Bowers e cols. tem a vantagem de ter podido utilizar uma mesma palavra como seu próprio controle (devido ao contrabalanceamento das listas de treinamento entre os participantes) o que diminui a influência de quaisquer outras diferenças que possam existir entre as palavras. No entanto, essa metodologia também tem suas limitações. Uma delas é que se pode interpretar que a pseudopalavra aprendida interferiu no processamento das palavras alvo porque funcionou como um “*priming* de longo prazo” e sendo assim, afetou a dinâmica natural de processamento lexical. De fato, em estudos que utilizaram o mesmo tipo de paradigma (*priming* de longo prazo), o estudo prévio (p.ex., na sessão de treinamento) de um vizinho ortográfico das palavras alvo, tendeu a prejudicar o reconhecimento posterior (p.ex., na sessão experimental) das palavras alvo (Ratcliff & McKoon, 1996; 1997; Zeelenberg, Wagenmakers & Raaijmakers, 2002). Outra dúvida que se pode levantar é quanto ao *status* da pseudopalavra “aprendida” pelos participantes: ela deveria ser considerada um vizinho ortográfico de maior frequência que a palavra alvo ou não? Em outras palavras, o efeito observado deveria ser atribuído a NF ou a N?

Uma possível explicação para tantos achados inconsistentes na tarefa de categorização semântica provêm do estudo de Pecher, Zeelenberg e Wagenmakers (2005). Nesse estudo, os pesquisadores hipotetizaram que uma variável que pode afetar o desempenho na tarefa de categorização semântica é o grau de congruência semântica entre a palavra alvo e seus vizinhos ortográficos. Por exemplo, em uma tarefa em que o participante deve decidir se uma palavra representa uma cor ou não, a palavra alvo ‘roxo’ tem um vizinho ortográfico semanticamente congruente que é a palavra ‘roxa’ e tem um vizinho ortográfico semanticamente incongruente que é a palavra ‘rolo’. Dessa forma, uma palavra com muitos vizinhos ortográficos semanticamente congruentes deveria ser reconhecida mais rapidamente do que uma com vizinhos ortográficos semanticamente incongruentes. Para investigar essa

hipótese, os pesquisadores desenvolveram uma série de experimentos, porém apresentar-se-á aqui apenas o terceiro experimento, porque esse é o mais diretamente relevante para a discussão dos efeitos de vizinhança ortográfica em tarefas de categorização semântica tradicionais. Em um estudo que contou com a participação de 74 estudantes universitários falantes da língua inglesa, Pecher e cols. (2005), manipularam o grau de congruência semântica dos vizinhos ortográficos das palavras alvo (congruentes, neutros e incongruentes), enquanto controlaram o número de vizinhos ortográficos, o número de letras e a frequência de ocorrência das palavras alvo. Os participantes deviam pressionar uma tecla quando a palavra representava algo animado e outra tecla quando a palavra representava algo inanimado. As palavras classificadas como semanticamente congruentes em relação a seus vizinhos ortográficos foram reconhecidas mais rapidamente do que as palavras classificadas como neutras e as palavras classificadas como semanticamente incongruentes foram reconhecidas mais lentamente que todas as anteriores. Porém, o efeito de congruência interagiu com o tipo de resposta (animado/inanimado) porque as palavras classificadas como semanticamente incongruentes não diferiram das neutras quando a resposta envolvia palavras inanimadas. Esse estudo é interessante porque indica que uma variável (o grau de congruência semântica entre a palavra alvo e seus vizinhos) não controlada nos estudos de vizinhança ortográfica realizados com a tarefa de categorização semântica, pode ter interferido no resultado desses estudos. No entanto, como os próprios autores ressaltaram esse estudo não traz qualquer evidência acerca dos efeitos de N (ou NF) *per se*, uma vez que o número de vizinhos ortográficos foi mantido constante entre as condições experimentais (palavras semanticamente congruentes, neutras e incongruentes).

Uma explicação alternativa para tantos efeitos conflitantes na tarefa de categorização semântica é advogar que, na realidade, os efeitos são nulos. A explicação proposta por Carreiras e cols. (1997) tem como base essa estratégia e propõe que dois critérios guiam a

resposta dos participantes na tarefa de categorização semântica. No caso das respostas positivas, uma resposta seria emitida se uma determinada quantidade de ativação no sistema semântico, relativa à categoria alvo, fosse detectada. Por exemplo, em uma tarefa em que o participante tivesse que decidir se algo é um animal ou não, uma resposta positiva seria emitida se fosse detectado no sistema semântico um aumento na ativação de unidades que codificassem características como ‘anda’, ‘respira’, entre outras. Já as respostas negativas seriam emitidas quando a ativação no sistema semântico não atingisse o nível crítico em um determinado período de tempo (de forma análoga a da estratégia *deadline* na tarefa de decisão lexical). Essa explicação prevê que as respostas positivas serão mais rápidas que as negativas e que as respostas negativas não serão guiadas por informações provenientes do sistema lexical (como as respostas negativas são disparadas quando um determinado limite de tempo expira essas nada têm a ver com o acesso lexical genuíno). Dessa forma, a proposta de Carreiras e cols. (1997) também explicaria porque os efeitos de vizinhança ortográfica não têm sido encontrados nas tarefas de categorização semântica, uma vez que, em todos os estudos com o paradigma tradicional, a manipulação experimental concentrou-se nas respostas negativas. Assim sendo, no que diz respeito à tarefa de categorização semântica, ainda é difícil se chegar a qualquer conclusão quanto aos efeitos de N ou NF.

3.6 Interpretações dos Resultados das Pesquisas e *Rationale* do Presente Estudo

De uma forma geral, pode-se dizer que nas tarefas de leitura em voz alta se têm encontrado um efeito facilitador do número de vizinhos ortográficos (N) de forma consistente. Já no caso das tarefas de identificação perceptual e nos estudos sobre o movimento dos olhos

na leitura, observaram-se efeitos inibidores de NF, embora em alguns casos, também tenham sido encontrados efeitos nulos. Dos estudos que investigaram o efeito de vizinhança ortográfica com tarefa de categorização semântica, três apresentaram resultados nulos ou conflitantes e um apresentou um efeito inibidor, mas devido ao pequeno número de estudos realizados com essa tarefa e a possível interferência de variáveis como a congruência semântica (Pecher & cols., 2005) e estratégias do tipo *deadline* (Carreiras & cols., 1997), pode-se dizer que qualquer conclusão seria precipitada. Talvez, os resultados mais controversos sejam provenientes das tarefas de decisão lexical, posto que algumas pesquisas têm encontrado um efeito facilitador de N e outras um efeito inibidor de NF, embora também tenham sido observados efeitos nulos de ambas as variáveis.

Uma das maneiras de se interpretar tais resultados é a proposta por Grainger e Jacobs (1996) e Carreiras e cols. (1997). Para esses pesquisadores, o efeito facilitador do número de vizinhos ortográficos na tarefa de decisão lexical poderia ser interpretado como resultado do uso de uma estratégia incomum (“estratégia de adivinhação rápida”), encorajada por um viés nas instruções dadas aos participantes ou pelas características das pseudopalavras. Essa estratégia obscureceria o papel do processo de identificação lexical que, em condições normais, ocorre na tarefa de decisão lexical. Se essa hipótese for aceita, poder-se-ia interpretar que a existência de um vizinho ortográfico de maior frequência de ocorrência teria um efeito inibidor nas situações em que o processo de identificação lexical estivesse mais presente, enquanto no caso de tarefas como a de leitura em voz alta, a existência de um maior número de vizinhos ortográficos facilitaria o desempenho. Uma possível explicação para esses diferentes efeitos seria argumentar que o efeito de vizinhança ortográfica ocorre em dois momentos e de maneiras diferentes (Perea & Pollatsek, 1998; Perea & Rosa, 2000). Em um primeiro momento, no qual o importante seria decodificar o estímulo visual em um conjunto de letras, o efeito de vizinhança ortográfica seria facilitador (em especial o efeito de N). Já em

um segundo momento, quando o importante seria achar a representação lexical correspondente ao estímulo alvo, o efeito seria inibidor (nesse caso, o efeito de NF seria preponderante). Como no caso das tarefas de leitura em voz alta a ênfase não é no acesso ao léxico, mas na conversão de um conjunto de grafemas em um conjunto de fonemas, o efeito facilitador do número de vizinhos ortográficos tornar-se-ia mais evidente. Já nas tarefas de identificação perceptual e versões “normais” da tarefa de decisão lexical, em que o acesso ao léxico está mais envolvido (pois a pessoa tem que “escolher” a palavra certa dentre as concorrentes), o efeito inibidor de NF seria preponderante (muitas vezes cancelando o efeito facilitador que veio primeiro).

No entanto, é importante ressaltar que a interpretação exposta acima é controversa. Andrews (1997), por exemplo, considera dúbia a análise de quais processos estariam envolvidos em cada uma das tarefas propostas por Carreiras e cols. (1997) e Perea e Pollatsek (1998), preferindo enfatizar que as contradições observadas nas tarefas de decisão lexical se devem, principalmente, às diferenças existentes entre a língua inglesa e a língua francesa e espanhola. De acordo com Andrews, na língua inglesa, as palavras monossílabas que têm muitos vizinhos ortográficos, têm uma maior probabilidade de também terem vizinhos que compartilham a mesma rima (*body*). A relação entre essas variáveis é importante, pois alguns trabalhos têm demonstrado que palavras que têm vizinhos que compartilham a mesma rima (*body neighbors*) são reconhecidas mais rapidamente na língua inglesa (Treiman, Mullennix, Bijeljac-Babic & Richmond-Welty, 1995; Ziegler & Perry, 1998). Isso aconteceria porque, no inglês, a pronúncia da vogal nos monossílabos é bastante inconsistente e a rima, nessas palavras, ajudaria a prever a pronúncia dessa vogal (Treiman & cols., 1995). Tem especial relevância para esse argumento de Andrews, o estudo de Ziegler e Perry (1998). Nesse estudo, quando se controlou o número de vizinhos ortográficos, observou-se um efeito facilitador do número de vizinhos que compartilhavam a rima com as palavras alvo (*body neighbors*) e, ao

mesmo tempo, quando o número de vizinhos que compartilhavam a rima com as palavras alvo foi controlado, encontrou-se uma tendência inibidora de N (Ziegler & Perry, 1998).

No entanto, essa proposta de Andrews (1997) também enfrenta dificuldades. Um dos problemas com a proposta é que não é claro porque as diferenças entre essas línguas se manifestariam, principalmente, nas tarefas de decisão lexical e não em outras tarefas. Estudos dos movimentos dos olhos na leitura têm apontado efeitos inibidores de NF, tanto na língua inglesa (Perea & Pollatsek, 1998; Pollatsek & cols., 1999), quanto na francesa (Grainger & cols., 1989) e experimentos com a tarefa de identificação perceptual, utilizando os mesmos estímulos de Andrews (1989), levaram a resultados contrários aos que seriam esperados (Snodgrass & Mintzer, 1993 – note-se também que N apresentou um efeito inibidor geral na condição de maior visibilidade do estudo de Sears & cols., 1999). Além disso, se a diferença nos resultados encontrados na tarefa de decisão lexical deve-se, principalmente, ao fato de que, na língua inglesa, as palavras com muitos vizinhos ortográficos teriam uma pronúncia mais previsível, era de se esperar uma diferença ainda mais saliente entre as línguas inglesa, francesa e espanhola, na tarefa de leitura em voz alta, já que nessa tarefa a pronúncia dos estímulos é obrigatória e, no entanto, isso não ocorre (Andrews, 1989, 1992; Carreiras & cols., 1997; Peereman & Content, 1995).

Um agravante que dificulta a interpretação dos resultados dos estudos sobre o efeito de vizinhança ortográfica é que boa parte dos trabalhos discutidos incorrem em um erro estatístico conhecido como “a falácia da linguagem como um efeito fixo”, o que pode ter resultado em um maior índice de erros do Tipo I nesses trabalhos (Clark, 1973; Forster & Dickinson, 1976; Raaijmakers, 2003, Raaijmakers, Schrijnemakers & Gremmen, 1999; Wickens & Keppel, 1983). Assim sendo, é importante que os trabalhos sobre vizinhança ortográfica lidem explicitamente com esse problema estatístico de forma a evitar um índice elevado de erros do Tipo I. Além disso, uma forma de se contribuir mais diretamente para a

clarificação desse debate é o desenvolvimento de estudos empíricos que tentem evitar que componentes próprios de cada tarefa obscureçam o processo de identificação lexical que se configura, quase sempre, como o interesse principal do pesquisador. Nesse sentido, um dos pontos positivos do trabalho de Grainger e Jacobs (1996) foi desenvolver um modelo que especifica as possíveis estratégias envolvidas na tarefa de decisão lexical. Assim, com base nesse modelo, pode-se tentar controlar as estratégias identificadas de forma a tornar mais provável que o processo de identificação lexical esteja sendo um dos principais responsáveis pelo desempenho na tarefa. Se, além disso, for possível especificar as estratégias que possam estar presentes em outra tarefa que também envolva acesso lexical, pode-se tentar controlar essas estratégias e comparar os resultados das duas tarefas. Caso esses resultados sejam compatíveis, obteríamos evidências ainda mais convincentes sobre a influência das variáveis estudadas no acesso lexical (Andrews & Heathcote, 2001). Outra linha de investigação que pode convergir para ajudar na interpretação dos dados é enfatizar o uso de estratégias nessas tarefas e comparar os resultados com versões dessas tarefas em que esses efeitos foram controlados (Grainger & Jacobs, 1996). Assim, pode-se comparar até que ponto a estratégia em questão afeta o desempenho na tarefa e/ou até que ponto ela é inócua, podendo ser considerada apenas um artefato interpretativo.

Tendo essas considerações em mente, o presente estudo investigou os efeitos das variáveis N e NF no reconhecimento visual de palavras em falantes do português do Brasil. Para tal, o modelo de Dupla Rota em Cascata (DRC) de Coltheart e cols. (2001) foi adotado como o arcabouço teórico utilizado para se definir as condições iniciais (Chow, 1998; Popper, 1972) em cada um dos experimentos desenvolvidos.

A escolha do modelo DRC (Coltheart & cols., 2001) tem uma justificativa objetiva. Primeiramente, é necessário considerar que ele incorpora os mecanismos propostos pelo modelo de Leitura Múltipla de Grainger e Jacobs (1996). Como ressaltado anteriormente, o

modelo de Grainger e Jacobs (1996), tem mecanismos que permitem identificar e prever as estratégias que estariam presentes na tarefa de decisão lexical. Outra vantagem do modelo de Grainger e Jacobs (1996) é o fato de ele ter sido desenvolvido para lidar especificamente com os efeitos de N e de NF. No entanto, uma das limitações desse modelo é que ele se aplica mais diretamente, apenas, às tarefas de identificação perceptual e decisão lexical (Grainger & Jacobs, 1996). Já o modelo DRC, além de ser aplicável a essas tarefas, também se aplica diretamente à tarefa de leitura em voz alta (Coltheart & cols., 2001). Sendo assim, a justificativa para adotar o modelo DRC é que ele incorpora os mecanismos propostos pelo modelo de Leitura Múltipla e, no entanto, é aplicável a mais tarefas, o que permite uma investigação mais abrangente do efeito de vizinhança ortográfica. Os outros modelos discutidos no capítulo anterior ainda não têm mecanismos que os permitam prever componentes estratégicos nas diferentes tarefas e por isso, no que diz respeito ao efeito de vizinhança ortográfica, são menos completos do que o modelo de Grainger e Jacobs (1996).

Tendo como base as previsões do modelo DRC (Coltheart & cols., 2001), o presente estudo propôs o desenvolvimento de quatro experimentos: dois com a tarefa de leitura em voz alta (um com controle de estratégias e outro em que as estratégias foram enfatizadas); um com a tarefa de identificação perceptual; e, por último, um com a tarefa de categorização semântica. A hipótese é que os resultados dos experimentos que implicam em acesso lexical sejam compatíveis e que esses, por sua vez, se diferenciem dos resultados dos experimentos onde efeitos estratégicos podem estar presentes. Para contrastar e/ou combinar os resultados dos diversos experimentos realizados nessa tese e também para compará-los com outros estudos realizados com falantes do português brasileiro, técnicas meta-analíticas foram utilizadas.

Os capítulos sobre os experimentos realizados trazem, de forma detalhada, a descrição das diversas tarefas que foram utilizadas no presente trabalho, bem como, a discussão das

possíveis estratégias envolvidas nessas e como se tentou controlá-las de acordo com as previsões do modelo DRC. O capítulo ‘Método’, discute o problema da “linguagem como um efeito fixo” e é fundamental para o entendimento dos procedimentos empregados na análise dos dados dos experimentos propostos nesse estudo. Já o capítulo ‘Meta-Análise e discussão geral’ traz um comparativo dos diversos estudos realizados com falantes do português brasileiro, incluindo os dessa tese, tendo em mente a hipótese de que os efeitos de N e NF podem ser modulados pelo tipo de demanda das tarefas utilizadas nos estudos.

3.6.1 Objetivos

O objetivo mais geral do presente trabalho foi o de investigar o efeito de vizinhança ortográfica no reconhecimento visual de palavras em adultos falantes do português do Brasil. Para tanto, o modelo DRC de Coltheart e cols. (2001) foi utilizado como um modelo explicativo dos componentes que estariam presentes nas tarefas utilizadas no presente estudo. Tal escolha teve a finalidade de permitir identificar os processos de acesso lexical e as estratégias de resolução de problemas que poderiam estar presentes nessas tarefas.

A partir disso, o presente estudo objetivou investigar os efeitos de N e de NF, tanto nos processos de identificação lexical, quanto no processamento estratégico em falantes do português brasileiro. Para tanto, traçou-se uma comparação entre as tarefas nas quais o processo de identificação lexical deveria estar mais implicado e depois, essas foram contrastadas com as tarefas que favoreceram a utilização de processamento estratégico.

Por último, em decorrência da teoria adotada, o objetivo mais específico foi avaliar se o DRC é um modelo explicativo adequado para descrever os efeitos de N e NF no processo de acesso ao léxico ortográfico no português do Brasil.

3.6.2 Justificativas

Pode-se dizer que o presente trabalho contribui para a literatura na área de reconhecimento visual de palavras de diferentes formas. Por um lado, testa se um importante modelo de reconhecimento visual de palavras descreve, adequadamente, o efeito de vizinhança ortográfica no processo de identificação lexical em falantes do português brasileiro, contribuindo assim, com informações sobre a generalidade desse modelo. Além disso, o método empregado no presente trabalho visou revelar os componentes e as estratégias de processamento presentes em diversas tarefas que são, tradicionalmente, utilizadas em estudos de reconhecimento visual de palavras. Conseqüentemente, com base nos resultados desse estudo, outros pesquisadores podem avaliar qual variedade dessas tarefas melhor atende aos objetivos de suas pesquisas. Por fim, ao investigar os efeitos de N e NF, os resultados desse trabalho lançam luz sobre a organização e o acesso ao léxico ortográfico em falantes do português do Brasil e servem, por conseguinte, para impor restrições a futuros modelos de reconhecimento visual de palavras que visem explicá-los. Desse modo, indiretamente, essas informações podem contribuir para esclarecer se o efeito de vizinhança ortográfica varia de acordo com a língua estudada ou se ele é independente dessa.

4 ANÁLISE DE CORPUS

A finalidade desse capítulo é descrever algumas análises que foram desenvolvidas para dar suporte à seleção dos estímulos experimentais e por isso precede o capítulo ‘método’ que versa sobre os procedimentos comuns aos experimentos desenvolvidos. A primeira questão que se levantou diz respeito a qual lista de contagem de frequência de ocorrência de palavras do português brasileiro deveria ser utilizada. Uma vez selecionada essa lista, outra questão importante que se impôs foi investigar a distribuição e a correlação das diferentes medidas de vizinhança ortográfica (N, NF e TLN) nas palavras do português brasileiro. Essa investigação serviu de suporte empírico para a decisão acerca de qual definição de vizinhança ortográfica deveria ser adotada nos experimentos do presente estudo. Além disso, esse corpo de dados permitiu comparar a estrutura da vizinhança ortográfica das palavras do português com a das palavras da língua inglesa. A seguir, encontra-se a descrição e a discussão desses estudos.

4.1 Estudo de Familiaridade

Existem, pelo menos, duas listas contendo uma contagem da frequência de ocorrência das palavras do português do Brasil: a lista de Pinheiro (1996) e a lista do *Corpus* NILC / São Carlos (NILC, 2005). A lista de Pinheiro foi desenvolvida a partir de amostras de palavras das publicações utilizadas mais frequentemente nas redes de ensino de Belo Horizonte/BR para crianças do primeiro ao quinto ano do ensino fundamental e tem sido utilizada com razoável

sucesso, mesmo com estudantes universitários (ver, por exemplo, Capovilla, Macedo, Duduchi & Sória, 1997; Justi & Pinheiro, 2006). Já a lista de palavras do NILC foi desenvolvida a partir de uma amostra mais ampla de textos (envolvendo textos brasileiros do registro jornalístico, didático, epistolar e redações de alunos), porém o material mais significativamente aproveitado no corpus decorre de textos do jornal ‘A Folha de São Paulo’ (LINGUATECA, 2009). A escolha entre uma dessas duas listas de palavras se tornou imperiosa no presente estudo, pois nenhuma delas traz estatísticas sobre as variáveis de vizinhança ortográfica e assim sendo, seria necessário computar essas estatísticas para uma amostra extensa de palavras dessas listas para que os estímulos experimentais desse estudo pudessem ser selecionados adequadamente. Desse modo, no presente estudo, optou-se por um critério empírico para se decidir entre essas duas listas.

Pode-se dizer que existem, pelo menos, duas formas de se estimar a frequência de ocorrência das palavras de uma língua: por meio de medidas objetivas e por meio de medidas subjetivas. Medidas objetivas, como a lista de Pinheiro (1996) e do NILC (2005), envolvem a contagem de palavras em amostras de textos aos quais, determinada população, normalmente é exposta. Já as medidas subjetivas envolvem o julgamento dos leitores sobre o quão frequentemente ouvem ou encontram em sua forma escrita, determinada palavra – esse tipo de medida é conhecido como *índice de familiaridade* (Balota, Piloti & Cortese, 2001). Dessa forma, para se decidir sobre qual lista de palavras utilizar nesse estudo, optou-se por selecionar uma amostra de palavras comuns às duas listas (a lista de Pinheiro e a lista do NILC) e correlacionar a frequência de ocorrência calculada por essas listas para esse grupo de palavras com o índice de familiaridade estimado por um grupo de universitários para essas mesmas palavras. Assim, o critério empírico adotado no presente estudo foi o de utilizar a lista que apresentasse a maior correlação com o julgamento subjetivo de frequência realizado pelo grupo de universitários.

4.1.1 Método

4.1.1.1 Participantes

Participaram dessa pesquisa 22 alunos do curso de Psicologia da UFPE. Os dados de dois alunos foram descartados, pois esses deixaram de julgar a frequência de ocorrência de mais de 20% das palavras. A média de idade para os 20 participantes remanescentes foi de 21 anos e 1 mês (desvio padrão de 4 anos e 3 meses). Sendo 16 do sexo feminino e 4 do sexo masculino. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Pernambuco / UFPE (número de registro CEP/CCS/UFPE: 189/07)

4.1.1.2 Material

Quatrocentas palavras de quatro a seis letras comuns às listas de Pinheiro (1996) e do NILC (2005) foram utilizadas. Optou-se por palavras com esse número de letras porque essas representam a maioria dos estímulos utilizados nas pesquisas sobre vizinhança ortográfica (Andrews, 1997).

4.1.1.3 Procedimentos

Como a lista de Pinheiro (1996) contém menos palavras do que a lista do NILC (2005) utilizou-se o programa ‘*SPSS for Windows*’ versão 16.0 para fazer a seleção randômica de 400 palavras de quatro a seis letras da lista de Pinheiro (1996) referentes ao quinto ano do ensino fundamental. A partir disso, as palavras que também estavam presentes na lista do NILC (2005) foram mantidas e as palavras que constavam apenas na lista de Pinheiro descartadas. Foram então sorteadas, novamente, palavras de quatro a seis letras da lista de Pinheiro (1996), porém, dessa vez, a amostra de palavras selecionadas correspondia ao número de palavras que faltava para que se completassem as 400 palavras comuns às duas listas. Esse procedimento foi repetido até que se obtivessem 400 palavras de quatro a seis letras que fossem comuns às duas listas. As 400 palavras selecionadas foram então divididas em quatro listas, sendo que, em cada lista, a ordem de disposição das palavras foi sorteada. Para que possíveis efeitos de ordem de apresentação pudessem ser reduzidos, a ordem das listas foi contrabalanceada entre os participantes.

A aplicação das listas foi coletiva e ocorreu em uma sala de aula da UFPE em um horário previamente acordado com os participantes. Os participantes receberam as listas e então as seguintes instruções (que foram adaptadas do estudo de Balota & cols., 2001, p.641):

As palavras diferem em quão comumente ou freqüentemente são encontradas. Algumas palavras são encontradas muito freqüentemente, enquanto outras são encontradas com pouca freqüência. O propósito desse estudo é estimar o quão freqüentes, em sua forma escrita, são as palavras de uma lista. Essa lista contém 400 palavras. Acreditamos que as suas estimativas serão importantes para estudos futuros envolvendo o reconhecimento visual de palavras. Você deve basear suas estimativas de acordo com uma escala de 7 pontos (use apenas números inteiros). Utilize o seguinte raciocínio:

Encontro essa palavra em sua forma escrita ...

1 = nunca; 2 = uma vez por ano; 3 = uma vez por mês; 4 = uma vez por semana;
5 = a cada dois dias; 6 = uma vez por dia; 7 = muitas vezes por dia.

Dadas essas instruções, quaisquer dúvidas quanto as mesmas foram redimidas e os participantes foram instruídos a assinalar os valores referentes ao seu julgamento quanto à frequência de ocorrência da palavra em um espaço delimitado ao lado dela. Os participantes foram encorajados a prover suas estimativas para todas as palavras. Não se estabeleceu nenhum tempo limite para o preenchimento das listas, mas os participantes levaram aproximadamente de 30 a 35 minutos para preenchê-las.

4.1.2 Resultados

A média da frequência de ocorrência estimada pelos participantes foi calculada para cada uma das 400 palavras. Com base nesses valores, o índice de familiaridade médio para as 400 palavras foi de 3,9 (com desvio padrão de 1,12). Para facilitar a comparação entre as listas e a comparação com os resultados de outros estudos, a frequência de ocorrência bruta relatada nas listas de Pinheiro (1996) e do NILC (2005), para as 400 palavras, foi convertida em um índice que expressa a frequência de ocorrência por milhão de palavras (que é o padrão mais utilizado em estudos sobre o reconhecimento visual de palavras). Com base no índice de ocorrência por milhão de palavras também foi calculada a contrapartida logarítmica ($\log_{10}(x+1)$) dessa nova medida de frequência para ambas as listas, pois, de acordo com Whaley (1978), a frequência logarítmica tem uma relação mais forte com o tempo de reação do que a frequência absoluta. A Tabela 1 apresenta a média e o desvio padrão desses valores para as 400 palavras do estudo.

Tabela 1 – Estatísticas descritivas para a o índice de familiaridade médio (IFM), frequência de ocorrência por milhão de palavras (FpM) e frequência logarítmica (FLog) de 400 palavras

	Média	Desvio Padrão
IFM	3,90	1,12
FpM (lista Pinheiro, 1996)	66,94	177,05
FpM (lista NILC, 2005)	38,19	97,37
FLog (lista Pinheiro, 1996)	1,23	0,66
FLog (lista NILC, 2005)	1,01	0,67

Pode-se observar na Tabela 1, que as palavras da lista de Pinheiro (1996) apresentam uma maior frequência de ocorrência do que as mesmas palavras na lista do NILC (2005). Além disso, em média, os participantes julgaram ser expostos a essas palavras em sua forma escrita, pelo menos, uma vez por semana. Análises de correlação entre essas diferentes medidas de frequência de ocorrência foram calculadas tendo como base essa amostra de 400 palavras comuns as duas listas e são apresentadas na Tabela 2.

Tabela 2 – Correlações entre diferentes medidas de frequência de ocorrência para uma amostra de 400 palavras

	IFM	FpM (Pinheiro, 1996)	FLog (Pinheiro, 1996)	FpM (NILC, 2005)
FpM (Pinheiro, 1996)	0,45			
FLog (Pinheiro, 1996)	0,56	0,70		
FpM (NILC, 2005)	0,45	0,68	0,56	
FLog (NILC, 2005)	0,69	0,55	0,73	0,68

Nota. IFM = índice de familiaridade médio; FpM = frequência de ocorrência por milhão de palavras; FLog = frequência logarítmica. Todas as correlações foram significantes para $p \leq 0,01$.

