ELIAS FERREIRA

PALAVRA FREQÜENTE, PRONÚNCIA DIFERENTE:
A Lingüística de Corpus auxiliando o ensino da pronúncia do inglês como língua estrangeira

MESTRADO EM
LINGÜÍSTICA APLICADA E ESTUDOS DA LINGUAGEM

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
2006

ELIAS FERREIRA

eliregbr@yahoo.com.br

PALAVRA FREQÜENTE, PRONÚNCIA DIFERENTE:
A Lingüística de Corpus auxiliando o ensino da pronúncia do inglês como língua estrangeira

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de MESTRE em Lingüística Aplicada e Estudos da Linguagem, sob orientação do Prof. Dr. Antonio Paulo Berber Sardinha.

PUC - SP
2006

BANCA EXAMINADORA

_____

_____

_____

À minha mãe, Antonia, que tudo fez e tudo suportou por amor a mim.

À minha amada esposa, Regina Aiko, que sempre esteve ao meu lado com seu amor, paciência, incentivo e servidão.

Ao meu pai (in memoriam), pelo amor profundo que recebi nos três primeiros anos de minha vida e pelas maravilhosas lembranças que carrego comigo até hoje desde sua partida.

...

Elaine: I just don't enjoy being with' im.

Jerry: Well, that's what's important.

Elaine: How do you pronounce S-C-O-U-R-G-E?

Jerry: /skɜrdʒ/.

Elaine: You see? I said /skɔrdʒ/.
And then Owen makes this really big deal about it
in front of this other couple. He really embarrassed me.
This is it! This is it!
I cannot date and watch my grammar at the same time!

(Laughs)

Jerry: It's not grammar. It's pronunciation.

Elaine: And don't you get smart!

(Seinfeld, 1991, The Alternate Side, extra scenes)

Agradecimentos

Ao Professor Dr. Tony Berber Sardinha, meu orientador, pela maneira que sempre via o que eu não via e que sempre me mostrou melhor caminho;

À CAPES, pelo apoio financeiro, fundamental para a realização deste trabalho;

À Professora Dra. Ângela Brambilla Cavenaghi Themudo Lessa, por seu excelente trabalho, que serviu de inspiração para esta dissertação;

À Professora Dra. Aglael Gama Rossi, pelo apoio e dedicação durante todo o curso;

Aos Professores Doutores Adauri Brezolin e Helena Gordon e à Professora Zaina Abdalla Nunes, pelo apoio na qualificação e defesa;

A Osvaldo Succi, cuja excelente dissertação de mestrado serviu como guia para o desenvolvimento desta;

Aos muitos amigos que fiz no LAEL, companheiros de aulas, seminários de orientação, InPLA's, enfim, companheiros de batalha: Carlos Kauffman, Cláudia Garcia, Giseli, Daniela, Renata Picasso, Denise, Renata Condi, Roberto, Adriana Rossini, Adriana Passoni, Lindinalva, Marta, Cláudia Rocha, Glauce, Lílian, Mauro, Marta, Maurício, Fabíola, Helmara e Elisa;

Aos funcionários do LAEL, Maria Lúcia, Márcia, Paulo, Rosangeles e Ricardo, pela dedicação.

Resumo

Este trabalho tem como objetivo descobrir quais são os vocábulos da língua inglesa que apresentam uma relação atípica entre a ortografia e a pronúncia e que têm freqüência de uso relevante, observada por meio de um corpus.

O resultado deste trabalho poderá ter posterior aplicação na formação de professores brasileiros de inglês, orientando a preparação dos mesmos em relação à área de pronúncia de vocábulos a partir da forma escrita, indicando quais palavras necessitam receber maior atenção durante o processo de formação acadêmica, atuando assim como um trabalho de referência.

Os resultados poderão também posteriormente ser utilizados como referência por elaboradores de material teórico e didático, oferecendo maior especificidade para o caso falante brasileiro.