Como se pode observar na Tabela 2, no geral, as correlações entre as medidas foram de razoáveis para fortes, sendo todas significantes (para todas $p \leq 0,01$). As correlações entre as frequências logarítmicas das duas listas e o índice de familiaridade foram maiores do que as correlações entre a frequência de ocorrência por milhão de palavras das duas listas e o índice de familiaridade. A correlação entre a frequência logarítmica das duas listas foi de 0,73. No que diz respeito à correlação com o índice de familiaridade calculado com base no julgamento subjetivo dos participantes, a medida que mais fortemente se correlacionou foi a

freqüência logarítmica da lista do NILC (2005), seguida pela freqüência logarítmica da lista de Pinheiro (1996), sendo as correlações de 0,69 e 0,56, respectivamente.

4.1.3 Discussão

Como esperado, devido aos achados de Whaley (1978), as correlações entre as freqüências logarítmicas e o índice de familiaridade foram mais fortes do que a correlação desse índice com as freqüências de ocorrências por milhão de palavras. As duas listas de palavras apresentaram uma forte correlação, o que pode explicar o relativo sucesso que os pesquisadores têm tido ao utilizar a lista de Pinheiro (1996) em estudos com adultos (Capovilla & cols., 1997; Justi & Pinheiro, 2006). No entanto, a lista de palavras que apresentou a correlação mais forte com o índice de familiaridade foi a do NILC (2005). Assim sendo, do ponto de vista empírico, essa parece ser a lista de palavras mais adequada para estudos com adultos e, por conseguinte, foi a utilizada no presente estudo.

4.2 A Estrutura de Vizinhança Ortográfica das Palavras do Português Brasileiro

Tendo optado pela lista de palavras do *corpus* NILC (2005), outra questão que se tornou de fundamental importância no presente estudo foi o cômputo de estatísticas quanto à distribuição dos índices de vizinhança ortográfica nas palavras do português do Brasil. Como discutido no capítulo ‘Modelos de Reconhecimento Visual de Palavras’, alguns autores, como

Grainger e Whitney (2004) e Whitney e Lavidor (2005), questionam as definições mais tradicionais de vizinhança ortográfica baseadas em N e NF. Para esses autores, as medidas tradicionais pecam por não incluir em seu cômputo vizinhos ortográficos formados pela transposição de duas letras (p.ex.: ‘acesos’ e ‘acesso’), conhecidos como *transposed letter neighbors* – TLN, em inglês. Como existe certa evidência de que palavras com vizinhos transpostos são reconhecidas mais lentamente do que palavras sem vizinhos transpostos (Andrews, 1996), há a possibilidade de que os efeitos de N e NF sejam meros subprodutos dos efeitos de TLN. Uma forma de se averiguar essa questão é observar até que ponto N e NF se correlacionam com TLN, pois, se a correlação entre essas variáveis e TLN for muito pequena, dificilmente poder-se-ia explicar os efeitos encontrados de N e NF com base em sua correlação com TLN. Colocando de outra forma, se N e NF não se correlacionam com TLN, deve-se admitir que essas variáveis podem ter efeitos genuínos e independentes de TLN e que, muito provavelmente, são esses efeitos que têm sido reportados nos estudos de vizinhança ortográfica.

Outra questão a se considerar é que existem argumentos razoáveis de que as diferenças entre as diversas línguas nas quais as pesquisas de vizinhança ortográfica foram desenvolvidas podem explicar as variações encontradas nos efeitos de N e NF (Andrews, 1997; Siakaluk & cols., 2002; Ziegler & Perry, 1998). A sugestão mais comum e que foi discutida no capítulo ‘Revisão da Literatura’, é que a diferença na consistência do mapeamento grafema-fonema entre as línguas pode ser responsável por tais diferenças. É bastante comum nos estudos sobre a leitura, classificar as diferentes línguas de acordo com o seu grau de profundidade ortográfica (*orthographic depth*), normalmente em raso (*shallow*) – quando o mapeamento grafema-fonema é bastante consistente ou inequívoco – ou profundo (*deep*) – quando o mapeamento grafema-fonema é mais inconsistente. Um exemplo desse tipo de classificação é a proposta por Seymour (2005) que pode ser observada na Tabela 3.

Tabela 3 – Exemplo de classificação das línguas européias de acordo com a profundidade ortográfica (Tabela adaptada de Seymour, 2005)

Profundidade Ortográfica				
Rasa				Profunda
Finlandês	Grego	Português	Francês	Inglês
	Italiano	(Portugal)		
	Espanhol			

Tendo essa classificação de Seymour (2005) em mente, pode-se retomar, brevemente, o argumento exposto no capítulo ‘revisão da literatura’. No que diz respeito ao efeito de vizinhança ortográfica, o que tem sido proposto é que no inglês, que é uma língua mais profunda do ponto de vista ortográfico, a rima teria um papel importante, pois ela ajuda a prever a pronúncia das vogais que é bastante inconsistente nos monossílabos (Treiman & cols., 1995). Uma variável que parece ajudar no processamento da rima é o número de vizinhos da palavra que compartilham com ela a mesma rima (*body neighbors* – Treiman & cols., 1995; Ziegler & Perry, 1998) e, como é provável que palavras com muitos vizinhos ortográficos tenham também muitos *body neighbors*, isso faria com que o efeito de N na língua inglesa tendesse a ser facilitador (Andrews, 1997). Já em línguas com mapeamentos menos inconsistentes (como o espanhol e o português, por exemplo), existe a possibilidade do número de *body neighbors* não desempenhar um papel tão especial quanto o desempenhado no inglês, porque nessas línguas um mapeamento direto grafema-fonema já seria bastante eficiente para gerar a pronúncia das palavras (Andrews, 1997).

É interessante notar que a explicação exposta no parágrafo anterior explora diferenças entre as línguas no que diz respeito ao grau de consistência do mapeamento grafema-fonema. No entanto, outra possível fonte de diferenças entre as línguas pode advir de diferenças na estrutura da vizinhança ortográfica dessas línguas, isso é, diferentes distribuições de N e NF nas palavras da língua. Por exemplo, Siakaluk e cols. (2002) argumentam que, como mais da metade das palavras de língua inglesa de três a cinco letras têm vizinhos ortográficos de maior frequência de ocorrência, seria bastante contra-intuitivo imaginar um mecanismo de

processamento da informação que prejudicasse o desempenho de leitura na maioria das palavras. Ora, o que Siakaluk e cols. parecem estar argumentando é que, de alguma forma, a própria distribuição estatística de N e NF nas palavras da língua inglesa pode levar o nosso sistema cognitivo a tirar proveito dessas variáveis e, no caso do inglês, levar a um efeito facilitador. Dessa forma, uma hipótese diretamente relacionada seria a de que, em uma língua em que a estrutura da vizinhança ortográfica fosse diferente e a maioria das palavras não tivesse vizinhos de maior frequência de ocorrência, o efeito de N e NF deveria ser diferente do encontrado na língua inglesa (sendo, provavelmente, inibidor). Essa hipótese torna importante que se compare a estrutura da vizinhança ortográfica das palavras do português com a das palavras da língua inglesa, para se averiguar se diferenças nessas distribuições também poderiam explicar as variações encontradas nos resultados das pesquisas.

Tendo essas considerações em mente, o presente estudo desenvolveu uma análise de corpus lingüístico com a finalidade de investigar a distribuição e a correlação das diferentes medidas de vizinhança ortográfica (N, NF e TLN) nas palavras do português brasileiro. Essa investigação teve duas finalidades: embasar a decisão acerca de qual definição de vizinhança ortográfica deveria ser adotada nos experimentos do presente estudo; e, permitir comparar a estrutura da vizinhança ortográfica das palavras do português com a das palavras da língua inglesa.

4.2.1 Método

4.2.1.1 Material

Foram utilizadas no presente estudo 8465 palavras de quatro a seis letras do corpus NILC (2005).

4.2.1.2 Procedimentos

Foram selecionadas para a análise, inicialmente, todas as palavras do corpus NILC (2005) que tinham de quatro a seis letras. Então, a frequência de ocorrência bruta relatada para essas palavras pelo corpus NILC foi convertida em um índice que expressa a frequência de ocorrência por milhão de palavras e foram selecionadas apenas as palavras com frequência de, pelo menos, uma ocorrência por milhão. Posteriormente, todas as palavras estrangeiras e hifenizadas foram descartadas e a análise final foi desenvolvida com base nas 8465 palavras restantes. Essas 8465 palavras foram organizadas em um banco de dados e o programa ‘*N_Watch*’ (Davis, 2005) foi utilizado como auxílio para o cômputo das seguintes estatísticas: FLog – contrapartida logarítmica ($\log_{10}^{(x+1)}$) da frequência por milhão de ocorrências da palavra; N – número de vizinhos ortográficos da palavra; NPV ou *spread* – número de posições da palavra em que, alterando-se uma letra, vizinhos ortográficos são gerados (p.ex. ‘*missa*’ tem um NPV igual a dois, pois tem vizinhos formados por mudança na segunda e terceira letras, ‘*massa*’ e ‘*mista*’, respectivamente); NVmF – número de vizinhos de menor frequência que a palavra; NF – número de vizinhos de maior frequência de ocorrência que a palavra; P_NF – presença (1) ou ausência (0) de, pelo menos, um vizinho de maior

freqüência; e, P_TLN – presença (1) ou ausência (0) de, pelo menos, um vizinho ortográfico formado pela transposição de duas letras da palavra.

4.2.2 Resultados

Para facilitar a visualização e a comparação com os dados existentes para a língua inglesa, optou-se por apresentar estatísticas descritivas para o total de palavras e também estatísticas separadas para palavras de quatro, cinco e seis letras (como no trabalho de Andrews, 1997). O mesmo ocorreu para as análises de correlação. A Tabela 4 apresenta as estatísticas descritivas para o total de 8465 palavras analisadas.

Tabela 4 – Estatísticas descritivas para uma amostra de 8465 palavras de quatro a seis letras retiradas do *corpus* NILC (2005)

	Média	Desvio Padrão
Frequência de ocorrência por milhão de palavras (FpM)	31,02	173,82
Frequência Logarítmica (FLog)	0,85	0,60
Número de letras (L)	5,39	0,71
Número de vizinhos ortográficos (N)	2,76	2,95
Número de posições em que se geram vizinhos (NPV)	1,65	1,28
Número de vizinhos de menor freqüência (NVmF)	1,52	2,01
Número de vizinhos de maior freqüência (NF)	1,24	1,8
Porcentagem com vizinhos de maior freqüência (%NF)	54,11	49,83
Porcentagem com vizinhos formados pela transposição de duas letras (%TLN)	3,86	19,27

No que diz respeito às principais variáveis de vizinhança ortográfica, pode-se observar na Tabela 4 que as palavras do português brasileiro de quatro a seis letras têm, em média, aproximadamente, três vizinhos ortográficos (N) e um vizinho ortográfico de maior freqüência de ocorrência (NF), sendo que, geralmente, os vizinhos das palavras são gerados por mudanças em uma ou duas posições das palavras (NPV). Além disso, pelo menos, 54%

das palavras têm vizinhos ortográficos de maior frequência de ocorrência (%NF), enquanto, aproximadamente, apenas 4% do total das palavras de quatro a seis letras têm vizinhos formados pela transposição de duas letras (%TLN). No caso da língua inglesa, Andrews (1997) apresentou apenas estatísticas específicas pelo número de letras, mas como a pesquisadora relatou o número de palavras de quatro, de cinco e de seis letras de sua amostra, tornou-se possível o cálculo da média ponderada para essas estatísticas de vizinhança ortográfica a exceção de %TLN, pois o estudo de Andrews não apresentou dados quanto à porcentagem de palavras que têm vizinhos ortográficos formados pela transposição de duas letras na língua inglesa. Sendo assim, a média ponderada para N, NF, NPV e %NF na língua inglesa é respectivamente de: 2,81 vizinhos ortográficos; 1,33 vizinhos de maior frequência de ocorrência; 1,53 posições, em média, nas quais os vizinhos podem ser gerados; e, 47,85% das palavras de quatro a seis letras têm vizinhos ortográficos de maior frequência de ocorrência. As análises de correlação para o total das palavras de quatro a seis letras do português do Brasil encontram-se na Tabela 5.

Tabela 5 – Correlações entre medidas de vizinhança ortográfica, frequência e número de letras para uma amostra de 8465 palavras do português do Brasil

	FpM	FLog	L	N	NPV	NVmF	NF
FLog	0,44						
L	-0,10	-0,14					
N	0,08	0,15	-0,49				
NPV	0,06	0,15	-0,33	0,82			
NVmF	0,19	0,42	-0,37	0,80	0,68		
NF	-0,08	-0,23	-0,39	0,74	0,59	0,20	
TLN	0,03	0,03	-0,12	0,10	0,11	0,07	0,08

Nota. FpM = frequência de ocorrência por milhão de palavras; FLog = frequência logarítmica; L = número de letras; N = número de vizinhos ortográficos; NPV = número de posições em que podem ser gerados vizinhos; NVmF = número de vizinhos de menor frequência; NF = número de vizinhos de maior frequência; TLN = presença de um vizinho ortográfico formado pela transposição de duas letras. Todas as correlações foram significantes para $p \leq 0,05$.

Como pode ser observado na Tabela 5, todas as correlações foram significantes. No entanto, uma análise cuidadosa indica que isso se deveu mais ao tamanho da amostra do que,

propriamente, a força das correlações. As correlações entre N e TLN e entre NF e TLN, foram bastante pequenas: 0,1 para a correlação entre N e TLN e 0,08 para a correlação entre NF e TLN. Dessa forma, apenas 1% da variação em N poderia ser explicada pela variação em TLN e apenas 0,6 % da variação em NF poderia ser explicada pela variação em TLN. O número de vizinhos ortográficos (N) apresentou uma forte correlação positiva com o número de vizinhos ortográficos de maior frequência (NF). O número de posições em que se podem gerar vizinhos (NPV) apresentou uma forte correlação positiva com N e uma correlação positiva moderada com NF. O número de letras apresentou uma correlação negativa moderada com N e uma correlação negativa, um pouco menor, com NF. Já a frequência logarítmica apresentou uma correlação positiva moderada com o número de vizinhos de menor frequência (NVMF) e uma pequena correlação negativa com o número de vizinhos de maior frequência (NF).

Para facilitar a visualização comparativa dos dados estatísticos da estrutura da vizinhança ortográfica das palavras do português do Brasil e das palavras da língua inglesa, esses foram ambos organizados na Tabela 6. Essa Tabela provê as estatísticas descritivas da estrutura da vizinhança ortográfica das duas línguas por número de letras, sendo que os dados da língua inglesa foram extraídos do trabalho de Andrews (1997). Além disso, como no trabalho de Andrews não foram relatadas estatísticas quanto aos vizinhos que podem ser formados pela transposição de duas letras (TLN), nem quanto à frequência logarítmica, e como as estatísticas dessas variáveis já foram descritas nas análises gerais (ver Tabela 4 e Tabela 5), optou-se por não incluir estatísticas para essas variáveis na Tabela 6.

Tabela 6 – Estatísticas descritivas de medidas de vizinhança ortográfica e de frequência de ocorrência para palavras do português do Brasil e do Inglês (adaptado de Andrews, 1997)

	Português	Inglês
<i>Palavras de quatro letras</i>		
	1106 palavras	1895 palavras
	Média (DP)	Média (DP)
Frequência de ocorrência por milhão de palavras (FpM)	73,0 (415,0)	100,3 (549,8)
Número de vizinhos ortográficos (N)	6,0 (4,4)	7,2 (4,9)
Número de posições em que se geram vizinhos (NPV)	2,4 (1,2)	2,5 (1,1)
Número de vizinhos de menor frequência (NVmF)	3,2 (3,2)	3,4 (3,8)
Número de vizinhos de maior frequência (NF)	2,8 (2,9)	3,5 (3,4)
Porcentagem com vizinhos de maior frequência (%NF)	74,5	80,3
<i>Palavras de cinco letras</i>		
	2968 palavras	2895 palavras
	Média (DP)	Média (DP)
Frequência de ocorrência por milhão de palavras (FpM)	33,3 (107,5)	34,2 (153,3)
Número de vizinhos ortográficos (N)	3,3 (2,8)	2,4 (2,3)
Número de posições em que se geram vizinhos (NPV)	2,0 (1,3)	1,5 (1,2)
Número de vizinhos de menor frequência (NVmF)	1,8 (2,0)	1,1 (1,6)
Número de vizinhos de maior frequência (NF)	1,5 (1,8)	1,1 (1,5)
Porcentagem com vizinhos de maior frequência (%NF)	62,1	52,0
<i>Palavras de seis letras</i>		
	4391 palavras	4166 palavras
	Média (DP)	Média (DP)
Frequência de ocorrência por milhão de palavras (FpM)	18,9 (80,6)	16,5 (62,6)
Número de vizinhos ortográficos (N)	1,6 (1,6)	1,1 (1,6)
Número de posições em que se geram vizinhos (NPV)	1,2 (1,1)	0,8 (0,9)
Número de vizinhos de menor frequência (NVmF)	0,9 (1,2)	0,5 (1,1)
Número de vizinhos de maior frequência (NF)	0,7 (1,0)	0,5 (1,0)
Porcentagem com vizinhos de maior frequência (%NF)	43,6	30,2

Nota. DP = Desvio Padrão. Os dados da lingual inglesa foram extraídos do trabalho de Andrews (1997) e reproduzidos com a permissão da *Psychonomic Bulletin and Review*.

Pode-se observar na Tabela 6 que as estatísticas descritivas de vizinhança ortográfica para as duas línguas são razoavelmente similares. No entanto, devido ao tamanho excessivamente grande das amostras, ao se calcular testes *t* para amostras independentes com base nos dados fornecidos, praticamente todas as médias diferiram significativamente. As únicas médias que não diferiram significativamente foram as de frequência de ocorrência (devido aos grandes desvios padrão) e as do número de vizinhos de menor frequência (NVmF) para palavras de quatro letras. Em situações como essas (em que mesmo pequenas diferenças são estatisticamente significativas devido ao tamanho da amostra) a magnitude das diferenças acaba sendo mais informativa do que os valores de *p* (Cohen, 1994). Destarte, optou-se por calcular o *d'* de Cohen (1988) como uma estimativa da magnitude da diferença

entre as médias. Com base nos valores de d' , todas as diferenças entre as médias ficaram abaixo do valor de 0,5, que é o critério proposto por Cohen (1988) para que uma diferença seja considerada razoável (média). As maiores diferenças entre as línguas ocorreram para as palavras de cinco e de seis letras e foram entre o número de posições em que se geram vizinhos (NPV), em ambos os casos o valor de d' foi de 0,4. Outras diferenças que merecem menção ocorreram entre o número de vizinhos de menor frequência (NVmF) e o número de vizinhos ortográficos (N), também no caso das palavras de cinco letras e de seis letras. Para palavras de cinco letras, o valor d' para a diferença em NVmF entre as línguas foi de 0,39 e para a diferença em N foi de 0,35. No caso das palavras de seis letras, o valor d' para a diferença em NVmF entre as línguas foi de 0,35 e para a diferença em N foi de 0,31. A diferença entre o número de vizinhos de maior frequência de ocorrência (NF) entre as línguas manteve-se razoavelmente constante, independente do número de letras (valores d' de 0,22, 0,24 e 0,20 para palavras de quatro, cinco e seis letras, respectivamente). No que diz respeito às palavras de quatro letras, o valor d' para a diferença entre as línguas no número de vizinhos ortográficos (N) foi de 0,26 que é um valor considerado pequeno, enquanto as demais diferenças foram ínfimas ($d' < 0,1$). Por fim, Andrews (1997) não relatou o desvio padrão para a porcentagem de palavras com vizinhos de maior frequência de ocorrência (%NF) em seu estudo, fato que levou a omissão de tal dado na Tabela 6. Porém, utilizando o desvio padrão calculado com base nos dados do presente estudo, os valores d' para as diferenças entre as línguas em %NF foi pequeno (0,13 para palavras de quatro; 0,21 para palavras de cinco e 0,27 para palavras de seis letras).

A Tabela 7 apresenta as correlações entre as diferentes medidas de vizinhança ortográfica e frequência de ocorrência no português do Brasil e no inglês de acordo com o número de letras. Os dados da língua inglesa encontram-se entre parênteses e foram retirados do trabalho de Andrews (1997) que não relatou quais correlações foram significantes ou não.

Tabela 7 – Correlações entre medidas de vizinhança ortográfica e frequência, para palavras de 4 letras, 5 letras (itálico) e 6 letras (negrito) do português brasileiro e do inglês (entre parênteses – adaptado de Andrews, 1997)

		FpM	FLog	N	NPV	NVmF
FLog	4 letras	0,45 (0,45)				
	<i>5 letras</i>	<i>0,65 (0,52)</i>				
	6 letras	0,54 (0,62)				
N	4 letras	0,04 (0,03)	0,07 (0,24)			
	<i>5 letras</i>	<i>0,04 (0,05)</i>	<i>0,11 (0,20)</i>			
	6 letras	0,00 (0,01)	0,09 (0,08)			
NPV	4 letras	0,04 (0,03)	0,10 (0,21)	0,80 (0,72)		
	<i>5 letras</i>	<i>0,05 (0,03)</i>	<i>0,12 (0,16)</i>	<i>0,85 (0,82)</i>		
	6 letras	0,01 (0,01)	0,09 (0,09)	0,92 (0,87)		
NVmF	4 letras	0,18 (0,19)	0,48 (0,70)	0,74 (0,67)	0,62 (0,49)	
	<i>5 letras</i>	<i>0,22 (0,18)</i>	<i>0,44 (0,55)</i>	<i>0,77 (0,69)</i>	<i>0,66 (0,54)</i>	
	6 letras	0,10 (0,14)	0,32 (0,37)	0,79 (0,70)	0,74 (0,60)	
NF	4 letras	-0,14 (-0,15)	-0,42 (-0,36)	0,70 (0,65)	0,54 (0,45)	0,04 (-0,11)
	<i>5 letras</i>	<i>-0,18 (-0,11)</i>	<i>-0,32 (-0,21)</i>	<i>0,69 (0,70)</i>	<i>0,58 (0,58)</i>	<i>0,07 (0,01)</i>
	6 letras	-0,11 (-0,10)	-0,24 (-0,18)	0,67 (0,73)	0,60 (0,64)	0,08 (0,09)

Nota. FpM = frequência de ocorrência por milhão de palavras; FLog = frequência logarítmica; N = número de vizinhos ortográficos; NPV = número de posições em que podem ser gerados vizinhos; NVmF = número de vizinhos de menor frequência; NF = número de vizinhos de maior frequência. Os dados referentes à língua inglesa encontram-se entre parênteses e foram extraídos do trabalho de Andrews (1997), tendo sido reproduzidos com a permissão da *Psychonomic Bulletin and Review*.

Como observado anteriormente, Andrews (1997) não relatou quais correlações em seu estudo foram significantes, no entanto, devido ao tamanho das amostras, é esperado que todas as correlações relevantes tenham sido significantes. Afinal, no caso das palavras do presente estudo, só não foram significantes as correlações apresentadas na Tabela 7 que foram menores que 0,042, no caso das palavras de quatro letras, e as menores que 0,03, no caso das palavras de seis letras. De uma forma geral, as correlações entre as variáveis de vizinhança ortográfica no português do Brasil e no inglês foram muito similares, apresentando quase sempre a mesma magnitude e o mesmo padrão. As maiores correlações, para ambas as línguas, ocorreram entre o número de vizinhos ortográficos (N) e o número de posições onde se geram vizinhos (NPV), seguidas da correlação entre N e o número de vizinhos de menor frequência (NVmF) e da correlação entre N e o número de vizinhos de maior frequência de ocorrência (NF). No que diz respeito à correlação das medidas de vizinhança ortográfica com a frequência de ocorrência das palavras do português do Brasil, o número de vizinhos de menor

freqüência de ocorrência (NVmF) apresentou uma correlação positiva com a freqüência logarítmica (FLog) que foi de moderada (para palavras de quatro e cinco letras) à fraca (para palavras de seis letras). Enquanto o número de vizinhos de maior freqüência de ocorrência (NF) apresentou uma correlação negativa com FLog que também foi de moderada (para palavras de quatro letras) à fraca (para palavras de cinco e seis letras). No caso das palavras da língua inglesa, as correlações apresentaram o mesmo padrão (diminuíram com o aumento no número de letras), mas apresentaram maior magnitude que as correlações da língua portuguesa nas correlações entre NVmF e FLog e menor magnitude nas correlações de NF e FLog. Por fim, para ambas as línguas, as correlações entre o número de vizinhos ortográficos (N) e a freqüência logarítmica (FLog) e as correlações entre o número de posições em que vizinhos podem ser gerados (NPV) e FLog, foram pequenas.

4.2.3 Discussão

Considerando-se inicialmente a questão da definição de vizinhança ortográfica, torna-se necessário que se analise a relação entre N, NF e TLN. Como pôde ser observado nas análises desenvolvidas (especialmente Tabelas 4 e 5), as correlações entre N e TLN e entre NF e TLN são ínfimas, o que torna muito difícil a proposta de que TLN possa explicar os efeitos de N e NF. Sendo assim, é provável que N e NF tenham efeitos genuínos e não relacionados aos efeitos de TLN. Outra questão que coloca em dificuldades a proposta de se considerar TLN uma “melhor” medida de vizinhança ortográfica é que apenas 3,86% das palavras de quatro a seis letras no português do Brasil têm vizinhos formados pela transposição de duas letras, enquanto mais de 50% dessa mesma classe de palavras tem

vizinhos ortográficos de acordo com a métrica tradicional de Coltheart e cols. (1977), sejam eles de maior frequência de ocorrência ou não. Desse modo, no caso do português do Brasil, pode-se concluir que os efeitos de N e NF são muito provavelmente independentes dos efeitos de TLN e que afetam um grupo muito mais amplo de palavras do que TLN, o que torna N e NF medidas mais relevantes de vizinhança ortográfica, pelo menos, de um ponto de vista empírico.

Outra questão que merece avaliação é o quanto as diferenças na distribuição estatística de N e NF na língua portuguesa poderiam explicar diferenças nos efeitos dessas variáveis em relação à língua inglesa. O argumento levantado por Siakaluk e cols. (2002) é o de que, em uma língua em que a maioria das palavras tem vizinhos ortográficos de maior frequência de ocorrência, seria contra-intuitivo o desenvolvimento de um mecanismo de processamento da informação que prejudicasse a identificação da maioria das palavras da língua. A idéia de Siakaluk e cols. é que, como a maioria das palavras do inglês tem vizinhos ortográficos de maior frequência (note-se que ao se considerar as palavras de quatro a seis letras a porcentagem de palavras da língua inglesa com vizinhos de maior frequência cai para 47,85%, mas, de qualquer forma, continua bastante alta), o efeito de NF deveria ser facilitador e não inibidor. Em certo sentido, pode-se dizer que esse estudo trouxe evidências que entram em conflito com a proposta de Siakaluk e cols., pois a diferença entre a língua portuguesa e a língua inglesa foi sempre muito pequena no que diz respeito à distribuição estatística das variáveis de vizinhança ortográfica, sendo que as correlações entre essas variáveis tenderam a apresentar a mesma magnitude e padrão em ambas as línguas (ver Tabelas 6 e 7, por exemplo). Porém, pode-se tentar argumentar que, embora pequenas, existem diferenças entre a distribuição dos vizinhos ortográficos nas línguas em questão. Por exemplo, embora nas palavras de quatro letras as distribuições das variáveis de vizinhança ortográfica se sobreponham consideravelmente (diferenças com um valor de d' de até 0,2, indicam uma

sobreposição de 85%), no caso das palavras de cinco e seis letras, a distribuição das variáveis N, NPV e NVmF apresentaram uma menor sobreposição.

Esse seria um argumento até aceitável não fossem essas diferenças na direção oposta à predita por Siakaluk e cols. (2002), afinal essas diferenças indicam que as palavras do português do Brasil têm mais vizinhos ortográficos do que as palavras da língua inglesa. No entanto, a predição de Siakaluk e cols. diz respeito mais diretamente à proporção de palavras que tem vizinhos ortográficos de maior frequência de ocorrência e, no caso dessa variável, (%NF na Tabela 6) o valor d' médio foi de aproximadamente 0,2, indicando uma sobreposição muito alta das distribuições e, novamente, na direção oposta à predita. Destarte, o principal problema para o argumento de Siakaluk e cols. é que, no caso do português do Brasil, a maioria das palavras (54,11%) também tem vizinhos ortográficos de maior frequência de ocorrência e, ainda assim, o efeito de NF parece ser inibidor (Justi & Pinheiro, 2006; Justi & Pinheiro, 2008). Portanto, fica difícil sustentar que diferenças na proporção de palavras com vizinhos de maior frequência de ocorrência no português e no inglês poderiam explicar as diferenças nos efeitos de vizinhança ortográfica.

5 MÉTODO

Esse capítulo traz os procedimentos comuns a cada um dos experimentos desenvolvidos nesse estudo. Também apresenta o tipo de análise estatística utilizada e sua finalidade. Os aspectos específicos a cada experimento serão tratados em seus respectivos capítulos.

5.1 Participantes

Participaram dos quatro experimentos que compuseram essa pesquisa 32 alunos do curso de Psicologia da UFAL, sendo essa uma amostra de conveniência. A média de idade dos participantes foi de 21 anos e 2 meses (desvio padrão de 5 anos e 9 meses), sendo 24 do sexo feminino e 8 do sexo masculino. Esse estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Alagoas / UFAL (número de registro 018238/2008-06) e a participação nos experimentos foi condicionada à assinatura do Termo de Consentimento Livre e Esclarecido.

5.2 Material

Tendo como base as análises realizadas no capítulo ‘Análise de *Corpus*’, optou-se por trabalhar, no presente estudo, com a contagem de frequência de palavras do *Corpus* NILC / São Carlos (NILC, 2005), pois esta lista apresentou uma correlação positiva mais forte com o índice de familiaridade relatado por um grupo de universitários. Com base na amostra de 8465 palavras do *Corpus* NILC para as quais foram geradas estatísticas de vizinhança ortográfica, foram selecionadas para o presente estudo um conjunto de palavras que atendeu às seguintes características: ser dissílaba; ter baixa frequência de ocorrência (frequência menor ou igual a 20 ocorrências por milhão de palavras); ter uma correspondência grafema-fonema regular; e, ser um substantivo. Foram utilizadas palavras de baixa frequência de ocorrência, pois os efeitos de N e NF têm sido mais consistentemente relatados para essa classe de palavras (Andrews, 1997). No que diz respeito à regularidade do mapeamento grafema-fonema, optou-se pela utilização de palavras regulares, pois essas representam a maioria das palavras do português do Brasil. Como, de acordo com Whaley (1978), o número de letras ou o número de sílabas explicam proporções equivalentes da variância no tempo de reação, optou-se por controlar a extensão das palavras pelo número de sílabas, pois a maioria das tarefas do presente estudo envolveu respostas vocais. Por fim, foram utilizados apenas substantivos, como uma forma de controle da classe gramatical das palavras.

Definido esse conjunto de palavras, as palavras que compuseram as condições experimentais dos experimentos desenvolvidos foram selecionadas de acordo a manipulação fatorial 2 x 2 de N e NF a seguir:

- a) **Condição 1.1** (N nível 1, NF nível 1): palavras com apenas um vizinho ortográfico (N igual a um) e sem vizinhos de maior frequência de ocorrência (NF igual a zero);
- b) **Condição 1.2** (N nível 1, NF nível 2): palavras com apenas um vizinho ortográfico (N igual a um), sendo este vizinho de maior frequência de ocorrência (NF igual a um);
- c) **Condição 2.1** (N nível 2, NF nível 1): palavras com muitos vizinhos ortográficos (N maior ou igual a quatro) e sem vizinhos de maior frequência de ocorrência (NF igual a zero);
- d) **Condição 2.2** (N nível 2, NF nível 2): palavras com muitos vizinhos ortográficos (N maior ou igual a quatro), sendo um desses vizinhos de maior frequência de ocorrência (NF igual a um);

Tendo selecionado esse novo grupo de palavras, tomou-se o cuidado adicional de parear as palavras entre as condições experimentais (condições: 1.1; 1.2; 2.1; e, 2.2) de acordo com o seu fonema inicial. A única exceção foi o caso da palavra ‘ONZE’, que pertence a condição experimental 1.2. Nesse caso, o fonema inicial /õ/ foi pareado com o fonema inicial /a/ das palavras ‘ASTRO’, ‘ASAS’ e ‘ÁSIA’ que pertencem às condições experimentais 1.1, 2.1 e 2.2, respectivamente. No entanto, é importante ressaltar que o trabalho de Pinheiro e Reis (2004) não revelou diferenças estatisticamente significativas no tempo de reação entre palavras iniciadas com esses fonemas (/õ/ e /a/). Outro fator que foi controlado foi o número de palavras que continha vizinhos transpostos, esse número foi mantido constante entre as condições experimentais, sendo que apenas uma palavra por condição tinha um vizinho transposto. Dado esse extenso conjunto de controles experimentais, o conjunto final dos estímulos experimentais do presente estudo foi composto por 64 palavras, sendo 16 palavras por condição experimental. Essas palavras atenderam a uma manipulação fatorial 2 x 2 de N

e NF, sendo controladas as seguintes características: frequência de ocorrência; número de sílabas; grau de correspondência no mapeamento grafema-fonema; classe gramatical; natureza do fonema inicial; e, número de vizinhos transpostos. A Tabela 8 traz a média e o desvio padrão da frequência de ocorrência por milhão de palavras e das medidas de vizinhança ortográfica para as palavras por condição experimental (as outras variáveis foram mantidas constantes).

Tabela 8 – Média e Desvio Padrão (entre parênteses) da frequência de ocorrência e das medidas de vizinhança ortográfica dos estímulos experimentais

Cond. Exp.	Frequência	N	NF
	Média (DP)	Média (DP)	Média (DP)
1.1	10,1 (4,8)	1,0 (0,0)	0,0 (0,0)
1.2	10,1 (5,6)	1,0 (0,0)	1,0 (0,0)
2.1	10,2 (6,3)	4,9 (1,1)	0,0 (0,0)
2.2	10,1 (5,7)	5,6 (2,1)	1,0 (0,0)

Nota. Cond. Exp. = condição experimental; Frequência = frequência de ocorrência por milhão de palavras; N = número de vizinhos ortográficos; NF = número de vizinhos ortográficos de maior frequência de ocorrência.

Como pode ser observado na Tabela 8, a frequência de ocorrência das palavras nas diferentes condições experimentais é praticamente constante. As diferenças nas médias de N e NF seguem a manipulação fatorial e, no geral, não há diferença nas médias dessas variáveis dentro de um de seus níveis. A única exceção ocorre no segundo nível de N, onde as médias das condições 2.1 e 2.2 diferem um pouco no número de vizinhos, no entanto, um teste *T* indicou que essa diferença não é estatisticamente significativa ($p > 0,2$).

Tendo em mente o pareamento dos estímulos entre as condições de acordo com o fonema inicial e o fato da frequência de ocorrência das palavras ser equivalente entre as condições, tentou-se parear as palavras também de acordo com a frequência de ocorrência. No entanto, isso não foi possível, pelo menos, no nível das palavras. Porém, foi possível parear conjuntos de palavras de acordo com o fonema inicial e com a frequência de ocorrência média. Esse pareamento ocorreu porque as palavras de cada condição experimental puderam

ser divididas em quatro subconjuntos de quatro palavras. Dentro de cada condição experimental, cada um desses quatro subconjuntos recebeu a denominação de ‘lista 1’, ‘lista 2’, ‘lista 3’ e ‘lista 4’, respectivamente. A partir disso, os subconjuntos denominados como ‘lista 1’, nas quatro condições experimentais, foram pareados de acordo com o fonema inicial e com a média da frequência de ocorrência de suas palavras constituintes, o mesmo ocorrendo para os subconjuntos denominados ‘lista 2’, ‘lista 3’ e ‘lista 4’. Como todas as palavras já tinham as mesmas características, no que diz respeito ao número de sílabas, ao grau de correspondência no mapeamento grafema-fonema e a classe gramatical, pode-se dizer que os conjuntos de palavras com o mesmo nome (‘lista 1’, ‘lista 2’, ‘lista 3’ e ‘lista 4’) são equivalentes entre as condições experimentais, excetuando-se a manipulação de vizinhança ortográfica. Dessa forma, as quatro listas de palavras ao invés de serem “aninhadas” dentro das condições experimentais, puderam ser cruzadas entre as condições experimentais. Isso é interessante, pois ao se considerar o nível das listas tem-se um análogo do *design* intra-sujeitos (*within subjects*), porém, no caso desse estudo, o *design* é “intra-listas”, pois o pareamento ocorreu entre as listas e não no nível das palavras. Esse tipo de pareamento tem consequências relevantes para a análise estatística dos dados e será mais bem discutido na seção ‘Análise Estatística e *Design* Experimental’. A Tabela 9 apresenta a frequência de ocorrência das listas por condição experimental, lembrando-se que a frequência de ocorrência das listas corresponde à média da frequência de ocorrência das palavras que as compõem.

Tabela 9 – Frequência de ocorrência das listas por condição experimental

Listas	Condição Experimental			
	N1		N2	
	NF1	NF2	NF1	NF2
Lista 1	11,0	11,5	11,5	11
Lista 2	12,5	12	12,5	12,5
Lista 3	7,8	8,0	8,0	7,8
Lista 4	9,0	8,8	9,0	9,0

Nota. N1 = apenas 1 vizinho ortográfico; N2 = 4 ou mais vizinhos ortográficos; NF1 = sem vizinhos ortográficos de maior frequência de ocorrência; NF2 = um vizinho ortográfico de maior frequência de ocorrência.

Pode-se observar na Tabela 9 que a frequência de ocorrência de uma mesma lista se mantém bastante constante entre as condições experimentais. Dessa forma, considerando todos os outros controles já utilizados, pode-se dizer que os conjuntos de palavras que representam uma mesma lista nas diferentes condições experimentais são equivalentes à exceção da manipulação de N e NF. Assim sendo, essas listas de palavras configuraram-se como os estímulos experimentais em todos os quatro experimentos desenvolvidos no presente estudo (a lista com as 64 palavras que se configuraram como os estímulos experimentais encontram-se no Apêndice A). No caso dos dois primeiros experimentos, pseudopalavras foram construídas para compor as tarefas de leitura em voz alta. Para o terceiro experimento, por se constituir de uma tarefa de identificação perceptual, apenas os estímulos experimentais foram necessários. Já no último experimento, palavras que denominam cores foram incluídas para compor a tarefa de categorização semântica. Mais informações sobre as pseudopalavras e palavras que compuseram as tarefas de leitura em voz alta e de categorização semântica, estão inclusas nos capítulos que discorrem sobre cada um desses experimentos.

5.3 Procedimentos

Os participantes foram previamente divididos por sorteio em 4 grupos. Cada um desses grupos correspondia a uma ordem diferente de apresentação das listas, permitindo que cada participante fosse exposto a apenas um quarto dos estímulos (palavras alvo) por experimento. Dessa forma foi possível a participação de cada participante nos quatro experimentos, sendo que, cada um deles, foi exposto a uma ordem de listas diferente (e conseqüentemente, devido à natureza do pareamento envolvido, a um conjunto de palavras

diferentes). Mais informações sobre esse procedimento podem ser obtidas na seção ‘lidando com a falácia da linguagem como um efeito fixo’ e na Tabela 10 desse mesmo capítulo.

Todos os participantes foram testados, individualmente, em uma sala confortável e silenciosa previamente acordada com a direção do Instituto de Ciências Humanas Comunicação e Artes da Universidade Federal de Alagoas. Em cada sessão de testagem foram aplicadas duas das quatro tarefas experimentais. A ordem de execução das tarefas foi quase-aleatória porque se tomou o cuidado de evitar que a Tarefa de Leitura em Voz Alta do Tipo *Go/NoGo* (Experimento I) e a Tarefa de Leitura em Voz Alta com Palavras e Pseudopalavras (Experimento II) fossem realizadas na mesma sessão, dada a similaridade em suas demandas (o que poderia gerar confusão nos participantes). Assim sendo, todos os participantes foram testados em duas sessões diferentes, sendo que o intervalo entre uma sessão e outra foi de uma semana. Dentro de uma mesma sessão, o intervalo de tempo entre uma tarefa experimental e outra foi de aproximadamente 5 minutos e o tempo total da sessão foi de 15 a 20 minutos.

Em todas as sessões experimentais os participantes eram recepcionados por um colaborador do pesquisador. O colaborador, tal como os participantes, não tinha conhecimento das hipóteses experimentais, e assim sendo, todos os experimentos foram do tipo duplo-cego. O pesquisador não participou diretamente de nenhuma sessão de testagem, cabendo ao colaborador recepcionar os participantes, preparar o ambiente de testagem (calibrar o microfone e explicar o uso do *joystick* em tarefas nas quais esses eram necessários, digitar o nome do participante no programa, entre outros) e gravar os dados. O colaborador instruíu os participantes que o processo de testagem seria computadorizado e auto-explicativo. O instrumento utilizado para a testagem foi um computador de arquitetura compatível à IBM-PC e o *software* utilizado para a apresentação dos estímulos e coleta dos dados de tempo de reação e porcentagem de erros foi o DMDX (Forster & Forster, 2003). O *software* DMDX sincroniza a apresentação dos estímulos com o *refresh rate* do monitor o que permite uma

precisão de milissegundos na dinâmica de apresentação dos estímulos e no registro do tempo de reação dos participantes. Além disso, o *software* também foi utilizado para gerar uma ordem aleatória de apresentação dos estímulos para cada participante. Para iniciar o teste, o participante se acomodava na cadeira, ficando a uma distância aproximada de 40 centímetros da tela do computador, dependendo de sua postura. Após se acomodar, o colaborador pedia ao participante para falar algumas palavras com a finalidade de calibrar o microfone e/ou explicava o funcionamento do *joystick* (quando esse era o caso), instruindo-o então a ler as instruções experimentais na tela do computador. As instruções variavam de acordo com o tipo de tarefa realizada e, por isso, serão apresentadas nos capítulos específicos a cada experimento. Todas as tarefas contavam com uma sessão de treinamento que precedia a sessão experimental propriamente dita. O colaborador informava ao participante que ele podia redimir quaisquer dúvidas que tivesse sobre os procedimentos dos experimentos, antes ou depois do treinamento, porém antes de iniciar a sessão experimental propriamente dita. Considerando-se todos os experimentos realizados, os participantes nunca sentiram necessidade de redimir quaisquer dúvidas após o treinamento e a maioria das dúvidas que tinham antes de iniciar o treinamento referia-se a uma mera checagem de quais botões deveriam pressionar.

5.4 Análise Estatística e *Design* Experimental

No que diz respeito ao tratamento estatístico dos dados dos experimentos, pelo menos, duas considerações são necessárias: uma diz respeito ao tempo de reação (uma das variáveis

dependentes do estudo); e, a outra, ao tipo de teste estatístico que seria adequado para a análise dos dados. Essas duas questões são discutidas a seguir.

5.4.1 O Tempo de Reação como uma Variável Dependente

O Tempo de Reação (TR), além das porcentagens de erros, é uma das variáveis dependentes nos experimentos desse estudo. No que se refere ao TR, analisa-se a média do TR para cada participante por condição experimental e essa média é calculada excluindo-se o TR resultante de erros. Porém, um problema normalmente enfrentado com o tempo de reação é que a sua distribuição tende a ser positivamente assimétrica, mostrando um grande número de valores ao longo da média, mas com alguns valores maiores (provenientes de TRs mais lentos) formando uma longa cauda positiva, conseqüentemente “puxando a média para cima”.

Para lidar com esse problema, duas estratégias têm sido adotadas. Uma delas é o uso de pontos de corte fixo e a outra é o uso de medidas semi-restringidas (Perea, 1999; Perea & Rosa, 1999). O uso de pontos de corte fixos implica na exclusão de escores que ultrapassem dois valores limite estabelecidos *à priori*, por exemplo, 250 milissegundos para o limite inferior e 1500 milissegundos para o limite superior. Esse método é preferível de acordo com Perea (1999), pois não sofre nenhum viés devido a variações no tamanho amostral. No entanto, o emprego desse procedimento no presente estudo seria impraticável, pois não seria fácil precisar *à priori* quais seriam os limites (inferior e superior) mais adequados nos diversos experimentos que foram realizados. Portanto, a melhor opção, no caso do presente estudo, foi empregar o método de medidas semi-restringidas. Nesse método, todos os escores que se distanciam da média geral de um participante por, por exemplo, dois desvios padrão

para mais ou para menos, são tratados como se tivessem esse valor limite. De acordo com Perea (1999), o método de medidas semi-restringidas sofre um viés muito pequeno que tende a se tornar desprezível quando a amostra se torna maior. Sendo assim, o presente estudo, adotou o método das medidas semi-restringidas para lidar com os escores de tempo de reação dos participantes nos diferentes experimentos, limitando os escores a uma distância máxima de dois desvios padrão em relação à média.

5.4.2 A Linguagem como um Efeito Fixo: Problemas e Soluções

Recentemente, Raaijmakers (2003; e Raaijmakers & cols., 1999) trouxe a tona uma discussão muito importante a respeito do uso de testes estatísticos em experimentos onde os estímulos são lingüísticos. Essa discussão foi inicialmente proposta por Coleman (1964), mas foi com o trabalho de Clark (1973) que ela alcançou maior repercussão. Pode-se dizer que a controvérsia gerada por tais trabalhos gira em torno de duas questões: 1) ao analisarmos os dados de experimentos em que os estímulos são palavras que pertencem a uma determinada população de estímulos lingüísticos, devemos considerar o efeito dessas palavras como fixo ou como aleatório? 2) Qual o modelo estatístico adequado para analisar os dados de tais experimentos? O objetivo dessa seção é revisar essa discussão tendo como foco essas duas perguntas.

5.4.2.1 O que são um Efeito Fixo e um Efeito Aleatório?

A resposta a essa questão depende de se considerar em que medida o efeito da variável em questão é afetado ou não por conseqüências de amostragem. Por exemplo, em um experimento, o pesquisador X está interessado em saber o efeito do tempo de exposição a um estímulo sobre a probabilidade de esse estímulo ser reconhecido posteriormente. Para investigar tal questão, o pesquisador seleciona um conjunto de estímulos quaisquer e varia o tempo de exposição a eles, por exemplo, um segundo ou dois segundos. Em uma repetição futura desse experimento os dois níveis a serem comparados da variável 'tempo de exposição' poderiam ser exatamente os mesmos (um segundo versus dois segundos). Nesse sentido, se poderia dizer que a variável 'tempo de exposição' constitui um efeito fixo, pois envolve comparações entre categorias que não são afetadas pela amostragem (Lewis, 1995). Considere-se agora que o pesquisador X estivesse interessado em investigar se os substantivos em comparação aos verbos têm uma maior probabilidade de serem reconhecidos posteriormente. Para investigar tal hipótese, ele decide manter o tempo de exposição aos estímulos fixo em um segundo e apresenta aos participantes um conjunto de verbos e um conjunto de substantivos. Em uma repetição desse experimento, as palavras que compõem o conjunto de substantivos e o conjunto de verbos poderiam ser diferentes (afinal, o pesquisador quer que seus resultados sejam generalizáveis para comparações entre verbos e substantivos em geral e não, apenas, para a amostra presente em seu experimento). Nesse sentido, os dois níveis da variável de interesse (verbos e substantivos) estariam sujeitos a conseqüências de amostragem, devendo essa variável ser considerada um efeito aleatório (Lewis, 1995).

Pode-se argumentar que, em um experimento, os pesquisadores quase nunca selecionam as palavras que irão compor as condições experimentais aleatoriamente e que,

desse modo, o efeito desses estímulos poderia ser considerado fixo (Cohen, 1976). No entanto, é importante considerar que, mesmo as palavras-estímulo não sendo selecionadas de forma aleatória, é sempre possível definir uma população de palavras que atenda às características de seleção e para a qual o pesquisador esperaria que seus resultados pudessem ser generalizados. Uma forma de pensar sobre essa questão é imaginar a divisão entre a amostra de estímulos presente no estudo (p) e a população que efetivamente atende às características de seleção (P). Se o resultado da “fração amostral” p / P for próximo de 1, pode-se considerar o efeito como fixo. No entanto, se o resultado for próximo de 0, pode-se considerar o efeito como aleatório (Winer, 1971). Assim sendo, apenas quando os estímulos utilizados no experimento esgotam a população da qual eles foram selecionados, faz sentido considerar seu efeito como fixo e, mesmo assim, se o pesquisador não esperar que seus resultados possam ser generalizáveis para outras línguas. Como Clark coloca “... mesmo que o pesquisador saiba que as palavras de seu estudo não foram escolhidas aleatoriamente, ele deve tratar essa variável como um efeito aleatório, enquanto ele puder pensar em outras palavras que ele poderia ter escolhido ao invés dessas” (1973, p.349). A falta de um procedimento estritamente aleatório de seleção dos estímulos não impede que a caracterização mais adequada desses seja a de um efeito aleatório, afinal, os participantes de um experimento raramente são selecionados por um procedimento estritamente aleatório e mesmo assim, não se questiona o pressuposto de que os participantes devam ser considerados um efeito aleatório no modelo estatístico (Wickens & Keppel, 1983). De acordo com Lewis (1995) essa distinção entre um efeito fixo e um efeito aleatório não é trivial, posto que ela afeta a escolha do termo de erro para o cálculo da significância estatística, daí a segunda pergunta.

5.4.2.2 Qual o Modelo Estatístico Adequado para Analisar os Dados desses Experimentos?

Considere-se o exemplo de um pesquisador querendo testar se a classe gramatical das palavras (verbos versus substantivos) tem algum efeito sobre o tempo de reação dos participantes em uma tarefa de decisão lexical. Para tanto, ele seleciona quatro verbos e quatro substantivos e registra o tempo de reação de oito participantes para ambos os grupos de itens. Por meio de qual modelo estatístico o pesquisador deve avaliar se a diferença encontrada entre as médias é significativa? Nesse caso, o pesquisador está procurando um teste que possa determinar o quanto a diferença observada entre as médias experimentais é maior do que seria esperado com base nas outras fontes de variação que contribuem para a variabilidade entre essas médias. Uma forma de se conhecer essas outras fontes de variação é considerar quais aspectos poderiam ser diferentes em uma replicação independente do experimento (Raaijmakers, 2003).

No caso de uma replicação desse experimento, existem quatro razões pelas quais o resultado poderia ser diferente. A primeira deve-se a variabilidade inerente, isso é, mesmo quando tudo é mantido igual o resultado sempre pode ser um pouco diferente devido a algum erro aleatório. A segunda fonte de variação deve-se ao fato de que, em uma replicação, os participantes serão diferentes. Uma vez que o *design* emprega medidas repetidas (cada participante é exposto às duas condições experimentais), o fato de uns participantes serem mais rápidos do que outros não afetaria a diferença entre as médias das duas condições. No entanto, alguns participantes podem apresentar uma diferença maior entre as duas condições experimentais devido ao efeito da interação entre os sujeitos e o tratamento e isso pode afetar a diferença entre as médias. A terceira fonte de variação deve-se ao fato de que, em uma

replicação, novos itens (verbos e substantivos) podem ser utilizados. Nesse caso, pode ser que algumas palavras sejam, no geral, reconhecidas mais rapidamente (isso, no modelo estatístico, corresponderia ao efeito principal dos itens) o que pode afetar o resultado do experimento quando itens diferentes forem utilizados em uma replicação. Por fim, pode ser que alguns sujeitos reajam mais rapidamente a uns itens do que a outros devido ao efeito da interação entre os sujeitos e os itens. Essa interação constitui a quarta e última fonte de variação que pode afetar a diferença entre as médias em uma replicação desse experimento (Raaijmakers, 2003).

Pode-se perceber que nessa discussão sobre as fontes de variação assume-se que, em replicações do experimento, itens diferentes podem ser utilizados. É nesse sentido que os itens constituem um efeito aleatório e devem ser considerados uma fonte de variação no modelo estatístico. Tratar os itens como um efeito fixo significaria assumir que em qualquer replicação possível do experimento empregar-se-iam exatamente os mesmos itens (o que ocorreria se os itens esgotassem a população da qual foram selecionados). Nesse caso, no modelo estatístico, os itens não mais seriam considerados uma fonte de variação. É exatamente para esse tipo de consideração que Clark (1973) chamou atenção em seu artigo. De acordo com esse autor, muitas vezes os pesquisadores imaginam que seus resultados se aplicam a um conjunto de itens maior do que os utilizados no estudo (isso é, admitem o efeito dos itens como aleatório), mas não incluem a variação relativa ao efeito dos itens no termo de erro de seus testes estatísticos. Uma vez que o efeito principal dos itens pode contribuir para diferenças entre as médias experimentais, se essa variação também não for incluída no termo de erro do teste estatístico, o teste pode apresentar um índice de erros do tipo I maior do que o esperado (Forster & Dickinson, 1976; Wickens & Keppel, 1983). Daí a crítica de Clark (1973) de que estudos que não apresentam esse tipo de controle cometem a “falácia da linguagem como um efeito fixo”.

5.4.2.3 Lidando com a Falácia da Linguagem como um Efeito Fixo

Embora a principal forma proposta por Clark (1973) para se lidar com a variação resultante do efeito dos itens fosse estatística, essa variação também pode ser controlada de forma experimental (Raaijmakers & cols., 1999) e é exatamente essa forma de controle que o presente estudo adotou. Nesse caso, a preocupação é tentar impedir que a variação resultante do efeito dos itens possa afetar a média entre as diferentes condições experimentais. Tendo isso em mente, no presente estudo, foram utilizados diversos controles para tentar reduzir a variação entre os itens, afinal, em todas as condições experimentais foram mantidos constantes o número de sílabas das palavras (todas dissílabas), a correspondência grafema-fonema (todas regulares), a classe gramatical (apenas substantivos), o número de vizinhos transpostos (apenas um por condição experimental) e a frequência de ocorrência média (todas de baixa frequência de ocorrência, com média geral de 10,11 ocorrências por milhão de palavras). Além disso, entre as condições experimentais as palavras foram pareadas de acordo com o fonema inicial.

Pode-se argumentar que o extenso conjunto de controles utilizado, por si só, já é suficiente para reduzir sobremaneira a variação resultante do efeito dos itens. No entanto, nesse estudo, foi possível o uso de uma estratégia mais efetiva. As palavras-estímulo do presente trabalho foram subdivididas em quatro listas dentro de cada condição experimental, de forma que as palavras representando a ‘lista 1’, em uma determinada condição experimental, foram pareadas de acordo com o fonema inicial e balanceadas de acordo com a frequência de ocorrência média, com as palavras que representam a ‘lista 1’ nas demais condições experimentais, o mesmo ocorrendo para as palavras representantes das listas 2, 3 e 4. Dessa forma, pode-se considerar que no nível das listas, as mesmas foram pareadas de

acordo com a frequência de ocorrência e com o fonema inicial. Isso permite que as quatro listas de palavras, ao invés de serem aninhadas dentro das condições experimentais, possam ser cruzadas entre as condições experimentais, o que impede que o efeito principal das listas afete a diferença entre as médias. Isso pode ser observado no seguinte exemplo: imagine que a ‘lista 1’, no geral, seja reconhecida mais rapidamente do que as demais listas. Ora, como existem palavras representando a ‘lista 1’, em todas as condições experimentais, qualquer que seja a diferença de rapidez dessa lista de palavras em relação às outras, ela estará distribuída entre as condições experimentais e não afetará a diferença entre as médias. Sendo assim, da mesma forma que existem *designs* “entre-sujeitos” e “intra-sujeitos”, pode-se dizer que existem *designs* “entre-itens” e “intra-itens”. Porém, no caso desse estudo, o *design* é “intra-listas”, pois o pareamento entre os itens ocorreu no nível das listas e não no nível dos itens individuais.

Tendo subdividido os estímulos experimentais em quatro listas, além de controlar experimentalmente o efeito da variação dos itens, tornou-se também possível utilizar os mesmos participantes nos quatro experimentos desenvolvidos nesse estudo. Isso ocorreu devido ao esquema de contrabalanceamento das listas que foi adotado, em que cada participante foi exposto a apenas uma das listas por condição experimental. Assim, cada participante foi exposto a apenas um quarto dos estímulos (palavras alvo) por experimento, o que permitiu que ele participasse dos quatro experimentos, sendo que, em cada um deles, foi exposto a uma ordem de listas diferente (e conseqüentemente, devido à natureza do pareamento desenvolvido, a um conjunto de palavras diferentes). Para a análise dos dados desse *design* um tanto complexo, adotou-se a recomendação de Pollatsek e Well (1995 – anexo B, p.793) de incluir dois fatores entre sujeitos (‘D’ e ‘E’) na análise de variância para capturar a estrutura do contrabalanceamento das listas e permitir a extração da variação referente ao efeito principal das listas e da variação referente à interação entre listas e

tratamento. Dessa forma, conforme proposto por Pollatsek e Well, para todas as análises desenvolvidas, a variação referente ao efeito principal das listas foi extraída do termo de erro (afinal, devido ao esquema de contrabalanceamento utilizado essa variação não contribui para diferenças entre as condições experimentais) e a variação referente à interação entre listas e tratamento foi adicionada ao termo de erro (o que permitiu o acréscimo de dois graus de liberdade aos testes, aumentando o poder estatístico). É importante ressaltar que os fatores ‘D’ e ‘E’ não têm qualquer significado teórico, eles foram incluídos nas análises apenas para permitir que o esquema de contrabalanceamento das listas fosse levado em consideração na análise estatística, permitindo uma análise mais poderosa (com maior número de graus de liberdade). De qualquer forma, o efeito da interação entre as listas e o tratamento foi tão reduzido nesse estudo que os dois tipos de análise estatística desenvolvidos apresentaram virtualmente os mesmos resultados (ANOVA em que foram extraídas do termo de erro as variações referentes ao efeito principal das listas e a interação entre listas e tratamento; e, ANOVA em que foi extraída a variação referente ao efeito principal das listas, porém a variação referente à interação entre listas e tratamento foi adicionada ao termo de erro), ou seja, nenhum resultado que era estatisticamente significativo deixou de ser e vice-versa. Assim sendo, para facilitar a visualização dos dados, na análise dos resultados, foram considerados apenas os fatores ‘N’ e ‘NF’ e os testes estatísticos foram desenvolvidos com o auxílio do programa *SPSS for Windows* versão 16.0, sendo incorporadas nas análises estatísticas as recomendações de Pollatsek e Well (1995), pois representam a forma mais poderosa de análise dos dados. A Tabela 10 apresenta o esquema de contrabalanceamento utilizado bem como sua relação com ‘D’, ‘E’, ‘N’ e ‘NF’.