Ao aprender inglês, criamos padrões de pronúncia para certos grafemas ou seqüência de grafemas. Por exemplo, em palavras como *swear*, *sweat* e *sweet*, relacionamos facilmente os grafemas <sw> com os fonemas /sw/, e pronunciamos /swer/, /swet/ e /swit/. Porém, ao encontrarmos um vocábulo como *sword*, a tendência da grande maioria dos brasileiros é aplicar as mesmas regras de decodificação grafema-fonema e pronunciarmos erroneamente */swɔrd/, ao invés da forma correta /sɔrd/.

Outro exemplo é *bury* /beri/, "enterrar" em português, que, por causa de sua ortografia, parece conduzir a maioria dos falantes brasileiros a pronúncias como */bʌri/, */bjʊri/ ou */bʊri/. Entretanto, permanece a questão sobre a relevância da palavra *bury*: com base na freqüência de uso de *bury* na língua inglesa, é importante incluir essa palavra no processo de ensino da pronúncia do inglês?

Para responder a essa pergunta, lançamos mão da Lingüística de Corpus e observamos a freqüência de uso dos vocábulos na língua através do corpus britânico de língua geral BNC (British National Corpus). Descobrimos, assim, que *bury* é uma palavra de relação ortografia-pronúncia atípica de uso freqüente, com a qual professores muito provavelmente entrarão em contato.

O objetivo desta pesquisa é, portanto, identificar os vocábulos de relação atípica entre a ortografia e a pronúncia do ponto de vista do falante letrado de português brasileiro, tomando como base não apenas a freqüência da pronúncia desses vocábulos no léxico inglês, mas também sua freqüência de uso observada num corpus de língua inglesa (Hunston, 2002:3).

Buscamos também saber quais são os grafemas e seqüências de grafemas de maior atipicidade, o que poderá ajudar a orientar professores, elaboradores de material didático e demais interessados na área a desenvolver suas atividades, mostrando quais grafemas merecem uma abordagem com maior ênfase.

Palavras-chave: pronúncia, Lingüística de Corpus, formação de professores.

Abstract

The main objective of this study is to identify the English words that show an atypical grapheme-phoneme correspondence and a relevant frequency of use.

The results might be applied to the training of Brazilian teachers of English, helping them to improve their pronunciation of written words, showing them which words need to be focused on during the academic training. It is our wish thereby to provide a reference study that might also help English book designers to focus more on the Brazilian case.

When learning English, we create some pronunciation patterns for some graphemes. For example, in words such as *swear*, *sweat,* and *sweet*, we easily relate the graphemes <sw> to the phonemes /sw/, and pronounce /swer/, /swet/ and /swit/. However, when we have to pronounce a word like *sword*, most Brazilians tend to apply the same grapheme-phoneme correspondence rules and wrongly pronounce */swɔrd/ instead of the correct form /sɔrd/.

Another example is the word *bury* /beri/, which due to its spelling seems to lead most Brazilian speakers to pronunciations like */bʌri/, */bjʊri/ or */bʊri/. However, the question about the relevance of *bury* remains: based on the frequency of use of *bury*, is it important to include this word in the English pronunciation teaching process?

To answer this question we turned to Corpus Linguistics and observed the frequency of use of the words by means of the British general corpus of English BNC (British National Corpus). We discovered thereby that *bury* is a word of atypical spelling-pronunciation correspondence and is also a word of frequent use with which teachers are highly likely to get in contact.

It is our objective, therefore, to identify the words with atypical spelling-pronunciation relationship from the point of view of a well educated Brazilian Portuguese speaker, based not only on the frequency of their spellings in the English lexis, but also on their frequency of use observed in an English language corpus (Hunston, 2003:3).

We also wanted to know which were the most atypical grapheme and grapheme strings, which might orient teachers, English material designers and anyone interested in the area to develop their activities, showing which graphemes we should pay more attention to.

Keywords: pronunciation, Corpus Linguistics, teacher training.

Sumário

Lista de Quadros e Figuras

Quadros

Figuras

Introdução

Of course every good teacher is an avid learner of the subject she teaches.