Tabela 10 - Esquema de contrabalanceamento de listas para um design onde 'N' e 'NF' são fatores com medidas repetidas e 'D' e 'E' fatores entre sujeitos

Grupos	N1		N2	
	NF1	NF2	NF1	NF2
D1E1	L1	L2	L3	L4
D1E2	L2	L1	L4	L3
D2E1	L3	L4	L1	L2
D2E2	L4	L3	L2	L1

Nota. N1 = apenas 1 vizinho ortográfico; N2 = 4 ou mais vizinhos ortográficos; NF1 = sem vizinhos ortográficos de maior frequência de ocorrência; NF2 = um vizinho ortográfico de maior frequência de ocorrência; L = lista e seu respectivo número; D e E = fatores que capturam a estrutura do esquema de contrabalanceamento das listas.

6 TAREFA DE LEITURA EM VOZ ALTA DO TIPO *GO/NOGO*

Conforme discutido no capítulo ‘Revisão da Literatura’, não parece existir uma tarefa “pura” que avalie, exclusivamente, o processo de identificação lexical (Andrews, 1997). Geralmente, cada tarefa empregada nos diversos estudos discutidos envolve aspectos relacionados ao processo de identificação lexical, mas também envolve demandas a outros processos necessários ao desempenho na tarefa (Monsell & cols., 1989). Tendo essa complexidade em mente, tornam-se importantes, por um lado, modelos de reconhecimento visual de palavras que especifiquem, claramente, os diferentes processos que podem estar envolvidos nessas tarefas (Grainger & Jacobs, 1996) e por outro, estudos empíricos que tentem evitar que componentes próprios de cada tarefa obscureçam o processo de identificação lexical que se configura, quase sempre, como o interesse principal dos pesquisadores (Justi & Pinheiro, 2006; 2008).

O objetivo do experimento descrito nesse capítulo foi o de tentar controlar componentes estratégicos que poderiam mascarar o processo de identificação lexical de forma semelhante à que fizeram Justi e Pinheiro (2006; 2008). Nos estudos de Justi e Pinheiro (2006; 2008) diversos procedimentos foram utilizados para se controlar o uso de estratégias na tarefa de decisão lexical do tipo *Go/NoGo* (ver discussão a seguir) de forma a tornar a influência do processo de identificação lexical mais saliente. O presente experimento tentou fazer o mesmo, porém utilizando-se de outra tarefa que envolve identificação lexical: a tarefa de leitura em voz alta do tipo *Go/NoGo*. Dessa forma, se os controles utilizados no presente experimento foram adequados, pode-se assumir que o processo de identificação lexical também será saliente nesse estudo e assim sendo, caso os resultados desse experimento sejam compatíveis com os achados de Justi e Pinheiro (2006; 2008) relativos à tarefa de decisão

lexical do tipo *Go/NoGo*, pode-se assumir que o presente estudo traz evidências convergentes sobre a influência de N e NF no processo de identificação lexical (Andrews & Heathcote, 2001). A seguir serão discutidos os processos que estariam envolvidos na tarefa utilizada no presente experimento e como se tentou controlá-los, tendo como base o modelo DRC (Coltheart & cols., 2001).

6.1 Processos Presentes na Tarefa de Leitura em Voz Alta do Tipo *Go/NoGo*

Para se compreender adequadamente os processos que estariam envolvidos na tarefa de leitura em voz alta do tipo *Go/NoGo*, torna-se relevante discorrer primeiro sobre a tarefa de decisão lexical. De acordo com os modelos de Leitura Múltipla (Grainger & Jacobs, 1996) e DRC (Coltheart & cols., 2001), uma resposta pode ser emitida na tarefa de decisão lexical de três maneiras: 1) por meio do nível de ativação geral do sistema lexical, o que corresponde à estratégia de “adivinhação rápida”; 2) por meio de um limite de tempo para a resposta “não”, o que corresponde à estratégia “*deadline*” usada, principalmente, para responder à pseudopalavras; e finalmente, 3) por meio do processo de identificação lexical (quando o nível de ativação da palavra alvo atinge um limiar de ativação que é constante no sistema lexical). Como então, dificultar, o uso das estratégias “*deadline*” e de “adivinhação rápida”?

Existe uma versão da tarefa de decisão lexical conhecida como tarefa de decisão lexical do tipo *Go/NoGo*. Nessa tarefa, os participantes têm que apertar uma tecla quando um conjunto de letras é uma palavra e não apertar tecla alguma quando esse conjunto de letras não representa uma palavra. Como a estratégia “*deadline*” opera apenas para as respostas

negativas na tarefa padrão, nessa versão o uso dessa estratégia fica impossibilitado, pois apenas as respostas positivas são requeridas (Perea, Rosa & Gómez, 2003).

De acordo com Grainger e Jacobs (1996), a estratégia de “adivinhação rápida” basear-se-ia no nível de ativação geral do sistema lexical para gerar a resposta “sim”. Essa estratégia só funciona porque, na maioria das vezes, o nível de ativação lexical gerado por uma palavra e seus vizinhos excede, consideravelmente, o nível de ativação lexical gerado pelos vizinhos das pseudopalavras. Dessa forma, para que o uso dessa estratégia seja dificultado, é necessário que, de alguma forma, as pseudopalavras e as palavras tenham um índice de ativação lexical equiparável (Carreiras & cols., 1997). Em seus estudos Justi e Pinheiro (2006; 2008) utilizaram como uma medida do índice de ativação lexical, no caso das palavras, a soma da frequência de ocorrência da palavra e de seus vizinhos ortográficos e, no caso das pseudopalavras, a soma da frequência de ocorrência de seus vizinhos ortográficos. Porém, como argumentaram Justi e Pinheiro (2006), esse controle não, necessariamente, garante um alto nível de ativação lexical para as pseudopalavras, porque a presença de um vizinho ortográfico de alta frequência (sem a existência de um forte competidor) pode inibir rapidamente os outros vizinhos ortográficos da pseudopalavra, o que reduziria o nível de ativação geral do léxico. Isso não ocorreria facilmente com as palavras, porque a própria palavra representaria um forte competidor para o vizinho de alta frequência. Dessa forma, dada a interatividade entre o fluxo de ativação que vai do nível da letra ao nível da palavra e, ao mesmo tempo, ao funcionamento do mecanismo de inibição lateral no nível da palavra, medidas lineares do índice de ativação lexical, como as propostas por Justi e Pinheiro (2006; 2008), podem ser um pouco imprecisas. Uma forma de se minimizar esse problema é a utilização de critérios convergentes na escolha das pseudopalavras do estudo. Por exemplo, podem ser selecionadas para o estudo pseudopalavras que tenham, pelo menos, o mesmo número de vizinhos ortográficos que as palavras (critério formal) e que, além disso, sejam

consideradas difíceis de reconhecer por gerarem maior porcentagem de erros ou tempos de reação mais lentos em um estudo piloto (critério empírico). O uso desses dois critérios pode ajudar a garantir que em uma determinada população de pseudopalavras definida pelo critério formal (aquelas com muitos vizinhos ortográficos), sejam selecionadas as mais prováveis de serem, de fato, aquelas que possuem o maior nível de ativação lexical. Em suma, utiliza-se o procedimento tradicional de buscar medidas diferentes para avaliar um mesmo construto na esperança de que a precisão de medida seja maior. Desse modo, pode-se imaginar que as pseudopalavras irão gerar tanta ativação lexical quanto à das palavras do experimento. Portanto, ao tentar utilizar a estratégia de “adivinhação rápida”, o indivíduo obterá um índice de erros grande e, provavelmente, caso a precisão das respostas tenha sido enfatizada, abdicará de utilizar essa estratégia em prol da identificação completa do estímulo-alvo por meio do processo de identificação lexical.

Tendo essa discussão da tarefa de decisão lexical do tipo *Go/NoGo* em mente, na elaboração do presente experimento objetivou-se selecionar uma tarefa de leitura em voz alta que tivesse demandas semelhantes às da tarefa de decisão lexical do tipo *Go/NoGo*. Destarte, na tarefa de leitura em voz alta empregada no presente trabalho, palavras e pseudopalavras foram apresentadas aos participantes em ordem aleatória e eles foram instruídos a ler em voz alta apenas as palavras. Assim, pode-se considerar que essa tarefa incorpora uma demanda semelhante a da tarefa de decisão lexical do tipo *Go/NoGo* (Andrews & Heathcote, 2001), evitando, desse modo, o uso da estratégia “*deadline*”. Outro fator que torna essa tarefa interessante é que, no presente experimento, dois critérios foram utilizados para a seleção das pseudopalavras: ter um número de vizinhos ortográficos maior ou igual ao das palavras alvo; e, ter sido uma das pseudopalavras mais difíceis de rejeitar em uma tarefa de decisão lexical (Justi & Justi, em preparação) ou uma das que geraram maior porcentagem de erros no estudo de Justi e Pinheiro (2006). Dessa forma, pode-se esperar que na hora de decidir qual estímulo

ler em voz alta, os participantes tendam a procurar identificar completamente a palavra alvo antes de lê-la, tornando, dessa forma, o processo de identificação lexical um aspecto saliente nessa tarefa. Assim, a tarefa de leitura em voz alta utilizada nesse estudo torna-se comparável à tarefa de decisão lexical do tipo *Go/NoGo* dos estudos de Justi e Pinheiro (2006; 2008). No entanto, é importante ressaltar que por se tratar de uma tarefa de leitura em voz alta, essa tarefa incorpora processos relacionados à produção dos fonemas envolvidos na pronúncia das palavras, processos esses que não são estritamente necessários na tarefa de decisão lexical do tipo *Go/NoGo*.

Tendo essa tarefa em mente, pode-se agora focar o processo de identificação lexical no modelo DRC (Coltheart & cols., 2001). Nesse modelo, o processo de identificação lexical acontece por meio de uma “competição” entre as unidades ortográficas que são semelhantes ao *input*. Essa competição ocorre porque a cada padrão visual encontrado no *input*, as unidades que o contém recebem ativação e, ao mesmo tempo, inibem às unidades competidoras. Esse processo se repete, em condições normais, até que a unidade lexical correspondente ao *input* “vença” suas competidoras e a palavra alvo seja identificada (Coltheart & cols., 2001). Desse modo, uma palavra com muitos vizinhos ou com vizinhos de maior frequência receberá mais inibição de seus competidores e, conseqüentemente, será reconhecida mais lentamente. No entanto, talvez o número de vizinhos de maior frequência seja a variável mais importante nesse modelo, pois pode ocorrer que uma palavra mais frequente iniba, rapidamente, suas competidoras, não sofrendo, assim, tanta inibição. Sendo assim, nesse experimento, como o processo de identificação lexical deve ser saliente, a hipótese experimental é a de que serão encontrados efeitos inibidores de N e NF, porém é esperado que o efeito de NF seja mais forte. Isso ocorreria porque, conforme demonstrado pela simulação computacional de Reynolds e Besner (2002), ao gerar os fonemas necessários

à pronúncia de um estímulo, o número de vizinhos ortográficos (N) tem um efeito facilitador, o que diminuiria o efeito inibidor previsto por um processo de identificação lexical saliente.

6.2 Método

6.2.1 Participantes

Participaram desse experimento 32 alunos do curso de Psicologia da UFAL, sendo essa uma amostra de conveniência. A média de idade dos participantes foi de 21 anos e 2 meses (desvio padrão de 5 anos e 9 meses), sendo 24 do sexo feminino e 8 do sexo masculino.

6.2.2 Material

Os estímulos experimentais consistiram em 64 palavras que atenderam a uma manipulação fatorial 2 x 2 de N e NF, sendo controladas as seguintes características: frequência de ocorrência (em média 10 ocorrências por milhão de palavras); número de sílabas (todas dissílabas); grau de correspondência no mapeamento grafema-fonema (todas regulares); classe gramatical (todas substantivos); natureza do fonema inicial (pareados entre as condições experimentais); e, número de vizinhos transpostos (apenas 1 por condição experimental). Essas 64 palavras foram subdivididas em quatro listas (pareadas pela natureza

do fonema inicial e pela frequência de ocorrência média) que foram cruzadas entre as quatro condições experimentais. Por fim, a ordem de apresentação das listas foi contrabalanceada entre as quatro condições experimentais, de forma que cada participante foi exposto a apenas uma lista por condição experimental, configurando, assim, um total de 16 palavras (mais detalhes podem ser obtidos no capítulo ‘Método’).

Para compor a tarefa de leitura em voz alta do tipo *Go/NoGo*, foram selecionadas 16 pseudopalavras dissílabas. Essas, além de terem um número de vizinhos ortográficos, em média, maior que os das palavras alvo, foram também as pseudopalavras mais difíceis de rejeitar com base nos dados de um estudo piloto (Justi & Justi, em preparação) e nos dados do trabalho de Justi e Pinheiro (2006). Isso é, essas pseudopalavras produziram TRs de rejeição mais lentos ou maior porcentagem de erros do que outras pseudopalavras em uma tarefa de decisão lexical. Sendo assim, pode-se assumir que essas pseudopalavras geram bastante ativação no sistema lexical (ou seja, são bastante similares a palavras reais) o que é fundamental, pois o objetivo desse experimento é impedir que os participantes possam utilizar a estratégia de “adivinhação rápida”. As palavras utilizadas no experimento têm uma média de 3,16 (com desvio padrão de 2,46) vizinhos ortográficos e as pseudopalavras selecionadas para o experimento têm uma média de 5,88 (com desvio padrão de 2,99) vizinhos ortográficos. Ao se considerar apenas as palavras com muitos vizinhos ortográficos, a média (o desvio padrão está entre parênteses) passa para 5,31 (1,66) vizinhos ortográficos nessas palavras. A lista de pseudopalavras utilizadas nesse experimento encontra-se no apêndice B.

Por fim, esse experimento contou também com quatro palavras e quatro pseudopalavras que foram utilizadas na sessão de treinamento. É importante considerar que os estímulos utilizados nas sessões de treinamento nunca se repetiram de um experimento para outro. Esses estímulos tinham características semelhantes às das palavras e pseudopalavras utilizadas na sessão experimental. Assim sendo, cada participante foi exposto a oito estímulos

na sessão de treinamento (4 palavras e 4 pseudopalavras) e a trinta e dois estímulos na sessão experimental (16 palavras e 16 pseudopalavras).

6.2.3 Procedimentos

Os participantes foram testados, individualmente, em uma sala cedida pela UFAL, onde foram recebidos e instruídos por um colaborador do pesquisador quanto aos procedimentos. O instrumento utilizado para o teste foi um computador de arquitetura compatível à IBM-PC e o *software* utilizado para a apresentação dos estímulos e coleta dos dados de tempo de reação e porcentagem de erros foi o DMDX (Forster & Forster, 2003). O participante se acomodava na cadeira, ficando a uma distância aproximada de 40 centímetros da tela do computador, dependendo de sua postura. Após se acomodar, o colaborador pedia ao participante para falar algumas palavras com a finalidade de calibrar o microfone e então o convidava a ler as instruções experimentais na tela do computador. As instruções foram as seguintes: “Nesse teste você verá esse sinal + no centro da tela. Logo após, você verá uma palavra real ou uma palavra inventada. Fale o nome do item apenas quando for uma palavra real, do contrário espere ele desaparecer. Tente fazer isso o mais rápido possível, mas evite errar. Aperte o botão 3 para iniciar o treinamento”. Após ler essas instruções o participante tinha a oportunidade de redimir quaisquer dúvidas antes de iniciar a sessão de treinamento ou logo após essa, porém, antes de iniciar a sessão experimental.

Os estímulos foram apresentados em uma configuração de tela de 640x480 *pixels*, na fonte ‘*fixedsys*’ de tamanho 10 e em letras capitulares, sendo a cor da fonte branca e o fundo azul. Na sessão experimental, as palavras e pseudopalavras tiveram a seguinte dinâmica de

apresentação: a apresentação de cada estímulo alvo (palavra ou pseudopalavra) iniciava-se com uma marca de fixação (+) que aparecia na tela por 500 milissegundos (ms) e depois desaparecia deixando a tela sem estímulos por 500ms; aparecia então o estímulo alvo e esse permanecia na tela por 1500ms ou até que a pessoa emitisse uma resposta vocal. O intervalo entre o fim da apresentação de um estímulo alvo e o início da apresentação de outro foi de 500ms. O participante devia ler em voz alta o estímulo alvo apenas quando esse fosse uma palavra, do contrário deveria esperar que desaparecesse. A sessão de treinamento seguiu a mesma dinâmica de apresentação da sessão experimental, porém contou com *feedback* quanto à precisão das respostas dadas. Caso o participante apresentasse uma porcentagem de erros maior do que 15% na sessão de treinamento, essa era automaticamente repetida. O tempo total de testagem, incluindo as sessões de treinamento e experimental, foi de, aproximadamente, 5 minutos por participante.

6.3 Resultados

Devido a um erro de digitação a palavra estímulo ‘ÁSIA’ ficou sem o acento agudo (‘ASIA’) e, portanto, foi excluída das análises estatísticas realizadas em todos os experimentos desse estudo. Outros dados que tiveram de ser excluídos no presente experimento deveram-se a falhas do microfone no momento de captar a pronúncia dos participantes e registrar o tempo de reação (TR), porém isso implicou na exclusão de apenas 0,59% do total de escores. Como discutido no capítulo ‘método’, optou-se nesse estudo por utilizar o método das medidas semi-restringidas para lidar com a distribuição positivamente assimétrica do TR, por conseguinte, os escores de um participante que se desviaram de sua

média, por mais de dois desvios padrão em qualquer direção, foram limitados a esse valor (Perea, 1999). Esse procedimento foi aplicado apenas ao TR para as palavras e alterou 4,3% dos valores dos escores para as palavras desse experimento, resultando em uma distribuição normal do TR (Kolmogorov-Smirnov $Z = 1,016$, $p > 0,25$).

Os participantes apresentaram uma média de porcentagem de erros total de 5,27%, com um desvio padrão de 6,27%. Já a porcentagem de erros considerando-se apenas as palavras, foi de 1,95%, com um desvio padrão de 3,70%. Ao responder às palavras, os participantes apresentaram um TR médio de 835,49ms, com um desvio padrão de 154,92ms (não foi computado o TR médio para as pseudopalavras, uma vez que os participantes não deveriam lê-las em voz alta). A Tabela 11 apresenta as estatísticas descritivas para o TR médio dos participantes por condição experimental.

Tabela 11 – Média e desvio padrão do TR na tarefa de leitura em voz alta do tipo *Go/NoGo*

	N1		N2	
	NF1	NF2	NF1	NF2
Média	833,78	832,59	810,84	867,22
Desvio Padrão	168,10	198,11	159,35	165,65

Nota. N1 = apenas 1 vizinho ortográfico; N2 = 4 ou mais vizinhos ortográficos; NF1 = sem vizinhos ortográficos de maior frequência de ocorrência; NF2 = um vizinho ortográfico de maior frequência de ocorrência.

Uma análise de variância do tipo 2 x 2 para um *design* de medidas repetidas tendo como variável dependente o tempo de reação foi desenvolvida de acordo com as recomendações de Pollatsek e Well (1995) e conforme discutido no capítulo ‘método’. Considerando-se o efeito principal de N, as palavras com quatro ou mais vizinhos ortográficos foram reconhecidas ligeiramente mais devagar (média de 839,03ms) do que as palavras com apenas um vizinho ortográfico (média de 833,19ms), no entanto, essa diferença não foi estatisticamente significativa ($F(1,30) = 0,282$; $p = 0,6$). As palavras com um vizinho ortográfico de maior frequência de ocorrência também foram reconhecidas mais lentamente (média de 849,91ms) do que as palavras sem vizinhos de maior frequência de ocorrência

(média de 822,31ms), porém, o efeito principal de NF também não foi estatisticamente significativo ($F(1,30) = 1,896; p = 0,18$). Já a interação entre N e NF foi estatisticamente significativa ($F(1,30) = 4,152; p = 0,05$) e é ilustrada na Figura 1:

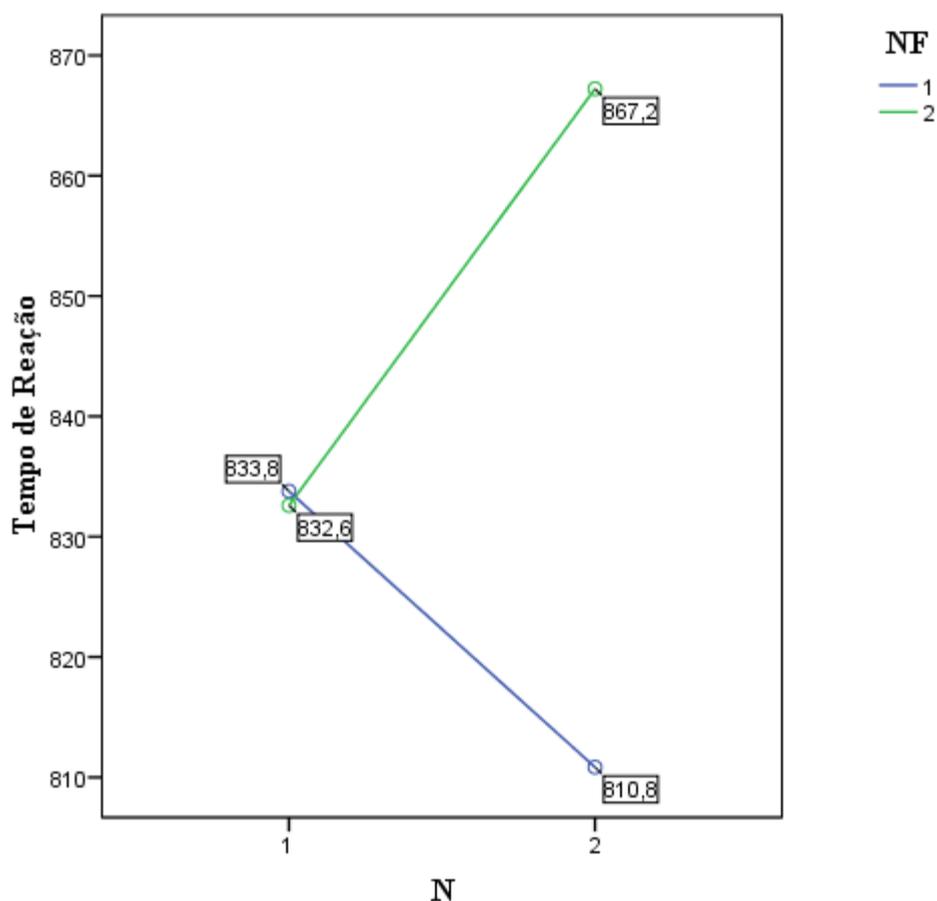


Figura 1 – Interação entre N (eixo x) e NF (linhas de cor azul e verde)

Para explorar a interação observada entre N e NF quatro testes *post hoc* foram desenvolvidos para comparar as seguintes médias da Tabela 11: NF1 x NF2 quando as palavras tinham apenas um vizinho ortográfico (N1); NF1 x NF2 quando as palavras tinham quatro ou mais vizinhos ortográficos (N2); N1 x N2 quando as palavras não tinham vizinhos ortográficos de maior frequência de ocorrência (NF1); e, N1 x N2 quando as palavras tinham um vizinho ortográfico de maior frequência de ocorrência (NF2). Como o programa *SPSS 16.0 for Windows* não calcula testes *post hoc* para *designs 2 x 2* de medidas repetidas, optou-

se por seguir as recomendações de Warner (2008) e computar testes *t* para medidas repetidas e ajustar o nível de significância de acordo com o método de Bonferroni. Como nas análises *post hoc* foram realizadas quatro comparações, de acordo com método Bonferroni, só serão consideradas estatisticamente significativas (em um nível α de 0,05) as comparações cujos valores de *p* forem menores que 0,0125. Assim sendo, os resultados das análises *post hoc* indicaram que a interação ocorreu porque quando as palavras tinham muitos vizinhos ortográficos (N2), a presença de um vizinho ortográfico de maior frequência de ocorrência aumentou a média de TR dos participantes em, aproximadamente, 56ms [NF1 x NF2: $t(31) = 2,750$; $p = 0,01$], mas quando as palavras tinham apenas um vizinho ortográfico (N1) o efeito de NF foi nulo [NF1 x NF2: $t(31) = 0,044$; $p = 0,97$]. No que diz respeito aos efeitos específicos de N, considerando-se apenas as palavras sem vizinhos ortográficos de maior frequência de ocorrência (NF1), aumentar o número de vizinhos ortográficos das palavras fez com que o tempo de reação dos participantes fosse reduzido em, aproximadamente, 23ms, porém essa diferença entre as médias não foi estatisticamente significativa [N1 x N2: $t(31) = 1,386$; $p = 0,18$]. Já aumentar o número de vizinhos ortográficos das palavras, quando todas essas tinham um vizinho ortográfico de maior frequência de ocorrência (NF2), fez com que o tempo de reação dos participantes aumentasse em, aproximadamente, 35ms, porém essa diferença entre as médias também não foi estatisticamente significativa [N1 x N2: $t(31) = 1,602$; $p = 0,12$]. Por fim, como a porcentagem de erros para a leitura das palavras foi muito pequena, os dados da porcentagem de erros por condição experimental não foram analisados.

6.4 Discussão

Embora os efeitos de N e NF tenham apresentado uma direção inibidora, os efeitos principais dessas variáveis não foram estatisticamente significantes. Isso ocorreu devido a uma interação entre N e NF, afinal, a presença de um vizinho ortográfico de maior frequência de ocorrência somente retardou o reconhecimento da palavra alvo quando essa tinha muitos vizinhos ortográficos. De uma forma geral, o modelo DRC (Coltheart & cols., 2001) é consistente com essa direção inibidora dos efeitos de N e NF, em especial, porque, como predito pelo modelo, o efeito inibidor de NF (aproximadamente, 28ms) foi numericamente maior do que o efeito inibidor de N (aproximadamente, 6ms) que, aliás, pode ser considerado nulo. No entanto, o modelo DRC parece predizer um efeito inibidor mais geral de NF e nesse sentido, é inconsistente com a interação observada entre N e NF.

Ao se comparar os resultados desse estudo com os resultados dos estudos de Justi e Pinheiro (2006; 2008) com a tarefa de decisão lexical do tipo *Go/NoGo*, pode-se dizer que eles são razoavelmente compatíveis, uma vez que os efeitos observados foram todos de natureza inibidora. No entanto, não parece haver uma explicação muito clara do porquê de N e NF terem apresentado um efeito interativo no presente experimento, principalmente, se esse padrão for comparado com o que foi observado no estudo de Justi e Pinheiro (2006). No estudo de Justi e Pinheiro, observou-se um efeito inibidor de NF quando as palavras tinham poucos vizinhos e um efeito inibidor de N quando as palavras não tinham vizinhos de maior frequência. O fato de efeitos inibidores específicos de N e NF terem sido estatisticamente significantes no estudo de Justi e Pinheiro (2006) não é muito surpreendente, posto que, diferentemente de tarefas de leitura em voz alta, em tarefas de decisão lexical a produção dos fonemas necessários à pronúncia dos estímulos não é um processo obrigatório (o que torna o

processo de identificação lexical ainda mais saliente). Porém, a interação observada entre N e NF no estudo de Justi e Pinheiro (2006) ocorreu porque, quando as palavras tinham muitos vizinhos ortográficos e um desses vizinhos tinha maior frequência de ocorrência do que a palavra alvo, o tempo de reação dos participantes não foi o mais alto. Esse padrão se opõe ao observado na interação entre N e NF nesse estudo, porque, no presente experimento, essa foi exatamente a condição experimental na qual os participantes apresentaram o maior tempo de reação. Além disso, no estudo de Justi e Pinheiro (2008), o efeito de NF foi específico às condições em que as palavras tinham poucos vizinhos ortográficos quando se levou em consideração a análise do TR, mas foi geral quando se levou em consideração a porcentagem de erros. Assim sendo, existe a possibilidade de que essa variação no efeito de NF (ora geral, ora específico) possa ser influenciada pela opção entre rapidez e precisão (*speed accuracy trade off*) dos participantes. Normalmente, quando isso ocorre, os participantes costumam apresentar uma maior porcentagem de erros nas condições experimentais em que foram mais rápidos e isso pode diminuir a precisão do TR nessas condições, já que o cálculo do TR basear-se-á em menos escores (os escores provenientes de erros são excluídos). No presente experimento, a porcentagem de erros para as palavras com apenas um vizinho ortográfico sendo esse de maior frequência de ocorrência foi, de fato, a mais alta (3,12% de erros em comparação a 1,56% de erros nas outras condições) e pode ter contribuído para aumentar o erro de medida nessa condição. Isso explicaria porque o efeito inibidor de NF não se estendeu também às condições em que as palavras tinham apenas um vizinho ortográfico no presente experimento. Se esse for o caso, o esperado é que a tarefa de identificação perceptual apresente um efeito geral de NF, já que a variável dependente nessa tarefa é a porcentagem de erros. Caso isso ocorra, talvez a interpretação mais parcimoniosa seja a de que o efeito de NF é geral, mas que devido a algum tipo de erro de medida ou pelo fato da tarefa de leitura em voz alta do tipo *Go/NoGo* ser uma tarefa menos poderosa do que outras tarefas que enfatizam

o processo de identificação lexical, ou ambos, ela falhou em detectar o efeito geral de NF. No entanto, uma resposta mais precisa para essa questão pode ser obtida por meio de uma comparação do padrão de efeitos de NF quando as palavras têm poucos e muitos vizinhos ortográficos nos diferentes estudos realizados com falantes do português brasileiro. Como isso será desenvolvido no capítulo ‘Meta-Análise e Discussão Geral’, por ora pode-se deixar em aberto se os efeitos inibidores de NF se restringem às condições em que as palavras têm muitos vizinhos ortográficos ou não.