Medgyes (1994:40)

Atuando como professor de inglês em empresas e instituições particulares de ensino há oito anos na cidade de São Paulo, percebo a necessidade que temos de preparar com maior precisão os professores de inglês brasileiros não-bilíngües em relação à pronúncia do inglês. Mais especificamente, refiro-me à pronúncia de palavras que apresentam uma relação atípica entre a ortografia e a pronúncia, como por exemplo *gross* /groʊs/[1] e *sword* /sɔrd/, que são pronunciadas de maneira diferente da maioria das palavras com a mesma seqüência de grafemas[2] <oss>[3] (*boss* /bɔs/, *cross* /krɔs/, *loss* /lɔs/, *toss* /tɔs/ etc.) e <ord> (*lord* /lɔrd/, *cord* /kɔrd/, *Ford* /fɔrd/ etc).

A correspondência grafema-fonema no inglês é muito irregular: o mesmo grafema pode representar mais de um fonema, e o mesmo fonema pode ser representado por grafemas diferentes, compostos por uma ou mais letras (Steinberg, 1985:62).

Professores e alunos vêem na pronúncia do inglês um grande desafio, uma barreira a ser transposta, quase como um inimigo a ser conquistado (Wanke, 1987). Isso se deve a fatores, tais como:

a) Transferência dos padrões de pronúncia do português para o inglês: usar padrões do português para transformar grafemas em fonemas, como por exemplo pronunciar o <b> em *doubt*, haja vista que não há <b> não-pronunciado em português.

b) Generalização da pronúncia dentro da língua-alvo: crer que a correspondência grafema-fonema segue apenas um padrão, como, por exemplo, pronunciar o <uce> de *lettuce* /ˈletəs/ da mesma maneira que o <uce> de *produce* /prəˈdus/, *reduce* /rɪˈdus/ e *deduce* /dɪˈdus/.

---

[1] Transcrições retiradas do dicionário eletrônico CMU. Ver seção 2.5.
[2] Para a definição de grafema, ver seção 1.2.1
[3] Ao referir-nos a grafemas, usaremos "<>", conforme Crystal (1997:257) e Mori (2004:150).

Acredito que os professores necessitam de uma formação mais aprofundada sobre a pronúncia de vocábulos que exibem esse comportamento.

Creio também na necessidade de abordarmos a questão com maior especificidade para as necessidades do falante de português brasileiro, e também elaborarmos materiais teóricos e didáticos com essa orientação.

Por conta disso, comecei a desenvolver uma lista de palavras que, a meu ver, apresentavam tal relação atípica entre a ortografia e a pronúncia, e criei também uma atividade para trabalhá-las em aula. Agrupei as palavras de acordo com o tipo de problema apresentado e as trabalhava com os alunos através de exercícios.

Os resultados foram impressionantes. Os alunos ficavam chocados com a pronúncia daquelas palavras, cuja ortografia não parecia tão problemática. As reações foram todas muito parecidas e se resumiam a um sentimento e a uma pergunta:

1. um sentimento de culpa por pensar em quantas vezes eles já haviam pronunciado aquelas palavras de maneira errada em palestras, conversas ao telefone e em reuniões.

2. uma pergunta que vinha inevitavelmente, com uma ponta de rancor: por que meus **professores** nunca me ensinaram isso?

Contudo, para mim, o professor, também surgiam duas perguntas:

1. Será que não estou ensinando muitas palavras, que talvez eles não terão a oportunidade nem a necessidade de utilizar em suas atividades profissionais ou sociais?

2. De todos esses casos de palavras com relação atípica entre ortografia e pronúncia, quais são os casos mais importantes, aos quais devo dar mais ênfase durante meu ensino?

Para responder a estas perguntas, este trabalho utilizou a Lingüística de Corpus (Sinclair, 1991; McEnery & Wilson, 1997; Biber, Conrad & Reppen, 1998; Hunston, 2002; Berber Sardinha, 2004), a qual estuda a linguagem empiricamente, coletando grandes quantidades de textos e analisando-os através de ferramentas computacionais.

Desenvolvemos também ferramentas computacionais de análise de correspondência grafofonêmica com o propósito específico de encontrar respostas para os problemas acima descritos.

Esta dissertação divide-se em seis partes:

No **capítulo 1**, encontra-se toda a Fundamentação Teórica que sustentou nossa pesquisa, dividida em cinco seções: Lingüística de Corpus, Relação entre Fala e Escrita, Correspondência Grafofonêmica, Ensino da Pronúncia do Inglês como Língua Estrangeira e Ortografia do Inglês .

O **capítulo 2** descreve a metodologia de pesquisa e as ferramentas utilizadas para a obtenção dos resultados.

No **capítulo 3**, encontram-se a apresentação e análise dos resultados.

A seguir vêm as **considerações finais**, as **referências bibliográficas** e, fechando o trabalho, os **anexos** em CD-ROM, contendo todos os dados colhidos em nossa pesquisa.

Capítulo 1 – Fundamentação Teórica

The use of machines in linguistic analysis is now established.

Firth (1957:31)

1.1 LINGÜÍSTICA DE CORPUS

1.1.1 Lingüística e Tecnologia

Vivemos em um século de muitas mudanças, em uma era dinâmica, repleta de avanços tecnológicos e de desenvolvimento.

Na Lingüística, várias transformações também estão ocorrendo: os avanços tecnológicos têm fornecido aos lingüistas recursos para atingir uma profundidade cada vez maior na coleta e análise de dados, descrevendo as línguas com precisão nunca antes atingida. Computadores, comparados com seres humanos, conseguem analisar quantidades maiores de dados com muito mais rapidez, não se fatigam com facilidade, têm muito mais *tolerância* a tarefas repetitivas e são infinitamente menos susceptíveis a erros, desde que bem programados.

Desses avanços tecnológicos advieram novas vertentes dentro da Lingüística, agora amparada por ferramentas computacionais, como a Lingüística Computacional, a Lingüística Informática, a Lingüística Quantitativa, a Estatística Lingüística, a Engenharia da Linguagem, o PLN (Processamento de Linguagem Natural) e a Lingüística de Corpus (Berber Sardinha, 2005:22). Nosso enfoque recai sobre esta última e sobre ela Leech (1992:106) afirma:

> *... computer corpus linguistics (henceforth CCL) defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject. The computer, as a uniquely powerful technological tool, has made this new kind of linguistics possible.* [1]

[1] Em português: A Lingüística de Corpus por computador (a partir de agora, CCL) define não somente uma nova metodologia emergente para o estudo da linguagem, mas também uma nova empreitada de pesquisa, e de fato, uma nova abordagem filosófica ao assunto. O computador, como uma ferramenta poderosa e singular,

Sem a tecnologia que permite a coleta e análise de milhares de textos, contendo centenas de milhões de palavras, a Lingüística de Corpus, como a conhecemos hoje, seria algo impraticável.

1.1.2 O que é corpus?

As definições de corpus abundam na literatura acadêmica. Em linhas gerais, um corpus é uma coletânea suficientemente grande de textos naturais[2] (Sinclair, 1995:171 *apud* Berber Sardinha, 2004:16) usados para descrever a linguagem. Por "natural" entende-se que os textos tiveram sua criação de maneira espontânea, ou seja, não foram criados para serem incluídos no corpus. Pode-se incluir também no conceito de "natural" o fato de terem sido produzidos por seres humanos, não incluindo, portanto, textos criados de maneira eletrônica através de programas geradores de textos.

Berber Sardinha (2004:3) afirma que o uso da palavra corpus data da Grécia Antiga (Corpus Helenístico de Alexandre, o Grande), sendo usado também na Idade Média (corpora[3] de citações da Bíblia). Obviamente, tais corpora não eram eletrônicos e a palavra corpus atinha-se a seu sentido original: um conjunto ou coletânea de documentos sobre determinado tema, conforme o dicionário Houaiss (2004).

Hunston (2002:2) diz que hoje um corpus é definido em termos de sua forma e de seu propósito. Ele tem sua construção planejada, isto é, há critérios para sua elaboração, e também tem o propósito de ser usado para investigação lingüística, e não simplesmente de viabilizar acesso a textos para leitura. Seu propósito é o de permitir a investigação da linguagem nele contida. Portanto, um arquivo (depósito de textos sem organização prévia) ou uma biblioteca eletrônica não

---

tornou esse tipo de Lingüística possível. (Todas as traduções para o português presentes nas citações desta dissertação foram feitas por mim).