Para concluir esse capítulo, cabe esclarecer ao leitor que devido ao fato das outras classes de modelos de reconhecimento visual de palavras discutidas no capítulo 2 não trazerem predições específicas para cada um dos experimentos desenvolvidos nesse estudo, esses não serão discutidos nos capítulos dedicados a cada experimento. Afinal, como as predições desses modelos são mais gerais, optou-se por discutir a compatibilidade desses modelos com os achados, apenas após se ter uma visão geral do padrão de resultados encontrados nos experimentos. Sendo assim, a adequação desses modelos será discutida no capítulo ‘Meta-Análise e Discussão Geral’.

7 TAREFA DE LEITURA EM VOZ ALTA

No experimento com a ‘Tarefa de Leitura em Voz Alta do Tipo *Go/NoGo*’, o objetivo principal foi tentar controlar as estratégias envolvidas na tarefa de leitura em voz alta de forma a fazer com que o processo de identificação lexical fosse um dos componentes mais salientes na tarefa. Uma das formas de avaliar se esse controle foi efetivo é comparar o resultado dessa tarefa com o de outra em que o processo de identificação lexical também seja saliente. Isso foi efetuado por meio da comparação estabelecida com a tarefa de decisão lexical do tipo *Go/NoGo* dos estudos de Justi e Pinheiro (2006; 2008) e também, poderá ser observado no experimento com a ‘Tarefa de Identificação Perceptual’. Porém, essa estratégia não resolve, por si só, o problema do processamento estratégico, porque é sempre possível argumentar que os resultados foram convergentes não porque estratégias foram controladas, mas sim, porque no português brasileiro esses são os efeitos genuínos de N e NF. Desse modo, torna-se necessário algum tipo de evidência de que os efeitos de N e NF podem, de fato, ser modulados por algum tipo de processamento estratégico. Assim sendo, o objetivo do presente experimento é enfatizar algum tipo de processamento estratégico que faça com que o processo de identificação lexical torne-se pouco influente na tarefa de leitura em voz alta. Caso esse objetivo seja alcançado, pode-se esperar que, nos resultados, o padrão dos efeitos de N e NF seja diferente do padrão observado em tarefas nas quais o processo de identificação lexical é um componente saliente (as tarefas de leitura em voz alta do tipo *Go/NoGo* e identificação perceptual do presente estudo). Assim, pode-se comparar até que ponto a estratégia em questão pode afetar realmente o desempenho na tarefa e/ou até que ponto é inócua, podendo ser considerada apenas um artefato interpretativo. Os processos que hipoteticamente estão envolvidos na tarefa utilizada no presente experimento são discutidos

logo a seguir, bem como a maneira encontrada para se enfatizar o processamento estratégico tendo como base o modelo DRC (Coltheart & cols., 2001).

7.1 Tarefa de Leitura em Voz Alta com Palavras e Pseudopalavras Misturadas

Em um trabalho influente, Coltheart (1978) argumentou que, se em uma tarefa de leitura em voz alta são misturadas palavras e pseudopalavras e se requer dos participantes respostas a ambos os estímulos, pode ser que esses tendam a ler “estrategicamente” todos os estímulos utilizando um processo de conversão grafema/fonema, sem, necessariamente, acessar a representação lexical específica da palavra alvo. A justificativa para tal estratégia seria a de que esse processo permite a pronúncia adequada de ambos os estímulos, caso as palavras sejam regulares em sua correspondência grafema/fonema. O modelo DRC (Coltheart & cols., 2001) lida com essa estratégia modulando o uso das duas rotas do modelo e o papel do mecanismo de inibição lateral presente na rota lexical. Isso é, quando muitas pseudopalavras são encontradas na tarefa de leitura em voz alta, os participantes podem diminuir o seu uso da rota lexical (porque ela não gera a pronúncia correta das pseudopalavras) e usar mais a rota fonológica (não-lexical), uma vez que essa rota sempre gera a pronúncia correta das pseudopalavras e também das palavras regulares em sua correspondência grafema/fonema. Portanto, para que a estratégia funcione, é fundamental que as palavras utilizadas no estudo tenham uma correspondência grafema/fonema regular. Nessas condições, em que a ênfase não é na identificação inequívoca de uma palavra, mas sim, na geração dos fonemas necessários à pronúncia de um estímulo-alvo (palavra ou pseudopalavra), o papel do mecanismo de inibição lateral presente na rota lexical pode ser

reduzido (Coltheart & cols., 2001). De acordo com Reynolds e Besner (2002), quando isso ocorre, o efeito do número de vizinhos ortográficos tende a ser facilitador e pode ocorrer em duas partes do modelo DRC. Em um primeiro momento, o efeito facilitador poderia ser proveniente da reverberação das ativações entre o nível da letra e o léxico de *input* ortográfico. Dessa forma, conjuntos de letras que ativam várias palavras também recebem ativação reverberativa dessas palavras, facilitando sua identificação. Em outras palavras, o estímulo alvo e seus vizinhos ortográficos tendem a facilitar o reconhecimento das letras que compartilham via ativação reverberativa. Já em um segundo momento, o efeito do número de vizinhos ortográficos seria proveniente da ativação que o léxico de *input* ortográfico e o léxico de *output* fonológico repassam para o nível de fonema, iniciando a ativação de fonemas que seriam comuns às palavras vizinhas. Nesse caso, palavras com muitos vizinhos ortográficos (léxico de *input* ortográfico), também tendem a ter muitos vizinhos fonológicos (léxico de *output* fonológico) e esses tenderão a auxiliar a ativação dos fonemas que têm em comum (nível de fonema). Como o funcionamento das duas rotas é paralelo e em cascata, a rota fonológica pode-se beneficiar da ativação reverberativa que facilita a identificação das letras e da pré-ativação de alguns fonemas ao gerar a pronúncia do estímulo-alvo, sem sofrer a influência negativa do processo de inibição lateral que ocorre na rota lexical (especialmente quando a ênfase no funcionamento desse mecanismo é reduzida). É interessante observar que nessas situações pode-se prever um efeito facilitador de N (Reynolds & Besner, 2002), no entanto, não necessariamente de NF, porque ter um vizinho ortográfico de maior frequência pode não contribuir tanto para a ativação de letras e/ou fonemas compartilhados quanto ter vários vizinhos ortográficos. Portanto, no presente experimento, palavras com uma correspondência grafema/fonema regular e pseudopalavras serão utilizadas e os participantes serão incentivados a ler o mais rápido possível todos os tipos de estímulo. Destarte, é possível que a leitura dos participantes seja guiada mais por um processo de conversão

grafema/fonema (uso da rota fonológica) do que pelo processo de identificação lexical. Sendo assim, nessa tarefa, a hipótese experimental é a de que se observará um efeito facilitador do número de vizinhos ortográficos (N).

7.2 Método

7.2.1 Participantes

Participaram desse experimento 32 alunos do curso de Psicologia da UFAL, sendo essa uma amostra de conveniência. A média de idade dos participantes foi de 21 anos e 2 meses (desvio padrão de 5 anos e 9 meses), sendo 24 do sexo feminino e 8 do sexo masculino.

7.2.2 Material

Os estímulos experimentais consistiram em 64 palavras que atenderam a uma manipulação fatorial 2 x 2 de N e NF, sendo controladas as seguintes características: frequência de ocorrência (em média 10 ocorrências por milhão de palavras); número de sílabas (todas dissílabas); grau de correspondência no mapeamento grafema-fonema (todas regulares); classe gramatical (todas substantivos); natureza do fonema inicial (pareados entre as condições experimentais); e, número de vizinhos transpostos (apenas 1 por condição

experimental). Essas 64 palavras foram subdivididas em quatro listas (pareadas pela natureza do fonema inicial e pela frequência de ocorrência média) que foram cruzadas entre as quatro condições experimentais. Por fim, a ordem de apresentação das listas foi contrabalanceada entre as quatro condições experimentais, de forma que cada participante foi exposto a apenas uma lista por condição experimental, configurando, assim, um total de 16 palavras (mais detalhes podem ser obtidos no capítulo ‘Método’).

Para compor a tarefa de leitura em voz alta foram selecionadas 16 pseudopalavras dissílabas (todas diferentes das utilizadas no experimento anterior). Essas pseudopalavras têm um número de vizinhos ortográficos, em média, maior que os das palavras alvo. A média do número de vizinhos ortográficos das pseudopalavras selecionadas para esse experimento é de 6,06 (desvio padrão de 2,94), sendo que a média do número de vizinhos ortográficos das palavras é de 3,16 (desvio padrão de 2,46). Se considerarmos apenas as palavras com muitos vizinhos ortográficos, a média passa para 5,31 (com desvio padrão de 1,66) vizinhos ortográficos nessas palavras. A lista de pseudopalavras utilizadas nesse experimento encontra-se no Apêndice B.

Por fim, esse experimento contou também com quatro palavras e quatro pseudopalavras que foram utilizadas na sessão de treinamento. Esses estímulos tinham características semelhantes às das palavras e pseudopalavras utilizadas na sessão experimental. Assim sendo, cada participante foi exposto a oito estímulos na sessão de treinamento (4 palavras e 4 pseudopalavras) e a trinta e dois estímulos na sessão experimental (16 palavras e 16 pseudopalavras).

7.2.3 Procedimentos

Os participantes foram testados individualmente em uma sala cedida pela UFAL, onde foram recebidos e instruídos por um colaborador do pesquisador quanto aos procedimentos. O instrumento utilizado para o teste foi um computador de arquitetura compatível à IBM-PC e o *software* utilizado para a apresentação dos estímulos e coleta dos dados de tempo de reação e porcentagem de erros foi o DMDX (Forster & Forster, 2003). O participante se acomodava na cadeira, ficando a uma distância aproximada de 40 centímetros da tela do computador, dependendo de sua postura. Após se acomodar, o colaborador pedia ao participante para falar algumas palavras com a finalidade de calibrar o microfone e então o convidava a ler as instruções experimentais na tela do computador. As instruções foram as seguintes: “Nesse teste você verá esse sinal + no centro da tela. Logo após, você verá uma palavra real ou uma palavra inventada. Você deve ler em voz alta ambos os tipos de estímulos. Tente fazer isso o mais rápido possível, mas evite errar. Aperte o botão 3 para iniciar o treinamento”. Após ler essas instruções o participante tinha a oportunidade de redimir quaisquer dúvidas antes de iniciar a sessão de treinamento ou logo após essa, porém, antes de iniciar a sessão experimental.

Os estímulos foram apresentados em uma configuração de tela de 640x480 *pixels*, na fonte ‘*fixedsys*’ de tamanho 10 e em letras capitulares, sendo a cor da fonte branca e o fundo azul. Na sessão experimental, as palavras e pseudopalavras tiveram a seguinte dinâmica de apresentação: a apresentação de cada estímulo alvo (palavra ou pseudopalavra) iniciava-se com uma marca de fixação (+) que aparecia na tela por 500 milissegundos (ms) e depois desaparecia deixando a tela sem estímulos por 500ms; aparecia então o estímulo alvo e esse permanecia na tela por 1500ms ou até que a pessoa emitisse uma resposta vocal. O intervalo

entre o fim da apresentação de um estímulo alvo e o início da apresentação de outro foi de 500ms. O participante devia ler em voz alta, o mais rápido possível, qualquer tipo de estímulo fossem essas palavras ou pseudopalavras. A sessão de treinamento seguiu a mesma dinâmica de apresentação da sessão experimental, porém contou com *feedback* quanto à precisão das respostas dadas. Caso o participante apresentasse uma porcentagem de erros maior do que 15% na sessão de treinamento, essa era automaticamente repetida. O tempo total de testagem, incluindo as sessões de treinamento e experimental, foi de, aproximadamente, 5 minutos por participante.

7.3 Resultados

Devido a um erro de digitação a palavra estímulo ‘ÁSIA’ ficou sem o acento agudo (‘ASIA’) e, portanto, foi excluída das análises estatísticas realizadas. Outros dados que tiveram de ser excluídos deveram-se a falhas do microfone no momento de captar a pronúncia dos participantes e registrar o TR, porém isso implicou na exclusão de apenas 1,46% do total de escores (0,59% dos escores referentes às palavras e 2,34% dos escores referentes às pseudopalavras). Como discutido no capítulo ‘método’, optou-se nesse estudo por utilizar o método das medidas semi-restringidas para lidar com a distribuição positivamente assimétrica do TR, por conseguinte, os escores de um participante que se desviaram de sua média, por mais de dois desvios padrão em qualquer direção, foram limitados a esse valor (Perea, 1999). Esse procedimento foi aplicado ao TR para as palavras e pseudopalavras e alterou 4,2% do total de escores desse experimento, resultando em uma distribuição normal do TR (Kolmogorov-Smirnov $Z = 0,494$; $p > 0,96$).

Os participantes apresentaram uma média de porcentagem de erros total de 0,98%, com um desvio padrão de 2,01%. A porcentagem de erros na leitura das palavras foi de 0,39% com um desvio padrão de 1,54% e a porcentagem de erros na leitura das pseudopalavras foi de 1,56% com um desvio padrão de 3,55%. O TR médio foi de 773,27ms, com um desvio padrão de 120,20ms. Considerando-se apenas as respostas às palavras, os participantes apresentaram um TR médio de 749,88ms, com um desvio padrão de 112,43ms. Já no caso da resposta às pseudopalavras o TR médio foi de 796,66ms, com um desvio padrão de 129,56ms. A Tabela 12 apresenta as estatísticas descritivas para o TR médio dos participantes por condição experimental.

Tabela 12 – Média e Desvio Padrão do TR na tarefa de leitura em voz alta

	N1		N2	
	NF1	NF2	NF1	NF2
Média	759,59	756,44	747,03	737,19
Desvio Padrão	125,71	119,94	118,61	110,64

Nota. N1 = apenas 1 vizinho ortográfico; N2 = 4 ou mais vizinhos ortográficos; NF1 = sem vizinhos ortográficos de maior frequência de ocorrência; NF2 = um vizinho ortográfico de maior frequência de ocorrência.

Uma análise de variância do tipo 2 x 2 para um *design* de medidas repetidas tendo como variável dependente o tempo de reação foi desenvolvida de acordo com as recomendações de Pollatsek e Well (1995) e conforme discutido no capítulo ‘método’. Considerando-se o efeito principal de N, as palavras com mais vizinhos ortográficos foram reconhecidas 15,91ms mais rapidamente do que as palavras com apenas um vizinho ortográfico, no entanto, essa diferença não foi estatisticamente significativa [$F(1,30) = 2,711; p = 0,11$]. As palavras com um vizinho de maior frequência de ocorrência também foram reconhecidas mais rapidamente do que as palavras sem vizinhos de maior frequência de ocorrência, porém, a vantagem no TR foi de apenas 6,5ms, sendo que o efeito principal de NF também não foi estatisticamente significativa [$F(1,30) = 1,085; p = 0,31$]. Por fim, a interação entre N e NF também não foi estatisticamente significativa [$F(1,30) = 0,18; p = 0,67$]. A

análise da porcentagem de erros não foi realizada porque o número de erros cometidos foi muito pequeno (apenas 0,39%).

7.4 Discussão

Conforme predito pelo modelo DRC (Coltheart & cols., 2001) as palavras que tinham muitos vizinhos ortográficos tenderam a ser reconhecidas mais rapidamente do que as palavras com apenas um vizinho ortográfico (cerca de 16ms mais rápido). É interessante notar que as análises estatísticas empregadas nesse estudo foram bicaudais, porque como existem predições opostas de diferentes modelos, esse pareceu ser o procedimento mais justificável. No entanto, não é menos importante ressaltar que se fossem consideradas apenas as predições do modelo DRC, análises unicaudais seriam plenamente justificadas. Afinal, o modelo faz predições direcionais bastante claras e assim sendo, em uma análise unicaudal, o efeito de N poderia ser considerado estatisticamente significativo ($p = 0,05$), enquanto o efeito de NF ou a interação entre N e NF continuariam longe de serem estatisticamente significantes, mesmo em uma análise unicaudal (todos os valores $p > 0,15$). Comparando-se os resultados desse experimento com os resultados do experimento realizado no capítulo anterior ('Tarefa de Leitura em Voz Alta do Tipo *Go/NoGo*'), pode-se perceber que o padrão dos efeitos de N e NF é invertido. Se na tarefa de leitura em voz alta do tipo *Go/NoGo* pode-se dizer que N e NF apresentaram efeitos numericamente inibidores (cerca de 6ms para N e cerca de 28ms para NF), no presente experimento, esses efeitos foram numericamente facilitadores (cerca de 16ms para N e cerca de 6ms para NF). Porém, pode-se argumentar que essa inversão no padrão dos efeitos de N e NF não é real, uma vez que, como os efeitos principais dessas

variáveis não foram estatisticamente significantes em nenhum dos experimentos, pode-se considerar que ambos os experimentos produziram efeitos nulos. Esse argumento tem o mérito de ser rigoroso do ponto de vista estatístico, porém peca por simplicidade. Afinal, mesmo que consideremos apenas os resultados que foram estatisticamente significantes, é necessário admitir que na tarefa de leitura em voz alta do tipo *Go/NoGo* o efeito de NF foi estatisticamente significante quando as palavras tinham muitos vizinhos ortográficos e, no presente experimento, se as análises tivessem sido unicaudais, o efeito de N poderia ser considerado estatisticamente significante. Em especial, no que diz respeito ao presente experimento, pode-se argumentar que esse falhou em detectar o efeito facilitador de N em uma análise bicaudal por falta de poder estatístico (daí esse efeito ter sido detectado em uma análise unicaudal, já que essa análise é mais poderosa). Se esse for, de fato, o caso, pode-se esperar que na meta-análise desenvolvida no capítulo ‘Meta-Análise e Discussão Geral’ detecte-se um efeito facilitador de N em tarefas que envolvam algum tipo de processamento estratégico, já que uma das características de uma meta-análise é a de reduzir a chance de erros do tipo II. Outro dado interessante é que, em nenhuma outra análise dos efeitos gerais de N e NF realizadas nos outros experimentos dessa tese, há qualquer mudança nos resultados ao se passar de uma análise bicaudal para uma unicaudal (todos os resultados estatisticamente não significantes continuam não significantes). Sendo assim, parece razoável argumentar que os efeitos de N e NF modificaram-se de um experimento para o outro e que a direção dessa mudança está de acordo com a predita pelo modelo DRC (Coltheart & cols., 2001).

Em suma, o presente experimento traz evidências de que os efeitos de N e NF podem, de fato, ser modulados por algum tipo de processamento estratégico e que, nas condições em que o mecanismo de inibição lateral é menos enfatizado, o efeito de N torna-se facilitador e o efeito de NF passa a ser nulo. No entanto, como o efeito de N foi nulo em uma análise bicaudal, é melhor deixar essa conclusão em suspenso até que as análises desenvolvidas no

capítulo ‘Meta-Análise e Discussão Geral’ possam esclarecer melhor se isso se deveu à falta de poder estatístico ou a um efeito realmente nulo.

8 TAREFA DE IDENTIFICAÇÃO PERCEPTUAL

Um dos objetivos desse estudo é investigar os efeitos de N e NF em diferentes tarefas de forma que seja possível observar esses efeitos em tarefas em que o processo de identificação lexical seja saliente e também em tarefas em que esse processo seja menos enfatizado devido a algum tipo de processamento estratégico. O experimento descrito no capítulo anterior procurou enfatizar o uso de uma estratégia de conversão grafema/fonema e permitiu contrastar os efeitos observados de N e NF com os de um experimento que tentou controlar o uso de estratégias (tarefa de leitura em voz alta do tipo *Go/NoGo*). Para que um dos objetivos desse estudo seja completamente alcançado, também é necessário que se comparem os efeitos de N e NF em duas tarefas em que o processo de identificação lexical é um componente saliente. Para tanto, o presente experimento investigou o efeito de N e NF em uma tarefa de identificação perceptual, já que, normalmente, o processo de identificação lexical é importante para o bom desempenho nessa tarefa. Isso permite a comparação dos resultados desse experimento com os da tarefa de leitura em voz alta do tipo *Go/NoGo* em que o processo de identificação lexical também pode ser considerado saliente e sendo assim, pode-se esperar que os efeitos de N e NF apresentem um padrão semelhante nessas tarefas. Os processos envolvidos na tarefa de identificação perceptual utilizada no presente experimento são discutidos logo a seguir, bem como a maneira encontrada para controlar o possível uso de estratégias nessa tarefa.

8.1 Processos Presentes na Tarefa de Identificação Perceptual

De acordo com Grainger e Jacobs (1996), a tarefa de identificação perceptual distingue-se da tarefa de decisão lexical, porque, na primeira, não estão operativas as estratégias *deadline* e de adivinhação rápida que costumam afetar o desempenho dos participantes na tarefa de decisão lexical. Isso ocorre porque na tarefa de identificação perceptual não existem respostas negativas e sendo assim, não é necessário se estabelecer qualquer *deadline* para essas respostas. A estratégia de adivinhação rápida também não é efetiva, porque os participantes não devem emitir uma resposta positiva (do tipo “sim” quando o estímulo é uma palavra), mas sim, identificar, efetivamente, qual foi o estímulo apresentado. Destarte, na análise de Grainger e Jacobs (1996) a principal forma que um participante tem para responder na tarefa de identificação perceptual é por meio do processo de identificação lexical. Como o modelo DRC (Coltheart & cols., 2001) incorpora os mesmos mecanismos do modelo de Grainger e Jacobs (1996), pode-se considerar que o modelo DRC também preveja que na tarefa de identificação perceptual o processo de identificação lexical seja bastante saliente.

Embora possa fazer sentido considerar que o processo de identificação lexical é um dos mais salientes na tarefa de identificação perceptual, nos estudos discutidos no capítulo ‘Revisão da Literatura’, levantou-se a hipótese de que essa tarefa também pode ser suscetível a estratégias. Isso poderia ocorrer porque na maioria dos estudos discutidos (Grainger & Jacobs, 1996; Grainger & Segui, 1990; Sears & cols., 1999; e em quase todos os experimentos de Snodgrass & Mintzer, 1993) não havia um tempo limite para a resposta dos participantes e assim, eles tinham bastante tempo para poder tentar “adivinhar” qual era a palavra mais compatível com o estímulo que viram. Além disso, como nesses estudos a

resposta consistia em digitar qual era a palavra que tinha sido apresentada, existe a possibilidade de que processos relacionados à grafia das palavras possam afetar de alguma forma a resposta dos participantes (lembre-se que o interesse principal dos pesquisadores é a leitura). Tendo essas considerações em mente, o presente experimento fez uso de uma tarefa de identificação perceptual tomando os seguintes cuidados: estabeleceu um limite de tempo (3 segundos) para a resposta dos participantes; e, a resposta requisitada consistia em dizer em voz alta qual era a palavra que tinha sido apresentada. Dessa forma, pode-se esperar que qualquer tentativa sofisticada de adivinhação dos estímulos será reduzida, porque os participantes terão um tempo limite para responder. Além disso, ao requisitar respostas vocais, esse experimento não só evita que processos relacionados à grafia das palavras possam afetar os resultados, como também, garante maior possibilidade de comparação com os experimentos dos capítulos anteriores, uma vez que todos empregaram respostas vocais. Cabe ressaltar que se optou por utilizar uma tarefa de identificação perceptual tradicional (*standard perceptual identification task*) nesse experimento porque apenas dois estudos utilizaram essa tarefa e apresentaram resultados relativamente conflitantes (Grainger & Jacobs, 1996; Sears & cols., 1999), sendo assim, parece importante investigar se os efeitos de N e NF podem ser generalizáveis para essa tarefa também. Por fim, é importante ressaltar que em versões padrão (como a adotada nesse estudo) da tarefa de identificação perceptual a porcentagem de erros constitui-se muitas vezes como a única variável dependente. Isso ocorre porque, como o objetivo é que a tarefa seja bastante difícil, procura-se fazer com que os participantes tenham porcentagens de erros próximas de 50% (Grainger & Jacobs, 1996; Sears & cols., 1999) e como é de praxe a exclusão do tempo de reação de respostas que resultaram em erros, muitos escores são excluídos, impossibilitando a análise do tempo de reação por condição experimental. O fato da variável dependente, nesse experimento, ser a porcentagem de erros, tem vantagens e desvantagens, porque, se por um lado, pode dificultar comparações

com os outros experimentos (uma vez que os efeitos terão métricas diferentes), por outro, pode ajudar a estabelecer a generalidade dos efeitos de N e NF em uma variável dependente diferente, o que é sempre bom (Kantowitz, Roediger III & Elmes, 2006).

Tendo-se essa tarefa de identificação perceptual em mente, pode-se agora passar para as previsões do modelo DRC (Coltheart & cols., 2001) relativas a essa. Na prática, como o processo de identificação lexical deve ser saliente nessa tarefa, pode-se esperar que N e NF devam apresentar efeitos semelhantes aos apresentados na tarefa de leitura em voz alta do tipo *Go/NoGo*. Isso é, pode-se esperar um efeito inibidor de NF e um efeito nulo de N.

8.2 Método

8.2.1 Participantes

Participaram desse experimento 32 alunos do curso de Psicologia da UFAL, sendo essa uma amostra de conveniência. A média de idade dos participantes foi de 21 anos e 2 meses (desvio padrão de 5 anos e 9 meses), sendo 24 do sexo feminino e 8 do sexo masculino.

8.2.2 Material

Os estímulos experimentais consistiram de 64 palavras que atenderam a uma manipulação fatorial 2 x 2 de N e NF, sendo controladas as seguintes características: frequência de ocorrência (em média 10 ocorrências por milhão de palavras); número de sílabas (todas dissílabas); grau de correspondência no mapeamento grafema-fonema (todas regulares); classe gramatical (todas substantivos); natureza do fonema inicial (pareados entre as condições experimentais); e, número de vizinhos transpostos (apenas 1 por condição experimental). Essas 64 palavras foram subdivididas em quatro listas (pareadas pela natureza do fonema inicial e pela frequência de ocorrência média) que foram cruzadas entre as quatro condições experimentais. Por fim, a ordem de apresentação das listas foi contrabalanceada entre as quatro condições experimentais, de forma que cada participante foi exposto a apenas uma lista por condição experimental, configurando, assim, um total de 16 palavras (mais detalhes podem ser obtidos no capítulo ‘Método’). Para essa tarefa não foram necessários outros estímulos. No entanto, esse experimento contou também com 12 palavras que foram utilizadas na sessão de treinamento. Esses estímulos tinham características semelhantes às das palavras utilizadas na sessão experimental. Assim sendo, cada participante foi exposto a doze palavras na sessão de treinamento e a dezesseis palavras na sessão experimental.

8.2.3 Procedimentos

Os participantes foram testados individualmente em uma sala cedida pela UFAL, onde foram recebidos e instruídos por um colaborador do pesquisador quanto aos procedimentos. O instrumento utilizado para o teste foi um computador de arquitetura compatível à IBM-PC e o *software* utilizado para a apresentação dos estímulos e coleta dos dados de tempo de reação e porcentagem de erros foi o DMDX (Forster & Forster, 2003). O participante se acomodava na cadeira, ficando a uma distância aproximada de 40 centímetros da tela do computador, dependendo de sua postura. Após se acomodar, o colaborador pedia ao participante para falar algumas palavras com a finalidade de calibrar o microfone e então o instruía a ler as instruções experimentais na tela do computador. As instruções foram as seguintes: “Nesse teste, é necessário que você preste muita atenção. Esse símbolo + indicará onde você deve olhar. Logo depois, esse símbolo: ##### aparecerá brevemente na tela. Em seguida, uma palavra aparecerá muito rapidamente, sendo coberta pelo símbolo ##### e depois por ??????. Você deve tentar dizer qual era a palavra o mais rápido que puder. Aperte o ENTER para iniciar o treinamento”. Após ler essas instruções o participante tinha a oportunidade de redimir quaisquer dúvidas antes de iniciar a sessão de treinamento ou logo após essa, porém, antes de iniciar a sessão experimental.