[2] Em inglês: *naturally occurring texts*.

[3] O plural de *corpus* é *corpora*.

podem ser chamados de corpus. Um corpus tem um desenho explícito e um propósito específico (Berber Sardinha, 2004:16).

Berber Sardinha (2004:18) cita a definição de corpus de Sanchez (1996:8) como sendo a mais completa por englobar todos os aspectos presentes em sua elaboração e em seus propósitos:

> *Um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos) sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.*

Essa definição inclui vários pontos importantes: origem (dados autênticos), propósito (estudo lingüístico), composição (conteúdo criteriosamente escolhido), formatação (eletrônica), representatividade (capaz de representar uma língua ou variedade) e extensão (suficientemente vasto para ser representativo).

1.1.3 Fundamentos da Lingüística de Corpus

Biber, Conrad & Reppen (1998:1) e Monaghan (1979:5) apontam que a Lingüística tradicionalmente deu ênfase à segmentação e à taxonomia, decompondo a linguagem em unidades menores, classificando-as (fonemas, morfemas, palavras, frases, classes gramaticais) e descrevendo de que maneira tais unidades se combinam para formar unidades maiores. Há, entretanto, uma perspectiva diferente de analisar a linguagem: pode-se centrar o foco da análise em como os falantes exploram os recursos oferecidos pela linguagem. Ao invés de teorizar sobre o que é possível ocorrer em uma língua, estuda-

se o que realmente ocorre, o que realmente é usado pelos falantes. A Lingüística de Corpus insere-se nessa segunda perspectiva.

Sinclair (1991) aponta cinco aspectos presentes em um corpus, que tornam a análise lingüística nele baseada diferente de outros métodos:

a) Os dados são autênticos;

b) Os dados não são pré-selecionados segundo critérios preestabelecidos pelo analista;

c) Há dados em grande quantidade;

d) Os dados estão sistematicamente organizados;

e) Os dados não são classificados conforme as teorias tradicionais, ou seja, de maneira a engessar os resultados, amoldando-os a teorias já existentes, bloqueando a descoberta de novos aspectos da linguagem, que tendem a surgir em pesquisas com corpora (Hunston, 2000:18-19).

A Lingüística de Corpus encontra-se em consonância com os princípios da visão neofirthiana de linguagem, descritos aqui por Stubbs (1993:2):

a) A natureza da Lingüística: ela é essencialmente uma ciência social e uma ciência aplicada com implicações práticas, especialmente na educação;

b) A natureza dos dados: os textos devem ser completos e autênticos; não devem ser sentenças isoladas ou fragmentos de texto; nenhum dado deve ser intuitivamente inventado;

c) O foco principal de estudo da Lingüística: a Lingüística deveria enfocar o sentido; forma e sentido são inseparáveis; léxico e sintaxe são interdependentes;

d) A natureza do comportamento lingüístico: a linguagem é o equilíbrio entre rotina e criação; a linguagem em uso transmite cultura;

Os princípios acima descritos refletem a visão empírica de linguagem, que se opõe diametralmente à visão racionalista de Noam Chomsky (1957), que se fundamenta na intuição do falante nativo e no subjetivismo.

Sampson (2001:2), um dos grandes defensores do empirismo no campo dos estudos lingüísticos, caracteriza a ciência empírica como sendo firmada em elementos que são interpessoalmente observáveis de modo que as diferenças de opinião possam ser resolvidas por meio da arbitragem neutra da experiência objetiva. Ele ainda afirma que, enquanto a ciência se esforçar para se fundamentar em dados interpessoalmente observáveis, ela sempre poderá seguir avante através do diálogo crítico dentro da comunidade de pesquisadores. Porém, conceder autoridade a evidências subjetivas e intuitivas significa podar essa possibilidade de progresso.

McEnery & Wilson (1996:12) afirmam que um corpus tem a vantagem de tornar público o ponto de vista usado para apoiar uma teoria. As observações baseadas em corpora são intrinsecamente mais verificáveis que julgamentos baseados em introspecção. Os autores dizem ainda que a Lingüística de Corpus pode ser descrita em termos simples como o estudo da linguagem baseado em seu uso real[4]. O lingüista de Corpus busca observar grandes porções de linguagem e firmar suas análises nessas observações ao invés de basear-se em suas intuições.

[4] Em inglês: *real-life language use.*

Para a Lingüística de Corpus, a evidência externa, isto é, evidência de uso real, é uma fonte melhor que a evidência interna, ou seja, a intuição do falante nativo (McCarthy, 2001:124). Schmitz (2005:4) afirma:

> *Corpus linguistics has shown native speaker judgments to be wrong in many cases. Native speakers as a group are not always reliable for they do not agree with one another about the grammaticality of sentences.*[5]

Leech (1992:112) dá mais algumas características da Lingüística de Corpus:

a) Falsificabilidade[6]: um modelo baseado em corpus pode ser testado em novas amostras de um outro corpus.

b) Completude: inclui todos os dados do corpus sem prévia seleção.

c) Simplicidade: a Lingüística de Corpus contabiliza os dados do corpus com um conjunto mais parcimonioso de conceitos sobre o domínio em investigação.

d) Força[7]: o autor considera os modelos baseados em corpus "mais fortes"[8] pelo fato de se limitarem firmemente aos dados que estão presentes no modelo, excluindo dados que intuitivamente deveriam figurar no corpus, mas que não estão presentes.

e) Objetividade: os modelos podem ser replicados e testados por observadores ou pesquisadores independentes, inclusive por aqueles que não têm nenhuma ligação emocional com o

[5] Em português: A Lingüística de Corpus já nos mostrou que os julgamentos do falante nativo são errados em muitos casos. Falantes nativos, como um grupo, nem sempre são confiáveis, pois eles não concordam uns com os outros sobre a gramaticalidade das sentenças.
[6] Em inglês: *falsifiability*. Tradução de acordo com Schmitz & Almeida Filho (1998:181). Ver também Popper (1968).
[7] Em inglês: *strength*.
[8] Leech usa a palavra *stronger*, também entre aspas.

sucesso ou fracasso do modelo. O subjetivismo tem muito pouco espaço na Lingüística de Corpus.

1.1.4 Lingüística de Corpus: Metodologia ou Disciplina?

A Lingüística de Corpus ocupa um território incerto na Lingüística Aplicada (McCarthy, 2001:125). A Lingüística de Corpus deve portar qual *status* dentro da Lingüística? De uma área de estudo definida, como a Sociolingüística e a Psicolingüística, ou de uma metodologia que veio para servir as outras áreas de pesquisa? Essas questões têm sido tema de muitos debates entre os praticantes da área.

Granger (2002:4) afirma:

> *Corpus linguistics can best be defined as a linguistic methodology which is founded on the use of electronic collections of naturally occurring texts, viz. corpora. It is neither a new branch of linguistics nor a new theory of language, but the very nature of the evidence it uses makes it a particularly powerful methodology, one which has the potential to change perspectives on language.*[9]

Leech (1992:106) chama a Lingüística de Corpus de um novo empreendimento na Lingüística, de uma nova abordagem filosófica da matéria, de um ”Abre-te, Sésamo” para uma nova maneira de pensar a linguagem.

Hunston (2000:14), com base no trabalho de Sinclair (1991), diz que a Lingüística de Corpus é uma maneira de investigar a linguagem

[9] Em português: A Lingüística de Corpus pode ser mais bem definida como uma metodologia lingüística que está fundada no uso de coleções de textos naturais, ou seja, corpora. Ela não é nem um ramo da Lingüística nem uma nova teoria de linguagem, mas a própria natureza da evidência que ela usa a torna uma metodologia particularmente poderosa, com potencial para mudar as perspectivas sobre a linguagem.

por meio de grandes quantidades de discurso, coletado naturalmente e armazenado eletronicamente, usando programas de computador que selecionam, separam, combinam, contam e calculam.