Os estímulos foram apresentados em uma configuração de tela de 640x480 *pixels*, na fonte ‘*fixedsys*’ de tamanho 10 e em letras capitulares, sendo a cor da fonte branca e o fundo azul. A exposição dos estímulos teve a seguinte dinâmica de apresentação na sessão experimental: uma marca de fixação (+) aparecia na tela por 500ms e depois desaparecia deixando a tela sem estímulos por 500ms; aparecia então uma máscara (#####) que durava 500ms, sendo apagada e imediatamente seguida pela apresentação da palavra alvo que durava

60ms; então a palavra alvo desaparecia sendo imediatamente substituída por uma máscara (#####) que durava 500ms; por fim, a máscara era substituída por sinais de interrogação (?????) que duravam até o tempo limite para uma resposta (3 segundos contados a partir da apresentação da palavra alvo). O participante devia dizer em voz alta qual era a palavra apresentada em até 3 segundos. A sessão de treinamento seguiu a mesma dinâmica, porém contou com *feedback* quanto à resposta dada. Além disso, o tempo de apresentação da palavra alvo na sessão de treinamento iniciou em 100ms, decrescendo em 20ms a cada duas tentativas até atingir o piso de 60ms nas 8 tentativas finais do treinamento. A finalidade desse procedimento foi facilitar a familiarização do participante com o tempo de apresentação das palavras alvo. O tempo total de testagem, incluindo as sessões de treinamento e experimental, foi de, aproximadamente, 4 minutos por participante.

8.3 Resultados

Devido a um erro de digitação a palavra estímulo ‘ÁSIA’ ficou sem o acento agudo (‘ASIA’) e, portanto, foi excluída das análises estatísticas realizadas. Nesse experimento, a variável dependente é a porcentagem de erros cometidos e, conseqüentemente, não foi efetuado o cômputo do tempo de reação por condição experimental. No entanto, foi possível o cômputo da média geral do tempo de reação dos participantes que foi de 1007,16ms, com um desvio padrão de 222,38ms (utilizou-se o método das medidas semi-restringidas para lidar com a distribuição positivamente assimétrica do tempo de reação, o que resultou na alteração de 4% dos escores e em uma distribuição normal do TR: Kolmogorov-Smirnov $Z = 0,67$, $p > 0,75$). No que diz respeito à porcentagem de erros, os participantes apresentaram uma média

geral de porcentagem de erros de 32,47%, com um desvio padrão de 26,76%, sendo essa uma distribuição normal (Kolmogorov-Smirnov $Z = 0,838$, $p > 0,48$). A Tabela 13 apresenta as estatísticas descritivas da porcentagem de erros por condição experimental para uma manipulação fatorial 2 x 2 de N e NF.

Tabela 13 – Média e Desvio Padrão da porcentagem de erros na tarefa de identificação perceptual

	N1		N2	
	NF1	NF2	NF1	NF2
Média	31,25	39,06	21,88	37,50
Desvio Padrão	35,36	34,16	28,93	35,04

Nota. N1 = apenas 1 vizinho ortográfico; N2 = 4 ou mais vizinhos ortográficos; NF1 = sem vizinhos ortográficos de maior frequência de ocorrência; NF2 = um vizinho ortográfico de maior frequência de ocorrência.

Uma análise de variância do tipo 2 x 2 para um *design* de medidas repetidas tendo como variável dependente a porcentagem de erros foi desenvolvida de acordo com as recomendações de Pollatsek e Well (1995) e conforme discutido no capítulo ‘método’. Essa análise indicou que o efeito principal de N foi facilitador e estatisticamente significativo [$F(1,30) = 4,31$; $p = 0,047$]. Isto é, as palavras com muitos vizinhos ortográficos foram reconhecidas de forma mais precisa (5,5%) do que palavras com poucos vizinhos ortográficos (porcentagem de erros média de 29,69% e 35,16%, respectivamente). Além disso, também se observou que o efeito principal de NF foi inibidor e estatisticamente significativo [$F(1,30) = 11,11$; $p < 0,01$], ou seja, os participantes apresentaram, aproximadamente, 12% a mais de erros nas palavras com vizinhos de maior frequência de ocorrência em relação às palavras sem vizinhos de maior frequência de ocorrência (porcentagem de erros média de 38,28% e 26,56%, respectivamente). A análise de variância não indicou uma interação estatisticamente significativa entre os fatores N e NF ($p > 0,35$).

8.4 Discussão

Nesse experimento, NF apresentou um efeito inibidor mais geral do que o observado na tarefa de leitura em voz alta do tipo *Go/NoGo*. Esse efeito mais geral de NF no presente experimento se coaduna com a hipótese levantada de que, devido a algum tipo de erro de medida e/ou por ser uma tarefa menos poderosa, a tarefa de leitura em voz alta do tipo *Go/NoGo* falhou em detectar o efeito geral de NF. Um dos argumentos levantados foi o de que como as palavras com um vizinho ortográfico, sendo esse vizinho de maior frequência de ocorrência, apresentaram uma maior porcentagem de erros, isso pode ter afetado a precisão de medida do tempo de reação nessa condição, prejudicando a detecção de um efeito mais geral de NF. Observando-se a porcentagem de erros no presente experimento, pode-se observar que esse apresentou o mesmo padrão. A condição em que as palavras tinham apenas um vizinho ortográfico, sendo esse de maior frequência de ocorrência, foi a que apresentou a maior porcentagem de erros: 39,06%. Porém, como a variável dependente, nesse experimento, foi a porcentagem de erros, esse foi capaz de detectar o efeito mais geral de NF. No entanto, é importante ressaltar também que, embora o efeito inibidor de NF tenha sido geral, ele foi mais acentuado quando as palavras tinham muitos vizinhos ortográficos (produzindo, aproximadamente, 16% a mais de erros) do que quando essas tinham apenas um vizinho ortográfico (produzindo, aproximadamente, 8% a mais de erros). Assim sendo, talvez não seja justo excluir de todo a possibilidade de interação entre N e NF e reservar a resposta a essa questão para o capítulo ‘Meta-Análise e Discussão Geral’, onde se pode comparar o padrão de efeitos de NF quando as palavras têm poucos e muitos vizinhos ortográficos nos diferentes estudos realizados com falantes do português brasileiro.

A principal diferença entre a tarefa de identificação perceptual e a tarefa de leitura em voz alta do tipo *Go/NoGo* encontra-se no efeito de N que foi nulo no experimento do capítulo 6 e facilitador no presente experimento. Porém, uma observação cuidadosa do efeito facilitador de N na tarefa de identificação perceptual indica que esse efeito foi muito mais forte nas condições em que as palavras não tinham vizinhos ortográficos mais frequentes (produzindo, aproximadamente, 9% a menos de erros) do que nas condições em que as palavras tinham vizinhos ortográficos mais frequentes (produzindo, aproximadamente, apenas 2% a menos de erros). Ora, se for considerado o efeito de N, apenas nas condições em que as palavras não tinham vizinhos ortográficos mais frequentes, pode-se perceber que, mesmo na tarefa de leitura em voz alta do tipo *Go/NoGo*, esse efeito apresentou uma tendência facilitadora de, aproximadamente, 23ms. Assim sendo, pode ser que a tarefa de identificação perceptual, por ser uma tarefa mais difícil, talvez tenha sido mais poderosa em detectar os efeitos de N e NF do que os outros experimentos desse estudo.

Tendo essas considerações em mente, pode-se dizer que os resultados desse experimento e os resultados da tarefa de leitura em voz alta do tipo *Go/NoGo* são, pelo menos, razoavelmente compatíveis. Destarte, o presente experimento traz evidências convergentes de que, em condições nas quais o processo de identificação lexical é saliente, o efeito de NF é inibidor. O efeito de N é mais difícil de precisar porque variou de nulo a facilitador nos experimentos dessa tese e foi inibidor nos experimentos de Justi e Pinheiro (2006; 2008). Uma diferença entre as tarefas dessa tese e as dos estudos de Justi e Pinheiro é que, nessa tese, as respostas exigiam uma resposta vocal e conseqüentemente, processos envolvidos na produção dos fonemas necessários à pronúncia de um estímulo, necessariamente, estão envolvidos nas tarefas de identificação perceptual e leitura em voz alta do tipo *Go/NoGo* que foram utilizadas.

No que diz respeito à compatibilidade entre os resultados dessa tarefa de identificação perceptual e as previsões do modelo DRC (Coltheart & cols., 2001), pode-se dizer que o efeito inibidor geral de NF está de acordo com o modelo DRC, ficando em aberto apenas se esse é mais acentuado ou não nas condições em que as palavras têm muitos vizinhos ortográficos. Já o efeito facilitador de N é mais difícil de ser explicado pelo modelo, porque o esperado seria um efeito nulo ou, até mesmo, um efeito inibidor fraco. No entanto, talvez esse efeito não seja tão difícil assim de ser acomodado pelo modelo, uma vez que a tarefa envolve leitura em voz alta e dessa forma, a reverberação da ativação proveniente do léxico ortográfico ao nível das letras e a ativação repassada pelo léxico de *output* fonológico ao nível de fonema poderiam, em conjunto, reverter o efeito inibidor do mecanismo de inibição lateral que, no caso de palavras sem vizinhos ortográficos mais frequentes, já se esperava ser fraco ou nulo. Dessa forma, a explicação para um efeito facilitador de N dependeria da força das conexões entre o léxico ortográfico e o nível da letra e entre o léxico de *output* fonológico e o nível de fonema. Se a força dessas conexões for razoável, elas poderiam reverter o efeito do mecanismo de inibição lateral. Isso não afetaria o efeito de NF, porque um único vizinho ortográfico, mesmo que mais frequente, enviaria pouca ativação para outros níveis do modelo, por se tratar de uma única fonte de informação (já palavras com muitos vizinhos ortográficos teriam várias fontes de informação ativando unidades em comum em diferentes níveis do modelo). Se essa hipótese estiver correta, é de se esperar que o efeito de N, mas não o de NF, seja modulado também pelo tipo de resposta envolvida na tarefa, porque tarefas que impliquem em uma demanda mais forte a processos fonológicos (como gerar os fonemas necessários à pronúncia de um estímulo) se beneficiariam mais da ativação repassada pelo léxico de *output* fonológico ao nível de fonema e esse processo é mais sensível a N do que a NF. Essa hipótese será avaliada no capítulo ‘Meta-Análise e Discussão Geral’.

9 TAREFA DE CATEGORIZAÇÃO SEMÂNTICA

No capítulo ‘Modelos de Reconhecimento Visual de Palavras’, pontuou-se que a maioria dos modelos de reconhecimento visual de palavras ainda não tem um sistema semântico (Modelos de Busca e Verificação - Forster, 1976; 1992; Murray & Forster, 2004; Paap & cols., 1982; Paap & Johansen, 1994; Modelos de Ativação Interativa e Competição - Coltheart e cols., 2001; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Modelos de Processamento Paralelo Distribuído - Plaut & cols., 1996; Seidenberg & McClelland, 1989). A exceção fica por conta do modelo de Harm e Seidenberg (2004). No entanto, no que diz respeito aos efeitos de N e NF, os autores desse modelo não desenvolveram nenhuma predição específica. Sendo assim, pode-se questionar a finalidade de se desenvolverem estudos com a tarefa de categorização semântica visando investigar os efeitos de N e NF. Existem, pelo menos, dois argumentos contra essa asserção: primeiro, durante a leitura, o acesso à semântica é o interesse principal; segundo, mesmo que não se possa estabelecer, inequivocamente, quais seriam as predições dos modelos de reconhecimento de palavras quanto à tarefa de categorização semântica, os achados de estudos com essa tarefa podem servir para impor restrições ao desenvolvimento futuro desses modelos. Destarte, tendo essas considerações em mente, procurou-se avaliar se N e NF teriam algum efeito em uma tarefa de categorização semântica realizada com falantes do português do Brasil.

9.1 Processos Presentes na Tarefa de Categorização Semântica

Considerando-se os estudos que investigaram efeitos de vizinhança ortográfica, existem, ao menos, duas explicações para como uma resposta pode ser levada a cabo na tarefa de categorização semântica. Uma dessas explicações foi proposta por Carreiras e cols. (1997) e fundamenta-se no uso de dois critérios para guiar as respostas dos participantes na tarefa de categorização semântica. No caso das respostas positivas, uma resposta seria emitida se uma determinada quantidade de ativação no sistema semântico, relativa à categoria alvo, fosse detectada. Por exemplo, em uma tarefa em que o participante tivesse que decidir se algo é um animal ou não, uma resposta positiva seria emitida se fosse detectado no sistema semântico um aumento na ativação de unidades que codificassem características como ‘anda’, ‘respira’, entre outras. Já as respostas negativas seriam emitidas quando a ativação no sistema semântico não atingisse o nível crítico em um determinado período de tempo (de forma análoga a da estratégia *deadline* na tarefa de decisão lexical). Essa explicação prevê que as respostas positivas serão mais rápidas que as negativas e que as respostas negativas não serão guiadas por informações provenientes do sistema lexical (como as respostas negativas são disparadas quando um determinado limite de tempo expira, essas nada têm a ver com o acesso lexical genuíno). Dessa forma, a proposta de Carreiras e cols. (1997) explicaria porque os efeitos de vizinhança ortográfica não têm sido encontrados nas tarefas de categorização semântica, uma vez que, em todos os estudos com o paradigma tradicional, a manipulação experimental concentrou-se nas respostas negativas.

Uma explicação alternativa para como uma resposta é emitida na tarefa de categorização semântica é a proposta por Pecher e cols. (2005). Na proposta desses autores, existiria apenas um mecanismo guiando a resposta dos participantes. Esse mecanismo seria

sensível ao acúmulo de evidências para uma resposta positiva ou negativa. A idéia é a mesma da proposta de Carreiras e cols. (1997) para as respostas positivas, porém essa é estendida às respostas negativas também. Por exemplo, Pecher e cols. (2005) propõem que em uma tarefa em que o participante tem que decidir se uma palavra representa algo animado ou inanimado, esse pode monitorar se em seu sistema semântico mais unidades relacionadas a coisas animadas estão sendo ativadas (p.ex., ‘respira’, ‘é orgânico’, ‘pode morrer’, etc.) ou se estão sendo ativadas mais unidades relacionadas a coisas inanimadas (p.ex., ‘feito de aço’, ‘é um artefato’, ‘movido a gasolina’, etc.). A resposta (animado ou inanimado) seria emitida assim que as evidências acumuladas a favor de uma das duas categorias ultrapassassem um determinado critério. Dessa forma, na proposta de Pecher e cols. (2005), a congruência semântica (ver capítulo ‘Revisão da Literatura’) entre a palavra alvo e seus vizinhos ortográficos torna-se um componente fundamental para a resposta dos participantes, porque os vizinhos ortográficos que compartilham com a palavra alvo uma mesma categoria semântica podem facilitar na ativação de unidades compartilhadas no sistema semântico. Assim, a proposta de Pecher e cols. (2005) prevê que uma resposta será mais rápida ou não na tarefa de categorização semântica não por ser positiva ou negativa, mas sim, pelo fato dos vizinhos ortográficos da palavra alvo serem semanticamente congruentes com ela ou não. Essa proposta também é capaz de explicar os efeitos inconsistentes de vizinhança ortográfica encontrados nos estudos com a tarefa de categorização semântica, pois é possível argumentar que esses ocorreram porque a congruência semântica não foi controlada nessas tarefas.

Tendo essas duas hipóteses alternativas em mente, procurou-se desenvolver um experimento fazendo uso da tarefa de categorização semântica no qual ambas as hipóteses pudessem ser avaliadas e que, mesmo assim, fosse possível fazer uso das mesmas palavras alvo utilizadas nos experimentos anteriores desse estudo. Para tanto, adotou-se uma tarefa de categorização semântica na qual os participantes tinham que decidir se uma palavra

representava o nome de uma cor ou não. Adotou-se o mesmo procedimento proposto por Forster e Shen (1996), isso é, todas as palavras que correspondiam à manipulação fatorial de N e NF pertenciam à categoria representada pela resposta “não” (nenhuma delas representava o nome de uma cor). A vantagem desse procedimento é que no caso das respostas “sim”, podem ocorrer efeitos de tipicidade (isso é, algumas palavras poderiam ser representantes mais característicos de nomes de cores do que outras) que seriam difíceis de controlar, enquanto se tentasse, ao mesmo tempo, manipular N e NF. Já no caso das respostas negativas, esses efeitos de tipicidade não ocorreriam (Forster & Shen, 1996). Para evitar que o efeito de congruência semântica prejudicasse a avaliação do efeito de vizinhança ortográfica, optou-se por manter constante a congruência semântica das palavras alvo no caso das respostas negativas. Assim sendo, todos os vizinhos ortográficos das palavras alvo, nessa condição, são semanticamente congruentes com essas (nenhum dos vizinhos ortográficos representa o nome de uma cor). Já no caso das respostas positivas (palavras que representam nomes de cores), todos os vizinhos ortográficos das palavras dessa condição eram incongruentes semanticamente com essas (nenhum dos vizinhos ortográficos representa o nome de uma cor). Considerando agora a proposta de Pecher e cols. (2005), se o que importa é a congruência semântica dos vizinhos ortográficos com as palavras alvo e não o tipo de resposta, os participantes devem apresentar tempos de reação mais rápidos para respostas negativas (onde todos os vizinhos ortográficos são semanticamente congruentes com as palavras alvo) do que para respostas positivas (onde todos os vizinhos ortográficos são semanticamente incongruentes com as palavras alvo). Além disso, no caso das respostas negativas, como todos os vizinhos ortográficos são semanticamente congruentes, deve-se observar um efeito facilitador de N, pois à medida que aumentam os vizinhos ortográficos, aumentam também os vizinhos semanticamente congruentes (a proposta de Pecher & cols., 2005 não parece prever um efeito de NF). Já, se a proposta de Carreiras e cols. (1997) estiver correta,

independentemente da congruência semântica dos vizinhos ortográficos, as respostas positivas devem ser mais rápidas do que as respostas negativas, porque as respostas negativas são emitidas quando um determinado limite de tempo é expirado (análogo à estratégia *deadline*). Além disso, como as respostas negativas não são guiadas por informações lexicais, não se deve esperar quaisquer efeitos de N ou NF nessa condição.

9.2 Método

9.2.1 Participantes

Participaram desse experimento 32 alunos do curso de Psicologia da UFAL, sendo essa uma amostra de conveniência. A média de idade dos participantes foi de 21 anos e 2 meses (desvio padrão de 5 anos e 9 meses), sendo 24 do sexo feminino e 8 do sexo masculino.

9.2.2 Material

Os estímulos experimentais consistiram de 64 palavras que atenderam a uma manipulação fatorial 2 x 2 de N e NF, sendo controladas as seguintes características: frequência de ocorrência (em média 10 ocorrências por milhão de palavras); número de sílabas (todas dissílabas); grau de correspondência no mapeamento grafema-fonema (todas

regulares); classe gramatical (todas substantivos); natureza do fonema inicial (pareados entre as condições experimentais); e, número de vizinhos transpostos (apenas um por condição experimental). Além disso, nenhum dos vizinhos ortográficos dessas palavras representava o nome de uma cor, sendo assim, devido à tarefa de categorização semântica utilizada nesse experimento, todos podem ser considerados congruentes semanticamente com as palavras. Essas 64 palavras foram subdivididas em quatro listas (pareadas pela natureza do fonema inicial e pela frequência de ocorrência média) que foram cruzadas entre as quatro condições experimentais. Por fim, a ordem de apresentação das listas foi contrabalanceada entre as quatro condições experimentais, de forma que cada participante foi exposto a apenas uma lista por condição experimental, configurando, assim, um total de 16 palavras (maiores detalhes podem ser obtidos no capítulo ‘Método’).

Para compor a tarefa de categorização semântica, além dos estímulos experimentais previamente descritos, 16 palavras que eram nomes de cores foram acrescentadas a essa tarefa. Todas essas palavras tinham duas sílabas, variando de quatro a seis letras, tendo uma média de frequência de ocorrência de 10,13 vezes por milhão de palavras (desvio padrão de 9,81), ou seja, uma frequência de ocorrência comparável à dos estímulos experimentais (10,11 ocorrências por milhão de palavras). O número médio de vizinhos ortográficos das palavras que representavam nomes de cor foi de 2,5, com um desvio padrão de 2,73 (bem próximo do número de vizinhos ortográficos das palavras correspondentes à manipulação experimental: média de 3,16, com desvio padrão de 2,46). Além disso, praticamente nenhum dos vizinhos ortográficos dessas palavras representava o nome de uma cor. Infelizmente, foi impossível selecionar apenas palavras com vizinhos ortográficos incongruentes semanticamente e assim sendo, duas palavras que tinham um vizinho ortográfico que também representava o nome de uma cor, tiveram que ser incluídas. Porém, cabe ressaltar que essas duas palavras tinham muitos vizinhos ortográficos e desse modo, eram muito mais incongruentes semanticamente

com esses do que congruentes. Para redimir quaisquer dúvidas, decidiu-se calcular o grau de congruência semântica dessas duas palavras e, para tanto, se adotou o seguinte procedimento: dividiu-se o número de vizinhos ortográficos da palavra que eram semanticamente congruentes com ela, pelo número total de vizinhos ortográficos da palavra. Assim, um valor próximo de um, indica uma palavra muito congruente semanticamente com seus vizinhos ortográficos e um valor próximo de zero, indica uma palavra muito incongruente. As duas palavras que tiveram um vizinho ortográfico semanticamente congruente tiveram valores de congruência semântica iguais a 0,25 e 0,17, indicando que, mesmo assim, são palavras bastante incongruentes semanticamente com seus vizinhos ortográficos. A lista de nomes de cores utilizada nesse experimento encontra-se no Apêndice B.

Por fim, esse experimento contou também com 10 palavras (cinco representavam nomes de cor e cinco não representavam nomes de cor) que foram utilizadas na sessão de treinamento. Esses estímulos tinham características semelhantes às das palavras utilizadas na sessão experimental. Assim sendo, cada participante foi exposto a 10 estímulos na sessão de treinamento e a trinta e dois estímulos na sessão experimental.

9.2.3 Procedimentos

Os participantes foram testados individualmente em uma sala cedida pela UFAL, onde foram recebidos e instruídos por um colaborador do pesquisador quanto aos procedimentos. O instrumento utilizado para o teste foi um computador de arquitetura compatível à IBM-PC equipado com um *joystick* ergonômico e o *software* utilizado para a apresentação dos estímulos e coleta dos dados de tempo de reação e porcentagem de erros foi o DMDX (Forster

& Forster, 2003). O participante se acomodava nas cadeiras, ficando a uma distância aproximada de 40 centímetros da tela do computador, dependendo de sua postura. Após se acomodar, o colaborador o convidava a ler as instruções experimentais na tela do computador. As instruções foram as seguintes: “Nesse teste, uma palavra aparecerá na tela do computador. Você deve apertar o botão 5 se ela for o nome de uma cor ou então você deve apertar o botão 6 se ela não for uma cor. Você deve tentar fazer isso o mais rápido possível, mas evite errar”. Após ler essas instruções o participante tinha a oportunidade de redimir quaisquer dúvidas antes de iniciar a sessão de treinamento.

Os estímulos foram apresentados em uma configuração de tela de 640x480 *pixels*, na fonte ‘*fixedsys*’ de tamanho 10 e em letras capitulares, sendo a cor da fonte branca e o fundo azul. A exposição dos estímulos teve a seguinte dinâmica de apresentação na sessão experimental: uma marca de fixação (+) aparecia na tela por 500ms e depois desaparecia, sendo imediatamente seguida pela apresentação da palavra alvo que durava até o tempo limite para uma resposta (3 segundos desde a apresentação da palavra alvo). O participante devia pressionar um botão, se a palavra alvo representasse uma cor e outro botão, caso não representasse uma cor, dentro do tempo limite para a resposta. As palavras alvo que representavam a manipulação experimental de N e NF desse estudo não representavam nomes de cores e assim sendo, sempre requeriam uma resposta negativa. A sessão de treinamento seguiu a mesma dinâmica, porém contou com *feedback* quanto à resposta dada. Além disso, um participante só avançava à sessão experimental quando cometia menos do que 15% de erros na sessão de treinamento, do contrário, a sessão de treinamento era repetida. O tempo total de testagem, incluindo as sessões de treinamento e experimental, foi de, aproximadamente, 5 minutos por participante.

9.3 Resultados

Devido a um erro de digitação a palavra estímulo ‘ÁSIA’ ficou sem o acento agudo (‘ASIA’) e, portanto, foi excluída das análises estatísticas realizadas. Como discutido no capítulo ‘método’, optou-se nesse estudo por utilizar o método das medidas semi-restringidas para se lidar com a distribuição positivamente assimétrica do TR e, por conseguinte, os escores de um participante que se desviaram de sua média, por mais de dois desvios padrão em qualquer direção, foram limitados a esse valor (Perea, 1999). Esse procedimento alterou 5,3% dos escores desse experimento e resultou em uma distribuição normal do TR (Kolmogorov-Smirnov $Z = 0,697$, $p > 0,71$).

Os participantes apresentaram uma média de porcentagem de erros de 4,59%, com um desvio padrão de 3,55%, sendo que a média da porcentagem de erros para responder às palavras que representavam nomes de cores foi de 3,71% (desvio padrão de 4,73%) e às palavras que não representavam nomes de cores foi de 5,57% (desvio padrão de 5,29%). O TR médio, considerando-se respostas positivas e negativas, foi de 771,53ms, com um desvio padrão de 116,28ms. A Tabela 14 apresenta as estatísticas descritivas para uma manipulação fatorial 2 x 2 de N e NF.

Tabela 14 – Média e Desvio Padrão do TR na tarefa de categorização semântica

	N1		N2	
	NF1	NF2	NF1	NF2
Média	831,84	823,69	828,25	831,94
Desvio padrão	143,48	157,28	183,92	202,26

Nota. N1 = apenas 1 vizinho ortográfico; N2 = 4 ou mais vizinhos ortográficos; NF1 = sem vizinhos ortográficos de maior frequência de ocorrência; NF2 = um vizinho ortográfico de maior frequência de ocorrência.

Uma análise de variância do tipo 2 x 2 para um *design* de medidas repetidas tendo como variável dependente o tempo de reação foi desenvolvida de acordo com as

recomendações de Pollatsek e Well (1995) e conforme discutido no capítulo ‘método’. Essa análise não indicou qualquer efeito estatisticamente significativo, seja de N, NF ou sua interação (todos os $p > 0,6$). Além dessa análise, um teste t para medidas repetidas foi utilizado para avaliar a comparação planejada entre o tempo de reação para responder às palavras que representavam cores (respostas positivas) e o tempo de reação para responder às palavras que não representavam cores (respostas negativas). A média do TR dos participantes para responder às palavras que representavam cores (média 713,44ms, com desvio padrão de 96,52ms) diferiu-se significativamente da média do TR para responder às palavras que não representavam cores (média de 829,62ms, com desvio padrão de 159,82ms), $t(31) = 5,26$ e $p < 0,001$. No geral, os participantes foram cerca de 116ms mais rápidos para responder que uma palavra representava o nome de uma cor (resposta positiva) do que para responder que não representava (resposta negativa). Como a porcentagem de erros foi pequena, os únicos dados analisados foram os referentes ao tempo de reação.

9.4 Discussão

No que diz respeito à proposta de Pecher e cols. (2005), essa prediz que as palavras com um maior grau de congruência semântica (as palavras que não representavam cores, no caso desse experimento) devem ser reconhecidas mais rapidamente do que as palavras com um menor grau de congruência semântica (no caso desse experimento, as palavras que representavam cores). No entanto, observou-se um padrão oposto nos resultados do presente estudo, pois as respostas positivas (palavras que representavam cores) foram mais rápidas do que as negativas (palavras que não representavam cores). Além disso, a proposta de Pecher e

cols. também prediz um efeito facilitador de N para as respostas negativas, uma vez que todos os vizinhos ortográficos das palavras dessa condição são semanticamente congruentes com essas. Porém, isso também não foi observado, pois o efeito de N foi nulo ($p > 0,6$) e na direção oposta (se é que se pode falar em algum efeito, N teve um efeito inibidor de, aproximadamente, 2ms). Já, ao contrário do que ocorre com a proposta de Pecher e cols., de uma forma geral, pode-se dizer que os resultados são claramente compatíveis com a proposta de Carreiras e cols. (1997), porque, como predito, as respostas positivas foram mais rápidas do que as negativas e não se observou qualquer efeito de N ou NF para as respostas negativas.

Para concluir, é necessário comparar os resultados desse experimento realizado com a tarefa de categorização semântica com os resultados dos outros experimentos desse estudo. Assumindo-se a interpretação de Carreiras e cols. (1997) como a mais apropriada, pode-se perceber porque o presente experimento, diferentemente dos outros desse estudo, não apresentou quaisquer efeitos de N ou NF. De acordo com Carreiras e cols., as respostas negativas na tarefa de categorização semântica não são emitidas com base em processos lexicais, mas sim, com base em uma estratégia *deadline* que não leva em consideração nenhuma característica relevante das palavras do ponto de vista ortográfico (afinal, a resposta é emitida quando um determinado limite de tempo é ultrapassado e a duração desse limite não parece ser derivada com base em N ou NF). Assim, é possível compreender a diferença dos resultados desse experimento para os resultados dos experimentos com as tarefas de identificação perceptual e de leitura em voz alta do tipo *Go/NoGo*, uma vez que os últimos utilizaram tarefas que nitidamente envolvem processos lexicais (como o processo de identificação lexical que ocorre no léxico ortográfico). Mesmo no caso da tarefa de leitura em voz alta, em que se procurou enfatizar uma leitura “estratégica” (não lexical), a diferença é compreensível, porque a estratégia empregada visava capitalizar na ativação compartilhada de fonemas e vizinhos fonológicos e nesse sentido, em certa medida, também envolvia processos

lexicais (ativação de vizinhos no léxico de *output* fonológico), embora fossem diferentes do processo de identificação lexical. Portanto, os resultados da tarefa de categorização semântica foram diferentes de todos os outros porque essa foi a única tarefa na qual a resposta provavelmente foi guiada por um processo exclusivamente não lexical. Por fim, parece sensato supor que qualquer modelo de reconhecimento visual de palavras que pretenda lograr sucesso em uma explicação da tarefa de categorização semântica terá que prever mecanismos capazes de possibilitar o uso da estratégia *deadline* nessa tarefa, conforme proposto por Carreiras e cols. (1997).