McEnery & Wilson (1996:2) não consideram a Lingüística de Corpus como sendo um ramo da Lingüística, como a semântica ou a sintaxe, as quais "descrevem/explicam algum aspecto do uso da linguagem". Os autores crêem que a Lingüística de Corpus pode ser descrita mais como uma metodologia do que um aspecto da linguagem que requer descrição ou explicação.

Berber Sardinha (2004:35) também não vê a Lingüística de Corpus como uma disciplina dentro da Lingüística, como o são a Sociolingüística ou a Psicolingüística, que têm seu objeto de pesquisa muito bem delimitado. Contudo, o autor também não a resume a uma simples metodologia, aqui vista como um conjunto de instrumentos, pelo fato de a Lingüística de Corpus ter fundamentos próprios que a norteiam. Além disso, os praticantes de Lingüística de Corpus produzem conhecimento novo, os quais, muitas vezes, divergem das práticas mais comuns no momento.

Hoey (1997) *apud* Berber Sardinha (2004:37) aparece com uma terceira possibilidade: a Lingüística de Corpus não é nem uma disciplina nem uma metodologia, ela pode ser considerada uma abordagem, ou seja, trata-se de uma perspectiva, uma maneira de enxergar a linguagem. Seria como uma janela que molda a visão que temos do mundo exterior à casa. Esta visão de Hoey, também apoiada por Berber Sardinha, é a que adotamos neste trabalho.

1.1.5 Breve Histórico da Lingüística de Corpus

McEnery & Wilson (1996:6) apontam para o fato de existirem registros de estudos no campo de aquisição de L1[10] baseados em

[10] L1 refere-se à primeira língua aprendida pela criança (também chamada de língua-mãe ou língua nativa) ou a língua preferida, quando se trata de indivíduos que moram em países onde se fala mais de uma língua (Crystal, 1997:108).

corpora realizados entre 1876 e 1926 – sem ainda receber o nome de *Lingüística de Corpus*.

Em 1897, Käding usou um corpus de impressionantes 11 milhões de palavras para analisar a seqüência e a freqüência de distribuição das letras do alemão. Para realizar essa tarefa, Käding usou o trabalho de 5 mil analistas (Berber Sardinha, 2004:4).

Nomes, como o do educador Thorndike e os lingüistas Boas e Fries, estão ligados à construção de corpora no início do século XX. Obviamente, tais corpora não eram eletrônicos, tendo sido coletados, mantidos e analisados manualmente. Tal fase também foi caracterizada pelo enfoque no ensino de línguas, contrastando com a Lingüística de Corpus moderna que focaliza mais a descrição de linguagem.

Em 1959, em Londres, Randolph Quirk e sua equipe iniciaram a compilação do SEU (Survey of English Usage), o último grande corpus processado manualmente, o qual serviu de referência para os corpora posteriores no que toca a número de textos e quantidade igual de palavras por texto. Desse trabalho, adveio a famosa *Comprehensive Grammar of the English Language* (Quirk et al., 1985).

A grande crítica aos corpora manuais, como o de Thorndike nos anos 40, com 18 milhões de palavras, era que o processamento de quantidades gigantescas de palavras por meios manuais não podia ser considerado confiável.

Em 1957, com o lançamento de *Syntactic Structures* de Chomsky, a Lingüística entra em um novo paradigma: o racionalismo. Nele, a intuição do falante nativo, a introspecção e o subjetivismo tomaram conta do cenário dos estudos lingüísticos, lançando a visão empirista e objetiva, e a observação num período de trevas. Tal mudança de paradigma obscureceu completamente o lançamento do primeiro corpus eletrônico do mundo em 1964: o Brown University Standard Corpus of

Present-Day American English, mais conhecido como o corpus Brown (Berber Sardinha, 2000:324).

A popularização dos computadores e das ferramentas de processamento nos anos 80 contribuiu decisivamente para o ressurgimento e fortalecimento da pesquisa lingüística baseada em corpus (Berber Sardinha, 2004:5).