10 META-ANÁLISE E DISCUSSÃO GERAL

De uma forma geral, pode-se dizer que o presente estudo teve três objetivos principais:

1) investigar os efeitos de N e NF em falantes do português brasileiro; 2) investigar se os efeitos dessas variáveis podem ser modulados pelo tipo de tarefa utilizada na investigação; 3) avaliar se o DRC (Coltheart & cols., 2001) é um modelo explicativo adequado para descrever os efeitos de N e NF em falantes do português do Brasil. Os quatro capítulos anteriores lidaram explicitamente com o último objetivo (avaliação do modelo DRC). No entanto, para uma efetivação mais adequada dos dois primeiros objetivos desse estudo, torna-se necessário que os resultados dos experimentos desenvolvidos sejam comparados e/ou combinados. Uma das formas mais conhecidas de se combinar ou comparar os resultados de diferentes experimentos é a realização de uma meta-análise. Assim sendo, no presente capítulo será realizada uma meta-análise dos estudos de vizinhança ortográfica realizados com falantes do português brasileiro. Essa meta-análise tem como objetivo responder a três perguntas principais que estão diretamente relacionadas aos dois objetivos mais gerais desse estudo: os efeitos de N e NF interagem um com o outro ou não? Os efeitos de N e NF variam de acordo com o tipo de tarefa utilizada? Se sim, em que tipos de tarefas os efeitos são inibidores e em que tarefas eles são facilitadores?

10.1 Meta-Análise

10.1.1 Critérios de Inclusão

Os estudos incluídos nessa meta-análise consistiram nos estudos de vizinhança ortográfica realizados com falantes do português brasileiro que anteriormente haviam sido discutidos na revisão da literatura (isso é, os dois estudos de Justi & Pinheiro, 2006; 2008) e os quatro experimentos desenvolvidos no presente estudo. Dessa forma, a meta-análise foi realizada com base em 8 experimentos (quatro provenientes dos estudos de Justi e Pinheiro e quatro do presente estudo) que investigaram os efeitos de N e NF em falantes do português brasileiro. Infelizmente, até o presente momento, esses parecem ter sido os únicos estudos que investigaram empiricamente tais efeitos em falantes do português do Brasil. Isso foi confirmado com base em uma pesquisa, realizada em setembro de 2009, nas bases de dados ‘BVS-PSI’ e ‘PsycInfo’, utilizando-se os termos “vizinhança ortográfica” para ambas as bases e “orthographic neighborhood” e “portuguese” para a base de dados ‘PsycInfo’.

10.1.2 Padronização dos Resultados

Uma vantagem dos estudos incluídos nessa meta-análise é que todos eles utilizaram uma manipulação fatorial 2 x 2 de N e NF e a análise de variância como teste estatístico (com correções para levar em conta a variação referente ao efeito dos itens na análise). Dessa forma, optou-se por estimar a magnitude do efeito de N e NF tanto no geral quanto em níveis específicos. Isso é, calculou-se a magnitude do efeito de N, independentemente, dos níveis de NF (efeito geral) e também se computou a magnitude do efeito de N quando as palavras não

tinham vizinhos de maior frequência de ocorrência e quando essas tinham vizinhos de maior frequência de ocorrência. No caso de NF, o procedimento foi isomórfico, isso é, calculou-se a magnitude do efeito de NF, independentemente, dos níveis de N (efeito geral) e também se computou a magnitude do efeito de NF quando as palavras tinham poucos vizinhos ortográficos e quando essas tinham muitos vizinhos ortográficos. Como medida da magnitude do efeito de N e NF, optou-se pelo cômputo de r pelas seguintes razões: esse é facilmente calculável com base no F da ANOVA; não precisa de ajustes no seu cômputo se o *design* for entre ou intra sujeitos; e, é mais simples de interpretar porque tem o mesmo significado de uma correlação (Rosenthal, 1984). Além disso, os níveis de significância referentes aos efeitos de N e NF foram todos convertidos em valores p unicaudais. A direção dos efeitos foi sinalizada sempre por um sinal negativo para os efeitos facilitadores e por um sinal positivo para os efeitos inibidores. Por fim, quando um mesmo experimento apresentava os efeitos de N e NF para dois tipos diferentes de variáveis dependentes, os valores r e p referentes a essas duas variáveis eram combinados em um só para evitar que um único experimento contribuísse duas vezes para a meta-análise.

10.1.3 Busca por Moderadores

Uma das principais perguntas que essa meta-análise visa responder é se os efeitos de N e NF variam dependendo da tarefa utilizada. Uma forma de se avaliar tal questão é testar a heterogeneidade da magnitude do efeito ou dos níveis de significância de N e NF. Se houver uma variação sistemática nos resultados dos oito experimentos como um todo, é esperado que o resultado do teste de heterogeneidade seja estatisticamente significativo. Além disso, os

resultados dos estudos podem ser agrupados de acordo com o tipo de tarefa ou predição teórica, nesse caso, o esperado é que o resultado do teste de heterogeneidade em um determinado agrupamento não seja estatisticamente significativo. Para se testar a heterogeneidade dos efeitos de N e NF, os seguintes procedimentos foram adotados: primeiro, os testes difusos de heterogeneidade para a magnitude do efeito foram calculados de acordo com a fórmula [4.15] proposta por Rosenthal (1984, p.77) e os testes difusos de heterogeneidade para o nível de significância foram calculados utilizando-se o programa *Meta-Analysis* versão 5.3, desenvolvido por Schwarzer (1989a; 1989b); segundo, os resultados foram considerados heterogêneos sempre que, pelo menos um desses dois testes foi estatisticamente significativo (ambos os testes têm como base a distribuição X^2). Assim sendo, sempre que um resultado for considerado heterogêneo, o teste mais estatisticamente significativo será relatado e quando ambos os testes não forem significantes, aquele mais próximo de ter sido será relatado. Tendo essas considerações em mente, optou-se por calcular o nível de significância combinado e a magnitude do efeito combinado, apenas para os agrupamentos de estudos que não pudessem ser considerados heterogêneos em relação ao nível de significância e/ou em relação à magnitude do efeito. O nível de significância foi combinado (p médio) de acordo com o método de Stouffer e a magnitude do efeito foi combinada transformando-se os valores r em sua contrapartida Z_r de Fisher, calculando-se o Z_r médio e transformando-o, finalmente, no valor r correspondente (r médio). Para ambos os procedimentos o programa *Meta-Analysis* versão 5.3 (Schwarzer, 1989a; 1989b) foi utilizado.

Outra questão importante a ser considerada é se os efeitos de N e NF são interativos. Para avaliar essa hipótese basta comparar a magnitude do efeito de N quando as palavras não têm vizinhos de maior frequência de ocorrência e a magnitude do efeito de N quando as palavras têm vizinhos de maior frequência de ocorrência. Se o efeito de N não interagir com NF, não devem existir diferenças estatisticamente significativas entre essas duas condições. O

mesmo raciocínio pode ser aplicado a NF. Ou seja, se o efeito de NF não interagir com N, não devem existir diferenças entre a magnitude do efeito de NF quando as palavras têm poucos vizinhos ortográficos e a magnitude do efeito de NF quando as palavras têm muitos vizinhos ortográficos. Para efetivar tais comparações, contrastes focais entre os níveis específicos de N e entre os níveis específicos de NF foram realizados, conforme a fórmula [4.27] proposta por Rosenthal (1984, p.84).

10.1.4 Resultados

A magnitude dos efeitos (r) de N e NF foram calculadas com base no valor da estatística F da análise de variância, conforme a fórmula [2.17] proposta por Rosenthal (1984, p.25) e todos os níveis de significância correspondentes foram transformados em sua contrapartida unicaudal. Apesar disso, embora não seja esse o procedimento padrão em uma meta-análise, a menos que seja explicitado o contrário, como os modelos de reconhecimento visual de palavras fazem previsões opostas para os efeitos de N e NF, todos os testes realizados foram bicaudais. A Tabela 15 apresenta o valor p unicaudal e a magnitude dos efeitos gerais de N e NF, de acordo com o tipo de tarefa empregada.

Tabela 15 – Magnitude do efeito (geral) e nível de significância de N e NF por tipo de tarefa

Estudo	Tarefa	VD	n	N		NF	
				<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Justi e Pinheiro (2006)	TDL (GNG)	TR	25	+0,163	0,070	+0,111	0,164
	TDL	TR+%ER	24	-0,205	0,005	+0,007	0,448
Justi e Pinheiro (2008)	TDL (GNG)	TR+%ER	34	+0,210	0,003	+0,207	0,003
	TDL	TR+%ER	28	-0,011	0,468	+0,103	0,089
	TLVA (GNG)	TR	32	+0,097	0,300	+0,244	0,089
TESE	TLVA	TR	32	-0,288	0,055	-0,187	0,153
	TIP	%ER	32	-0,354	0,023	+0,520	0,001
	TCS	TR	32	+0,027	0,442	-0,029	0,438

Nota. VD = variável dependente; n = número de participantes; *r* = magnitude do efeito (um sinal positivo indica efeitos inibidores e um sinal negativo indica efeitos facilitadores); *p* = nível de significância unicaudal; TDL = tarefa de decisão lexical; TLVA = tarefa de leitura em voz alta; TIP = tarefa de identificação perceptual; TCS = tarefa de categorização semântica; GNG = tarefa do tipo *Go/NoGo*; TR = tempo de reação; %ER = porcentagem de erros; TR+%ER = indica que os valores *r* e *p* relativos ao tempo de reação e a porcentagem de erros foram combinados.

Como se pode observar na Tabela 15, a magnitude dos efeitos de N e NF varia bastante ao longo dos oito experimentos, indo de -0,35 a +0,21 para N e de -0,19 a +0,52 para NF. Testes de heterogeneidade considerando todas as tarefas em conjunto, indicaram que esses resultados são claramente heterogêneos, seja para N [$X^2(7) = 22,98$, $p < 0,01$], seja para NF [$X^2(7) = 13,83$, $p = 0,05$]. Considerando-se a possibilidade de interação entre N e NF, dois testes focais foram realizados: um entre os níveis específicos de N e o outro entre os níveis específicos de NF. No primeiro, comparando a magnitude do efeito de N quando as palavras não tinham vizinhos ortográficos de maior frequência de ocorrência, com a magnitude do efeito de N quando as palavras tinham vizinhos ortográficos de maior frequência de ocorrência, não se observou uma diferença estatisticamente significativa ($Z = 0,49$, $p > 0,62$). No segundo, dessa vez comparando a magnitude do efeito de NF quando as palavras tinham poucos vizinhos ortográficos, com a magnitude do efeito de NF quando as palavras tinham muitos vizinhos ortográficos, também não se observou uma diferença estatisticamente significativa ($Z = 0,38$, $p > 0,70$). Assim sendo, optou-se por não incluir na Tabela 15 dados

relacionados aos níveis específicos de N e NF, mas apenas seu efeito geral. Por fim, um dado interessante é que a correlação entre magnitude do efeito (r) e o nível de significância (p) foi forte ($r = -0,79$, $p < 0,01$, $n = 16$). Isso ocorreu, muito provavelmente, devido ao número bastante similar de participantes em cada experimento.

Uma vez que ao se combinar os resultados de todos os estudos independentemente do tipo de tarefa, esses foram heterogêneos, passou-se a investigar se o tipo de tarefa poderia ser um moderador dos resultados. Os resultados foram então agrupados de acordo com o seguinte critério: tarefas que enfatizam o processo de identificação lexical (tarefa de identificação perceptual; tarefas de decisão lexical do tipo *Go/NoGo* e tarefa de leitura em voz alta do tipo *Go/NoGo*) e tarefas que enfatizam o uso de algum tipo de estratégia (tarefas de decisão lexical; tarefa de leitura em voz alta e tarefa de categorização semântica). Considerando-se apenas as tarefas que enfatizam o processo de identificação lexical, testes de heterogeneidade indicaram que os resultados são heterogêneos para N [$X^2(3) = 12,09$, $p < 0,01$], mas não para NF [$X^2(3) = 3,21$, $p > 0,36$]. O efeito de NF foi inibidor, tendo uma magnitude média de 0,28, sendo estatisticamente significante (p médio $< 0,001$). Os testes focais realizados para avaliar a possibilidade de interação entre N e NF não foram significantes (para todos $p > 0,9$). Já quando foram consideradas apenas as tarefas que enfatizam o uso de algum tipo de estratégia, testes de heterogeneidade indicaram que os resultados não são heterogêneos, seja para N [$X^2(3) = 4,99$, $p > 0,17$], seja para NF [$X^2(3) = 2,88$, $p > 0,40$]. O efeito de N foi facilitador, tendo uma magnitude média de -0,12, sendo estatisticamente significante (p médio $< 0,05$). Já o efeito de NF foi nulo (magnitude média de -0,03 e p médio $> 0,8$). Os testes focais realizados para avaliar a possibilidade de interação entre N e NF não foram significantes (para todos $p > 0,4$).

Pode-se argumentar que agrupar a tarefa de categorização semântica junto com as tarefas que enfatizam o uso de estratégias não é estritamente adequado, porque a estratégia

que estaria presente na tarefa de categorização semântica não é sensível a N ou NF. Sendo assim, considerando-se apenas as tarefas que enfatizam o uso de algum tipo de estratégia, porém sem incluir a tarefa de categorização semântica, testes de heterogeneidade continuaram indicando que os resultados não são heterogêneos, seja para N [$X^2(2) = 3,16, p > 0,2$] ou para NF [$X^2(2) = 2,81, p > 0,24$]. O efeito de N continuou sendo facilitador, tendo uma magnitude média de $-0,17$, e sendo estatisticamente significativo (p médio $< 0,01$). O efeito de NF continuou sendo nulo (magnitude média de $-0,03$ e p médio $> 0,7$). Os testes focais realizados para avaliar a possibilidade de interação entre N e NF não foram significantes (para todos $p > 0,26$).

Outro tipo de agrupamento de tarefas que foi analisado foi o das tarefas que enfatizam o processo de identificação lexical, porém, sem a inclusão da tarefa de identificação perceptual. Como essa tarefa foi a que apresentou os maiores efeitos, decidiu-se analisar novamente os resultados para ver se manteriam o mesmo padrão, após a exclusão dessa tarefa. Sendo assim, considerando-se apenas as tarefas que enfatizam o processo de identificação lexical, porém, sem incluir a tarefa de identificação perceptual, testes de heterogeneidade indicaram que os resultados não são heterogêneos, seja para N [$X^2(2) = 2,49, p > 0,28$], ou para NF [$X^2(2) = 1,74, p > 0,4$]. O efeito de N foi inibidor, tendo uma magnitude média de $0,16$, sendo estatisticamente significativo (p médio $< 0,01$). O efeito de NF continuou sendo inibidor, tendo uma magnitude média de $0,19$, sendo estatisticamente significativo (p médio $< 0,01$). Os testes focais realizados para avaliar a possibilidade de interação entre N e NF não foram significantes (para todos $p > 0,57$).

Por fim, o último agrupamento de tarefas analisado foi o das tarefas que requeriam algum tipo de resposta vocal (tarefa de leitura em voz alta do tipo *Go/NoGo*, tarefa de leitura em voz alta e tarefa de identificação perceptual). Considerando-se apenas esse grupo de tarefas, testes de heterogeneidade indicaram que os resultados não são heterogêneos para N

[$X^2(2) = 3,67, p > 0,15$], mas são para NF [$X^2(2) = 8,56, p < 0,02$]. Porém, o efeito de N não foi estatisticamente significativo (magnitude média de $-0,19$ e p médio = $0,08$). Os testes focais realizados para avaliar a possibilidade de interação entre N e NF não foram significantes (para todos $p > 0,14$).

10.1.5 Discussão

Uma das perguntas levantadas nessa meta-análise foi se os efeitos de N e NF interagem um com o outro ou não. Os resultados dessa meta-análise claramente indicam que os efeitos dessas variáveis não são interativos. Todos os testes realizados para avaliar a possibilidade de interação entre N e NF obtiveram resultados nulos, considerassem esses testes todos os experimentos realizados ou agrupamentos específicos desses. Dessa forma, uma interpretação mais parcimoniosa para a interação observada no experimento realizado com a tarefa de leitura em voz alta do tipo *Go/NoGo* é que, devido a algum tipo de erro de medida ou por ser uma tarefa menos poderosa do que a tarefa de identificação perceptual ou ambos, ela falhou em detectar os efeitos gerais de N e NF. Assim sendo, pode-se concluir que os efeitos de N e NF não são interativos, pelo menos, no que diz respeito aos estudos realizados com falantes do português brasileiro.

Outra questão levantada nessa meta-análise é: os efeitos de N e NF variam de acordo com o tipo de tarefa utilizada? Se sim, em que tipos de tarefas os efeitos são inibidores e em que tarefas eles são facilitadores? Observando-se os resultados da meta-análise, a resposta à primeira pergunta é claramente sim, afinal, ao serem considerados os resultados de N e NF, independentemente do tipo de tarefa, esses foram heterogêneos. Para responder à segunda

pergunta, considerou-se a proposta de Grainger e Jacobs (1996) e de Carreiras e cols. (1997) de que em tarefas que envolvessem mais fortemente componentes relacionados ao acesso lexical genuíno, seria possível observar efeitos inibidores de NF e em tarefas em que estratégias como a de adivinhação rápida pudessem estar mais envolvidas, deveria ser possível observar um efeito facilitador de N. Pode-se observar que essa hipótese foi corroborada nas análises desenvolvidas, porque em todos os agrupamentos de tarefas em que foram consideradas apenas as tarefas que envolveriam estratégias, o efeito de N foi facilitador e homogêneo e o efeito de NF foi nulo e homogêneo. Esse é um dado interessante, porque corrobora a hipótese levantada no capítulo ‘Tarefa de Leitura em Voz Alta’ de que o efeito facilitador de N não foi detectado no experimento desenvolvido por falta de poder estatístico. Por outro lado e também conforme predito pela hipótese de Grainger e Jacobs e de Carreiras e cols., em todos os agrupamentos de tarefas em que foram consideradas apenas tarefas que envolveriam mais fortemente o processo de acesso lexical (em especial, a identificação lexical), o efeito de NF foi inibidor e homogêneo. No que diz respeito a N, pode-se dizer que seu efeito foi menos claro, porque apenas quando a tarefa de identificação perceptual foi excluída é que seu efeito foi inibidor e homogêneo, do contrário esse foi nulo e heterogêneo. Pensou-se na hipótese de que o efeito de N também pudesse ser moderado pelo tipo de resposta exigida nas tarefas. Imaginou-se que em tarefas que exigissem respostas vocais, o efeito de N teria uma tendência mais forte a um efeito facilitador, devido à possibilidade de ativação de fonemas compartilhados entre a palavra alvo e seus vizinhos ortográficos. De certo modo, essa hipótese foi corroborada, porque o efeito de N foi facilitador ($r = -0,19$) e não heterogêneo para esse agrupamento de tarefas. Assim sendo, pode-se dizer que parte da heterogeneidade observada no efeito de N quando foram agrupadas as tarefas que envolviam acesso lexical, deveu-se a diferenças no tipo de resposta entre as tarefas (tarefas em que um botão deveria ser pressionado e tarefas em que uma resposta vocal deveria ser emitida). Já

com relação à NF, isso não ocorreu, porque quando foram agrupados os experimentos que exigiam respostas vocais independentemente do tipo de tarefa, o efeito de NF foi heterogêneo, indicando que, no seu caso, o tipo de tarefa é um moderador mais importante do que o tipo de resposta. Esses resultados corroboram a hipótese exposta no capítulo ‘Tarefa de Identificação Perceptual’ de que o processo de ativação de fonemas compartilhados entre a palavra alvo e seu vizinho ortográfico pode ser mais sensível a N do que a NF.

10.2 Discussão Geral

Tendo em mente o panorama geral sobre os efeitos de N e NF no português brasileiro, pode-se então avaliar se esse padrão de efeitos se assemelha ao padrão que seria predito pelos diferentes modelos de reconhecimento visual de palavras. Como o modelo DRC (Coltheart & cols., 2001) foi extensivamente discutido em cada um dos experimentos específicos dessa tese, ele não será discutido novamente aqui. Cumpre apenas assinalar que como na meta-análise os efeitos de N e NF não foram interativos, não parece ser necessária nenhuma modificação no modelo para explicar os efeitos inibidores de NF em tarefas nas quais a ênfase era no processo de identificação lexical. Já no que diz respeito aos efeitos facilitadores de N, o modelo é compatível com a existência desses efeitos em tarefas em que algum tipo de processamento estratégico foi enfatizado. No entanto, como a meta-análise indicou que os efeitos de N também podem ser modulados pelo tipo de resposta exigido nas tarefas, parece ser necessário que o modelo aumente a força da ativação repassada pelo léxico de *output* fonológico ao nível de fonema para acomodar os efeitos facilitadores de N em tarefas de leitura em voz alta em geral.

No que diz respeito aos outros modelos de reconhecimento visual de palavras, pode-se dizer que esses, devido aos resultados dessa meta-análise seriam, em princípio, modelos incompletos, já que a meta-análise, claramente, indicou que os efeitos de N e NF podem ser modulados pelo tipo de ênfase das tarefas utilizadas (em processos de acesso lexical ou em processos estratégicos) e nenhum desses modelos traz previsões explícitas a esse respeito. O que esses modelos precisam é desenvolver uma forma de prever e explicar os efeitos estratégicos e, embora seja possível argumentar que esse não é o objetivo dos modelos, esses poderão ser sempre considerados incompletos à luz de modelos que explicam ambas as situações (acesso lexical e processamento estratégico). Além disso, a menos que um modelo seja capaz de especificar quais tarefas enfatizariam mais o processo de identificação lexical, como definir, de forma não arbitrária, quais tarefas representam uma avaliação genuína do modelo e quais não? Destarte, ser capaz de fazer esse tipo de previsão não é uma característica trivial dos modelos, mas sim, uma condição fundamental para que esses possam ser falseáveis (Popper, 1972). Do contrário, será sempre possível argumentar que os efeitos desfavoráveis a um modelo foram encontrados em tarefas que enfatizam o uso de estratégias e que, dessa forma, não pertencem ao escopo desse modelo. Sendo assim, esses modelos são capazes de prover apenas explicações parciais para os dados obtidos com falantes do português brasileiro. De qualquer forma, é profícuo comparar as previsões desses modelos com os dados obtidos nessa meta-análise para averiguar até que ponto seriam necessárias modificações nos mecanismos centrais desses modelos.

No que diz respeito aos modelos de busca e verificação (MB de Forster, 1976; 1992; Murray & Forster, 2004 e MV de Paap, Newsome, McDonald & Schvaneveldt, 1982; Paap & Johansen, 1994) esses predizem um efeito inibidor de NF e um efeito nulo de N. Assim sendo, esses modelos seriam bastante adequados aos dados obtidos nas tarefas em que o processo de acesso lexical (em especial de identificação lexical) estaria mais envolvido, uma

vez que nessas observou-se um efeito inibidor de NF e um efeito nulo de N. No entanto, esses modelos teriam dificuldade em explicar a tendência a efeitos facilitadores de N observada em tarefas que envolviam respostas vocais e, como já ressaltado, nada têm a dizer sobre os efeitos de N e NF em tarefas que enfatizam o uso de estratégias. No entanto, talvez esses modelos não tenham muitos problemas para se adequar aos dados, uma vez que seus pressupostos mais centrais não precisam ser alterados para explicar o efeito inibidor de NF observado em tarefas que enfatizam o processo de identificação lexical, sendo necessária apenas a inclusão de um mecanismo que permita respostas estratégicas e que seja sensível a N. Como esse mecanismo poderia ser extra-lexical, a inclusão desse não precisaria alterar a ordem de busca/comparação nessa classe de modelos que é sensível a frequência de ocorrência dos estímulos e é responsável pelo efeito inibidor de NF (ver capítulo 2).

Ao se considerar os modelos PPD (Harm e Seidenberg, 2004; Seidenberg & McClelland, 1989; Plaut & cols., 1996), conforme discutido no capítulo 2, pode-se perceber que esses modelos predizem efeitos facilitadores de N e de NF, esteja envolvida na tarefa a necessidade de se computar um *output* fonológico (como nas tarefas de leitura em voz alta) ou um *output* ortográfico (como nas tarefas de decisão lexical). Assim sendo, esses modelos estão em claro contraste com os dados obtidos na meta-análise, porque, raramente, N e NF apresentaram os mesmos efeitos (nas tarefas classificadas como envolvendo mais fortemente o processo de identificação lexical, NF apresentou um efeito inibidor e N um efeito nulo; e, nas tarefas classificadas como envolvendo processamento estratégico NF apresentou um efeito nulo e N um efeito facilitador). A única vez em que as duas variáveis apresentaram o mesmo padrão de efeitos foi quando se analisou as tarefas que enfatizariam o processo de identificação lexical sem incluir a tarefa de identificação perceptual. Nesse caso, ambas variáveis apresentaram um efeito inibidor. Comparados aos modelos de Busca e Verificação, os problemas dos modelos PPD parecem ser mais centrais. Isso ocorre porque os efeitos de N

e NF são derivados de um mesmo mecanismo no modelo: o benefício que as unidades responsáveis por gerar o padrão de ativação de uma palavra alvo (unidades escondidas) obtêm do treino com palavras similares a essa palavra alvo. Desse modo, por princípio, os efeitos de N e NF nos modelos PPD devem sempre apresentar a mesma direção facilitadora, sendo difícil para o modelo acomodar, ao mesmo tempo, um efeito de N e a ausência de efeito de NF, por exemplo. Mesmo que se tentasse argumentar que os efeitos de N são mais fortes e que, portanto, seria possível observar um efeito de N e um efeito nulo de NF, esse argumento seria problemático, porque dependendo do tipo de tarefa, foi o efeito de NF que foi o mais forte. Lembre-se ainda que na única situação em que as duas variáveis apresentaram efeitos na mesma direção, esses efeitos foram inibidores. Sendo assim, parece que os modelos PPD encontram-se em pior situação, porque para se adequarem aos dados não é necessário apenas propor um mecanismo extra-lexical para acomodar possíveis efeitos estratégicos (a rigor, defensores dos modelos PPD devem argumentar que os efeitos estratégicos são os inibidores e não os facilitadores), mas também é necessário algum tipo de modificação em uma característica central dos modelos que é o fato dos efeitos de N e NF derivarem de um mesmo mecanismo de aprendizagem e por conseguinte, apresentarem sempre a mesma direção.

O último modelo de reconhecimento de palavras que falta ser avaliado no que diz respeito aos efeitos de N e NF em falantes do português brasileiro é o modelo SERIOL (Whitey, 2001; 2004; Whitney & Lavidor, 2005; Whitney & Cornelissen, 2008). Esse modelo, conforme discutido no capítulo 2 dessa tese, prevê que os efeitos de N e NF devem ser interativos. Isso ocorreria porque o efeito de vizinhança ortográfica depende de uma interação entre o mecanismo de *feedback* que existe entre o nível das palavras e o nível das letras e o mecanismo de inibição lateral que existe no nível das palavras (Whitney, 2004; Whitney & Lavidor, 2005). Para Whitney e Lavidor, quando existem muitos vizinhos, mas nenhum deles é muito similar à palavra alvo, o mecanismo de *feedback* deve dominar o

processo, pois a inibição lateral não é tão forte a ponto de cancelar os benefícios de uma ativação acelerada das letras e, desse modo, o resultado esperado é um efeito facilitador do número de vizinhos ortográficos. Já quando ocorre menos *feedback* (o número de vizinhos é pequeno) e a similaridade ortográfica entre o vizinho e a palavra alvo é maior (talvez pela presença de um vizinho de maior frequência), o mecanismo de inibição lateral domina o processo e o resultado predito é um efeito inibidor. Assim sendo, o modelo prediz um efeito facilitador de N quando as palavras não têm vizinhos de maior frequência de ocorrência e um efeito inibidor de NF quando as palavras têm poucos vizinhos ortográficos. O principal problema com esse modelo é que a meta-análise não trouxe evidências de que os efeitos de N e NF sejam interativos e assim sendo, alguma modificação na dinâmica de operação do mecanismo de *feedback* e do mecanismo de inibição lateral do modelo SERIOL torna-se necessária. Desse modo, embora o modelo, em princípio, tenha potencial para prever efeitos diferentes de N e NF (o que não ocorre com os modelos PPD), é difícil precisar que tipos de ajustes seriam necessários e, além disso, esse modelo também teria que incorporar mecanismos que lhe permitissem prever efeitos estratégicos também.

Uma vez discutida a adequação dos diferentes modelos de reconhecimento visual de palavras aos dados obtidos nesse estudo e em outros realizados com falantes do português brasileiro, pode-se também comentar sobre as duas hipóteses mais gerais propostas na literatura para explicar os resultados conflitantes dos diferentes estudos. Conforme discutido no capítulo revisão da literatura, uma das hipóteses aposta que diferenças nas demandas das tarefas em relação a processos lexicais ou a processos estratégicos poderiam modular os efeitos de N e NF (Carreiras & cols., 1997; Grainger & Jacobs, 1996); enquanto outra hipótese aposta em diferenças entre as línguas nas quais os estudos foram realizados, sejam essas diferenças no grau de consistência do mapeamento grafema-fonema (Andrews, 1997) ou na estrutura de vizinhança ortográfica das palavras das diferentes línguas (Siakaluk & cols.,

2002). Embora o objetivo dessa tese não seja decidir entre esses dois tipos de hipóteses mais gerais, já que para isso seria necessário um estudo translingüístico, essa tese traz evidências que indiretamente podem contribuir para esse debate.

No que diz respeito à proposta de Grainger e Jacobs (1996) e de Carreiras e cols. (1997), pode-se dizer que o padrão de resultados observado nos experimentos dessa tese corrobora bastante bem essa hipótese porque os efeitos de N e NF, de fato, variaram de acordo com o padrão predito pela hipótese. Porém, é importante ressaltar que o fato da hipótese de Grainger e Jacobs (1996) e de Carreiras e cols. (1997) ter sido corroborada no presente estudo, não exclui a possibilidade das hipóteses de Andrews (1997) e de Siakaluk e cols. (2002) também estarem certas, afinal, os dois tipos de hipóteses não são mutuamente excludentes. É logicamente possível que os efeitos de N e NF possam ser modulados por diferenças entre as tarefas e também por diferenças entre as línguas nas quais os estudos foram realizados. Desse modo, no que diz respeito à hipótese de que diferenças entre as línguas nas quais os estudos foram realizados poderiam modular os efeitos de N e NF, é necessário avaliar individualmente suas duas versões. A versão que atribui as diferenças nos resultados dos estudos a diferenças na estrutura de vizinhança ortográfica das palavras das diferentes línguas (Siakaluk & cols., 2002) foi, de certo modo, avaliada no capítulo ‘Análise de Corpus’ dessa tese. Essa hipótese propõe que, como na língua inglesa a maioria das palavras tem vizinhos de maior frequência de ocorrência, os efeitos de NF deveriam ser facilitadores nessa língua. Ao se comparar a estrutura de vizinhança ortográfica das palavras do português brasileiro à das palavras da língua inglesa, observou-se que as diferenças na proporção de palavras com vizinhos de maior frequência de ocorrência foram pequenas e, além disso, na direção oposta à predita pela hipótese de Siakaluk e cols. (2002). Isso é, no português brasileiro, uma proporção ainda maior de palavras tem vizinhos de maior frequência de ocorrência e mesmo assim o efeito de NF parece ser inibitório. Destarte, as

análises desenvolvidas nesse estudo não dão muito suporte a essa hipótese. Por fim, a hipótese de Andrews (1997) não foi avaliada diretamente nessa tese, mas, mesmo assim, é possível tecer algumas considerações sobre ela. Em certo sentido, pode-se dizer que a hipótese de Andrews (1997) situa o *locus* do efeito de N no processo de mapeamento grafema-fonema, já que pretende capitalizar em diferenças na transparência desse mapeamento entre as ortografias para explicar as diferenças no efeito de N no inglês (onde o mapeamento é mais inconsistente) em relação a línguas como o espanhol e o francês (onde o mapeamento é mais consistente). Por conseguinte, se, pelo menos, um dos *loci* do efeito de N relacionar-se ao processo de mapeamento grafema-fonema, o efeito dessa variável também deve ser modulado pelo tipo de resposta exigida nas tarefas, já que respostas vocais implicam em processos de produção de fonemas e outros tipos de resposta não, necessariamente, implicam nesses processos. Nesse sentido, pode-se dizer que a meta-análise desenvolvida nessa tese traz algum suporte a essa hipótese, já que o efeito de N foi homogêneo em tarefas que envolviam respostas vocais, a despeito do tipo de ênfase (seja em processamento lexical ou estratégico) presente nessas tarefas. No entanto, ao mesmo tempo, os achados dessa tese também trazem dificuldade à hipótese de Andrews (1997), porque o efeito de N nas tarefas que envolviam respostas vocais realizadas com falantes do português brasileiro tendeu a ser facilitador como no inglês, mesmo o português sendo uma língua na qual o mapeamento grafema-fonema é bem mais consistente do que no inglês. Destarte, na melhor das hipóteses, a proposta de Andrews (1997) obtém suporte apenas parcial do presente estudo.

11 CONCLUSÃO

O objetivo mais geral do presente trabalho foi o de investigar o efeito de vizinhança ortográfica no reconhecimento visual de palavras em adultos falantes do português do Brasil. Nesse sentido, um dos primeiros desafios que se encontrou foi o de definir qual seria a melhor medida de similaridade ortográfica a ser adotada nesse estudo. Para responder a essa questão, uma análise de corpus lingüístico foi desenvolvida. Nessa análise, foi possível observar que as medidas tradicionais de vizinhança ortográfica, N e NF, se correlacionam muito pouco com TLN, a medida rival de similaridade ortográfica, sendo que, além disso, afetam o processamento de uma proporção muito maior de palavras do que TLN. Assim sendo, um dos primeiros achados dessa tese é o de que, no que diz respeito ao português brasileiro, N e NF são melhores medidas de similaridade ortográfica do que TLN. Outro achado interessante da análise de corpus desenvolvida é que a estrutura de vizinhança ortográfica das palavras do português e do inglês é bastante similar, colocando em dúvida a proposta de Siakaluk e cols. (2002) de que diferenças no efeito de N e NF entre as línguas possam se dever a diferenças na estrutura de vizinhança ortográfica das palavras dessas línguas.

O passo seguinte para a efetivação do objetivo geral de investigar o efeito de vizinhança ortográfica no reconhecimento visual de palavras de falantes do português brasileiro envolveu tentar distinguir componentes relacionados ao acesso lexical e ao uso de estratégias que poderiam estar presentes em diferentes tarefas. Para tanto, o modelo DRC de Coltheart e cols. (2001) foi utilizado como um modelo explicativo dos componentes que estariam presentes nas tarefas utilizadas no presente estudo. Com base nesse modelo foram desenvolvidos quatro experimentos para investigar os efeitos de N e NF no reconhecimento visual de palavras: uma tarefa de leitura em voz alta do tipo *Go/NoGo*; uma tarefa de leitura

em voz alta onde os participantes deviam ler palavras e pseudopalavras em uma mesma sessão; uma tarefa de identificação perceptual e uma tarefa de categorização semântica. As tarefas utilizadas nessa tese foram agrupadas com tarefas desenvolvidas em outros estudos com falantes do português brasileiro e, o uso de técnicas meta-analíticas, permitiu que comparações fossem traçadas entre tarefas com ênfases similares e tarefas com ênfases divergentes. De uma forma geral, foram observadas duas classes principais de tarefas: tarefas que enfatizavam o processo de acesso lexical (tarefas de identificação perceptual; tarefas de decisão lexical e de leitura em voz alta do tipo *Go/NoGo* nas quais as pseudopalavras utilizadas tinham muitos vizinhos ortográficos); e, tarefas que enfatizavam o processamento estratégico (tarefas de decisão lexical em que as pseudopalavras tinham poucos vizinhos ortográficos; tarefas de leitura em voz alta em que os participantes deviam ler palavras e pseudopalavras em uma mesma sessão e tarefas de categorização semântica). Assim sendo, estudos futuros podem utilizar essas duas classes de tarefas para investigar os efeitos de outras variáveis, seja em tarefas que enfatizam o acesso lexical, seja em tarefas que enfatizam o processamento estratégico.

Considerando-se os efeitos de N e NF nos experimentos realizados nessa tese e em outros estudos desenvolvidos com falantes do português brasileiro, de uma forma geral, pode-se dizer que os efeitos de N e NF são inibidores em tarefas nas quais o processo de identificação lexical é predominante (embora o efeito de N seja mais fraco) e, em tarefas em que o processamento estratégico predomina, observam-se efeitos facilitadores de N e nulos de NF. Como o modelo DRC (Coltheart & cols., 2001) foi bastante consistente com esse padrão de resultados, pode-se dizer que esse é um modelo bastante adequado para descrever o acesso ao léxico ortográfico no português brasileiro. Já no que diz respeito aos outros modelos de reconhecimento visual de palavras que foram discutidos nessa tese, pode-se dizer que sua maior dificuldade é não terem desenvolvido predições específicas que lhes permitam prever

em que condições devem estar mais presentes processos de acesso lexical e/ou processos estratégicos.

Por fim, pode-se dizer que os resultados dessa tese se coadunam bem com a proposta de Grainger e Jacobs (1996) e Carreiras e cols. (1997) de que os efeitos de N e NF podem ser modulados pela demanda das tarefas utilizadas para se investigar o acesso lexical. No entanto, é importante se considerar também que esses dados se referem apenas a estudos realizados com falantes do português brasileiro e que, embora com base nesses dados seja legítimo postular que os efeitos de N e NF podem ser modulados pelo tipo de tarefa utilizada, nada impede que diferenças entre as línguas nas quais os estudos foram conduzidos também possam modular os efeitos de N e NF. Afinal, as duas explicações não são mutuamente excludentes. Sendo assim, estudos futuros poderiam estender a estratégia meta-analítica utilizada no presente estudo, para investigar se os efeitos de N e NF também podem ser modulados por diferenças entre as línguas nas quais os estudos foram desenvolvidos. Outra questão que o presente estudo também deixou em aberto é se os efeitos de N são sempre facilitadores em tarefas que envolvem respostas vocais, porque embora os efeitos de N tenham sido homogêneos nessas tarefas, o efeito facilitador observado nesse conjunto de tarefas não foi estatisticamente significante.

REFERÊNCIAS

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 15, 802-814.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 234-254.
- Andrews, S. (1996). Lexical retrieval and selection processes: effects of transposed-letter confusability. *Journal of Memory and Language*, 35, 775-800.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4, 439-461.
- Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: a matter of some moment? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 514-544.
- Arduino, L., & Burani, C. (2004). Neighborhood effects on nonword visual processing in a language with shallow orthography. *Journal of Psycholinguistic Research*, 33, 75-95.
- Balota, D., & Chumbley, J. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, 10, 340-357.

- Balota, D., Pilotti, M., & Cortese, M. (2001). Subjective frequency estimates for 2938 monosyllabic words. *Memory and Cognition*, 29, 639-647.
- Bowers, J., Davies, C., & Hanley, D. (2005) Interfering neighbours: the impact of novel word learning on the identification of visually similar words. *Cognition*, 97, B45-B54.
- Capovilla, F.; Macedo, E.; Duduchi, M. & Sória, R. (1997). Análise computadorizada de leitura em voz alta via rotas fonológica e lexical. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação*, 1, 81-140.
- Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 857-871.
- Chow, S. (1998). Précis of statistical significance: rationale, validity, and utility. *Behavioral and Brain Sciences*, 21, 169-239.
- Clark, H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- Coleman, E. (1964). Generalizing to a language population. *Psychological Reports*, 14, 219-226.
- Cohen, J. (1976). Random means random. *Journal of Verbal Learning and Verbal Behavior*, 15, 261-262.
- Cohen, J. (1988) *Statistical Power for Behavioral Sciences*. New York: Academic Press.
- Cohen, J. (1994) The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.

- Coltheart, M. (1978). Lexical access in simple reading tasks. Em: G. Underwood. (Ed.) *Strategies of information processing* (pp. 151-216). New York: Academic Press.
- Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In: S. Dornic (Ed.) *Attention and performance VI* (pp. 535-555). Hillsdale: Erlbaum.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.
- Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37, 65-70.
- Ellis, A. (2004). Length, formats, neighbours, hemispheres, and the processing of words presented laterally or at fixation. *Brain and Language*, 88, 355-366.
- Eysenck, M., & Keane, M. (2007) *Manual de Psicologia Cognitiva 5ª Ed.* Porto Alegre: Artmed.
- Forster, K. (1976) Accessing the mental lexicon. Em: R. Wales & E. Walker. (Orgs.) *New approaches to language mechanisms.* Amsterdam: North-Holland.
- Forster, K. (1992) Memory-addressing mechanisms and lexical access. Em: R. Frost & L. Katz. (Orgs.) *Orthography, phonology, morphology, and meaning.* Amsterdam: North-Holland.
- Forster, K., & Dickinson, R. (1976). More on the language-as-a-fixed-effect fallacy: Monte Carlo estimates of error rates for F1, F2, F', and min F'. *Journal of Verbal Learning and Verbal Behavior*, 15, 135-142.

- Forster, K., & Forster, J. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers*, 35, 116-124.
- Forster, K., & Shen, D. (1996). No enemies in the neighborhood: absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 696-713.
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29, 228-244.
- Grainger, J. (2008). Cracking the orthographic code: an introduction. *Language and Cognitive Processes*, 23, 1-35.
- Grainger, J., & Jacobs, A. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*, 103, 518-565.
- Grainger, J., O'Reagan, K., Jacobs, A., & Segui, J. (1989). On the role of competing word units in visual word recognition: the neighborhood frequency effect. *Perception & Psychophysics*, 45, 189-195.
- Grainger, J., O'Reagan, K., Jacobs, A., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics*, 51, 49-56.
- Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: a comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, 47, 191-198.
- Grainger, J., & Whitney, C. (2004). Does the human mind read words as a whole? *Trends in Cognitive Sciences*, 8, 58-59.

- Harm, M. & Seidenberg, M. (2004). Computing the meaning of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, 111, 662-720.
- Huntsman, L., & Lima, S. (1996). Orthographic neighborhood structure and lexical access. *Journal of Psycholinguistic Research*, 25, 417-429.
- Huntsman, L., & Lima, S. (2002). Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research*, 31, 289-306.
- Johnson, N., & Pugh, K. (1994). A cohort model of visual word recognition. *Cognitive Psychology*, 26, 240-346.
- Justi, F. R. R. & Pinheiro, A. M. V. (2006). O efeito de vizinhança ortográfica no português do Brasil: acesso lexical ou processamento estratégico. *Interamerican Journal of Psychology*, 40, 275-288.
- Justi, F. R. R. & Pinheiro, A. M. V. (2008). O efeito de vizinhança ortográfica em crianças brasileiras: estudo com a tarefa de decisão lexical. *Interamerican Journal of Psychology*, 42, 559-569.
- Justi, F. R. R., & Justi, C. N. G. (em preparação). Que tipo de pseudopalavra é mais difícil de rejeitar em uma tarefa de decisão lexical?
- Kantowitz, B., Roediger, H., & Elmes, D. (2006) *Psicologia Experimental*. São Paulo: Thomson Learning Edições.
- Kuhn, T. S. (1970). *A Estrutura das Revoluções Científicas*. São Paulo: Perspectiva.

- Laxon, V., Gallagher, A., & Masterson, J. (2002). The effects of familiarity, orthographic neighbourhood density, letter-length and graphemic complexity on children's reading accuracy. *British Journal of Psychology*, 93, 269-287.
- Laxon, V., Masterson, J., Pool, M., & Keating, C. (1992). Nonword naming: Further exploration of the pseudohomophone effect in terms of orthographic neighborhood size, graphemic changes, spelling - sound consistency, and reader accuracy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 730-748.
- Lewis, D. (1995). *Análise de Variância*. São Paulo: Harbra.
- LINGUATECA (2009). *Descrição do corpus NILC e sua descendência*. Disponível em: <<http://www.linguateca.pt/acesso/NILCsaocarlos.html>>. Acesso em 30 ago. 2009.
- McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: pt. 1, an account of basic findings. *Psychological Review*, 88, 375-407.
- Monsell, S., Doyle, M., & Haggard, P. (1989) Effects of frequency on visual word recognition tasks: where are they? *Journal of Experimental Psychology: General*, 118, 43-71.
- Moor, W., & Brysbaert, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research*, 63, 159-162.
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, 111, 721-756.
- NILC (2005). *Corpus NILC / São Carlos v.7.1*. Disponível em: <<http://www.nilc.icms.usp.br/nilc/tools/corpora.htm>>. Acesso em 30 ago. 2005.

- Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing frequency effect: a retest of the verification model. *Journal of Experimental Psychology: Human Perception & Performance*, 20, 1129-1157.
- Paap, K., Newsome, S., McDonald, J. & Schvaneveldt, R. (1982) An activation verification model for letter and word recognition: the word superiority effect. *Psychological Review*, 89, 573-594.
- Pecher, D., Zeelenberg, R., & Wagenmakers, E. (2005) Enemies and Friends in the neighborhood: orthographic similarity effects in semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 121-128.
- Peereman, R., & Content, A. (1995). Neighborhood size effect in naming: lexical activation or sublexical correspondences?. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 409-421.
- Perea, M. (1999). Tiempos de reacción y psicología cognitiva: dos procedimientos para evitar el sesgo debido al tamaño muestral. *Psicológica*, 20, 13-21.
- Perea, M., Carreiras, M., & Grainger, J. (2004). Blocking by word frequency and neighborhood density in visual word recognition: a task-specific response criteria account. *Memory & Cognition*, 32, 1090-1102.
- Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*, 24, 767-779.

- Perea, M. & Rosa, E. (1999). Psicología de la lectura y procesamiento léxico visual: una revisión de técnicas experimentales y de procedimientos de análisis. *Psicológica*, 20, 65-90.
- Perea, M., & Rosa, E. (2000). The effects of orthographic neighborhood in reading and laboratory word identification tasks: a review. *Psicológica*, 21, 327-340.
- Perea, M., Rosa, E., & Gómez, C. (2003). Influence of neighborhood size and exposure duration on visual-word recognition: evidence with the yes / no and the go / no-go lexical decision tasks. *Perception & Psychophysics*, 65, 273-286.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Pinheiro, A. M. V. (1996). *Contagem de frequência de ocorrência e análise psicolinguística de palavras expostas a crianças na faixa pré-escolar e séries iniciais do 1º grau*. São Paulo: Associação Brasileira de Dislexia.
- Pinheiro, A. M. V., & Reis, C. (2004) A influência da natureza fonética do fonema inicial da palavra na medida de tempo de reação em tarefas de leitura em voz alta. *Revista de Estudos da Linguagem*, 12, 19-36.
- Plaut, D., McClelland, J., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, 103, 56-115.
- Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*, 25, 1142-1158.

- Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: a suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 785-794.
- Popper, K. (1972). *A lógica da pesquisa científica*. São Paulo: Cultrix.
- Pugh, K. R., Rexer, K., Peter, M., & Katz, L. (1994). Neighborhood effects in visual word recognition: effects of letter delay and nonword context difficulty. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20, 639-648.
- Raaijmakers, J. (2003). A further look at the “language-as-fixed-effect fallacy”. *Canadian Journal of Experimental Psychology*, 57, 141-151.
- Raaijmakers, J., Schrijnemakers, J., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416-426.
- Ratcliff, R., & McKoon, G. (1996). Bias effects in implicit memory tasks. *Journal of Experimental Psychology: General*, 125, 403-421.
- Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review*, 104, 319-343.
- Rayner, K., Juhasz, B., & Pollatsek, A. (2005). Eye movements during reading. In: M. Snowling & C. Hulme (Eds.). *The science of reading: A handbook* (pp. 79-97). Oxford: Blackwell Publishing.
- Reichle, E., Pollatsek, A., Fisher, D., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125-157.

- Reichle, E., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: comparisons to other models. *Behavioral and Brain Sciences*, 26, 445-526.
- Reynolds, M., & Besner, D. (2002). Neighbourhood density effects in reading aloud: new insights from simulations with the DRC Model. *Canadian Journal of Experimental Psychology*, 56, 310-318.
- Roazzi, A., Justi, C. N. G., Justi, F. R. R. (2008). Da tinta à mente: uma discussão sobre os modelos computacionais de reconhecimento visual de palavras. Em: M. R. Maluf & S. R. K. Guimarães. (Orgs.) *Desenvolvimento da Linguagem Oral e Escrita* (pp. 95-121). Curitiba: Editora UFPR.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills: Sage.
- Rubenstein, H., Garfield, L., & Millikan, J. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9, 487-494.
- Rumelhart, D. & McClelland, J. (1982) An interactive activation model of context effects in letter perception: pt. 2, the contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60-94.
- Schwarzer, R. (1989a). *Meta-analysis Programs*. Disponível em: <<http://userpage.fu-berlin.de/~health/manual.exe>>. Acesso em 30 jun. 2008.
- Schwarzer, R. (1989b). *Meta-analysis 5.3*. Disponível em: <<http://userpage.fu-berlin.de/~health/meta53.exe>>. Acesso em 30 jun. 2008.

- Sears, C., Campbell, C., & Lupker, S. (2006). Is there a neighborhood frequency effect in English? Evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1040-1062.
- Sears, C., Hino, Y., & Lupker, S. (1995). Neighborhood size and neighbourhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, 16, 65-76.
- Sears, C., Hino, Y., & Lupker, S. (1999). Orthographic Neighbourhood effects in parallel distributed processing models. *Canadian Journal of Experimental Psychology*, 53, 220-229.
- Sears, C., Lupker, S., & Hino, Y. (1999). Orthographic Neighborhood effects in perceptual identification and semantic categorization tasks: a test of the multiple read-out model. *Perception and Psychophysics*, 61, 1537-1554.
- Seidenberg, M., & McClelland, J. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Seymour, P. H. K. (2005). Early reading development in European orthographies. In M. Snowling & C. Hulme (Orgs.) *The Science of Reading: a handbook* (pp. 296-315). Oxford: Blackwell.
- Siakaluk, P., Sears, C., & Lupker, S. (2002) Orthographic neighborhood effects in lexical decision: the effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 661-681.
- Snodgrass, J., & Minzer, M. (1993). Neighbourhood effects in visual word recognition: facilitatory or inhibitory? *Memory & Cognition*, 21, 247-266.

- Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, 124, 107-136.
- Warner, R. M. (2008) *Applied statistics: from bivariate through multivariate techniques*. London: Sage Publications.
- Whaley, C. P. (1978). Word-Nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143-154.
- Whitney, C. (2001) How the brain encodes the order of letters in a printed word: the SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, 8, 221-243.
- Whitney, C. (2004) Hemisphere-specific effects in word recognition do not require hemisphere-specific modes of access. *Brain and Language*, 88, 279-293.
- Whitney, C. & Cornelissen, P. (2008) SERIOL reading. *Language and Cognitive Processes*, 23, 143-164.
- Whitney, C. & Lavidor, M. (2005) Facilitative orthographic neighborhood effects: the SERIOL model account. *Cognitive Psychology*, 51, 179-213.
- Wickens, T., & Keppel, G. (1983). On the choice of design and of test statistic in the analysis of experiments with sampled materials. *Journal of Verbal learning and Verbal Behavior*, 22, 296-309.
- Winer, B. J. (1971) *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

Zeelenberg, R., Wagenmakers, E. M., & Raaijmakers, J. G. W. (2002). Repetition priming in implicit memory tasks: prior study causes enhanced discriminability, not only bias. *Journal of Experimental Psychology: General*, 131, 38-47.

Ziegler, J., & Perry, C. (1998). No more problems in Coltheart's neighborhood: resolving neighborhood conflicts in the lexical decision task. *Cognition*, 68, B53-B62.

GLOSSÁRIO

acesso lexical: no caso do reconhecimento visual de palavras refere-se ao acesso à representação mental das palavras que reconhecemos em um texto.

bigramas / bigramas abertos: pares de letras contíguos e sequenciais que ocorrem em determinadas posições nas palavras. No caso dos bigramas abertos a contigüidade é desprezada. Por exemplo, na palavra ‘gato’, ‘ga’, ‘at’, ‘to’, seriam bigramas no sentido tradicional, já no caso dos bigramas abertos, ‘gt’, ‘go’ e ‘ao’, também seriam considerados bigramas.

Bin: espécie de “região” no sistema de acesso ortográfico do modelo *Bin* (Murray & Forster, 2004) onde palavras ortograficamente similares estão agrupadas.

efeito de vizinhança ortográfica: efeito que as palavras similares em termos ortográficos desencadeiam no processo de reconhecimento de um estímulo alvo. Esse efeito pode ser facilitador ou inibidor. Efeitos facilitadores resultam em mais rapidez e precisão e os inibidores têm resultado contrário.

identificação lexical: seleção da forma no léxico mental que melhor corresponde ao estímulo-alvo.

léxico mental: espécie de “dicionário interno” que, dependendo do modelo, conteria a representação mental do significado, da ortografia e da fonologia das palavras.

ms: abreviatura de milissegundos.

N: refere-se ao número de vizinhos ortográficos da palavra-alvo.

NF: refere-se à frequência de ocorrência dos vizinhos ortográficos da palavra-alvo, isso é, se a palavra-alvo possui vizinhos de frequência maior do que ela ou não.

pseudopalavra: seqüências de letras, construídas com estruturas ortográficas possíveis na língua em questão, mas não associadas a nenhum significado. No caso do português do Brasil, ‘bresa’ e ‘paixe’, seriam exemplos de pseudopalavras.

pseudopalavra ilegal: seqüências de letras que desrespeitam as regras de construção ortográfica de palavras na língua em questão. No caso do português do Brasil, ‘aypqz’ e ‘buwha’, seriam exemplos de pseudopalavras ilegais.

TR: abreviatura de Tempo de Reação. É, normalmente, indexado pelo tempo despendido da apresentação de um estímulo até a resposta (ou início da resposta) de um participante.

transcodificação lexical: geração de um código no sistema semântico e/ou fonológico a partir de um *input*.

TLN: *transposed letter neighbors* – vizinhos ortográficos formados pela transposição de duas letras de uma palavra alvo. Por exemplo, ‘asco’ é um vizinho ortográfico formado pela transposição de duas letras da palavra ‘saco’.

vizinho ortográfico: um vizinho ortográfico de uma palavra é uma palavra que se pode formar a partir dessa, com a mudança de apenas uma letra, enquanto se mantêm a posição das demais letras. Por exemplo, ‘dedo’ e ‘gado’ são vizinhos ortográficos de ‘dado’.

APÊNDICE A – PALAVRAS CORRESPONDENTES A MANIPULAÇÃO

FATORIAL DE N E NF

PALAVRAS COM POUCOS VIZINHOS ORTOGRÁFICOS E SEM VIZINHOS DE
MAIOR FREQUÊNCIA

ASTRO – BISPO – BICHOS – CRENÇA – CACAU – CREDOR – CAPIM – DUPLAS
FAUNA – FIBRAS – FILTRO – MANHÃS – PADRES – SAQUE – TRUQUE – TAMPÃO

PALAVRAS COM POUCOS VIZINHOS ORTOGRÁFICOS E COM VIZINHOS DE
MAIOR FREQUÊNCIA

ONZE – BALCÃO – BONÉ – CRISES – CARNÊS – CORAIS – CRIVO – DIVÃ
FICHAS – FRANGO – FUSCA – MUNDOS – PATRÃO – SABOR – TRIBO – TRONO

PALAVRAS COM MUITOS VIZINHOS ORTOGRÁFICOS E SEM VIZINHOS DE
MAIOR FREQUÊNCIA

ASAS – BURRO – BRIGAS – CALMA – CASCA – CONDE – CRAVO – DICAS
FERAS – FITAS – FARDOS – MAPAS – PINGA – SUCO – TAMPA – TUMBA

PALAVRAS COM MUITOS VIZINHOS ORTOGRÁFICOS E COM VIZINHOS DE
MAIOR FREQUÊNCIA

ÁSIA – BARES – BARRAS – CABOS – CALDO – CALÇA – CABRA – DATAS
FADAS – FUMO – FUNDA – MUROS – PICOS – SACO – TRAVE – TOMBO

**APÊNDICE B – PSEUDOPALAVRAS E PALAVRAS UTILIZADAS COMO
ESTÍMULOS COMPLEMENTARES**

PSEUDOPALAVRAS UTILIZADAS NA TAREFA DE LEITURA EM VOZ ALTA DO
TIPO *GO/NOGO*

ADRA – BUCA – CALCO – CICO – CURJA – DACA – FARBAS – FARCAS
FARSO – MARGA – PAUNA – PINSA – RÉDIO – SANQUE – TANHA – TRONCA

PSEUDOPALAVRAS UTILIZADAS NA TAREFA DE LEITURA EM VOZ ALTA
BRACOS – CAPE – CARDAS – CONDA – CRADOS – DIRO – FANTOS – FIMA
LAGRAS – MUCA – SISTO – SUDA – SULDO – TASTO – TRACO – TRUPO

PALAVRAS UTILIZADAS NA TAREFA DE CATEGORIZAÇÃO SEMÂNTICA
ANIL – BEGE – BRONZE – CÁQUI – CHUMBO – CINZA – COBRE – CREME
GELO – LILÁS – MARFIM – MARROM – PRATA – ROXO – SALMÃO – VINHO