Em 1995, concluíram-se os trabalhos do BNC (British National Corpus), o primeiro corpus a romper a barreira dos 100 milhões de palavras. Esse megacorpus histórico está disponível para compra dentro da Comunidade Européia e ainda provê acesso pela Internet[11] a um concordanciador que gera 50 linhas de concordância randomicamente.

Hoje em dia, a Lingüística de Corpus tem grande influência nos estudos lingüísticos, estando os centros mais desenvolvidos situados na Europa, mais especificamente na Grã-Bretanha e Escandinávia.

No Brasil, a Lingüística de Corpus encontra-se ainda em estágio incipiente, sendo o Projeto Direct da Pontifícia Universidade Católica de São Paulo[12] sobre a linguagem do trabalho um dos expoentes na língua portuguesa.

1.1.6 A Lingüística de Corpus no Ensino de Línguas Estrangeiras

Berber Sardinha (2004:254) afirma que a Lingüística de Corpus se insere basicamente em quatro áreas do ensino de línguas:

1. Descrição de língua nativa: ainda de caráter acadêmico e não muito presente em sala de aula devido à distância entre o profissional de ensino e a academia;

[11] Endereço na Internet: http://sara.natcorp.ox.ac.uk/lookup.html
[12] Endereço na Internet: http://lael.pucsp.br/direct

2. Descrição da linguagem do aprendiz: trata-se dos corpora de aprendizes, que contêm a produção de alunos de língua estrangeira. Ainda restrita ao ambiente acadêmico, porém tem tomado bastante impulso;
3. Transposição de metodologia de pesquisa acadêmica para a sala de aula: trazer para sala de aula as concordâncias[13] e listas de palavras[14];
4. Desenvolvimento de materiais de ensino, currículos e abordagens: em termos de métodos e abordagens, podemos citar os três principais: o Currículo Lexical, de John Sinclair (1987), a Abordagem Lexical, de Michael Lewis (1993) e o Ensino Movido a Dados (DDL – Data Driven Learning), de Tim Johns (1994).

Hunston (2002:96) mostra que a Lingüística de Corpus tem revolucionado a elaboração de livros didáticos e dicionários de tal forma que hoje em dia tornou-se inconcebível uma editora publicar um dicionário ou gramática que não tenha suas bases em um corpus. Os materiais didáticos cada vez mais deixam de basear-se na intuição do autor e em linguagem por ele inventada e passam a refletir a linguagem usada na vida real contida num corpus.

1.1.6.1 Corpora de Aprendizes

Em franco crescimento, a área de corpora de aprendizes tem em Sylvianne Granger seu maior expoente na atualidade. Em seu trabalho, Granger (2002:4) assinala que apenas no final dos anos 80 a investigação lingüística baseada em corpus começou a desenvolver um interesse maior na linguagem de aprendizes de língua estrangeira, com a montagem dos primeiros corpora de aprendizes de inglês não-nativo.

[13] Ver seção 1.1.6.3
[14] Ver seção 1.1.6.4

Esse fato criou uma ligação entre esses dois campos anteriormente distantes: a Lingüística de Corpus e a pesquisa sobre aprendizagem de língua estrangeira. Segundo Granger, usando os princípios, ferramentas e métodos da Lingüística de Corpus, consegue-se melhorar a descrição da linguagem do aprendiz respondendo diversas questões sobre aprendizagem de língua estrangeira, tais como qual tipo de aluno tem mais dificuldade em qual ponto no processo de aprendizagem (Granger, 2002:21).

São informações importantes que têm influência na elaboração de material didático, na elaboração de currículo, no processo de formação de professores e no desenvolvimento de novas metodologias para sala de aula.

1.1.6.2 Padronização da Linguagem

Um dos grandes avanços trazidos pela Lingüística de Corpus para o campo do ensino de língua estrangeira foi a descrição da padronização da linguagem, ou seja, das combinações recorrentes entre as palavras.

Lewis (1993:82) descreve a padronização da linguagem como o fato de a ocorrência de certas palavras ou estruturas nos predispor a esperar outros itens lexicais específicos.

As nomenclaturas, entretanto, ainda não estão muito bem definidas na literatura acadêmica. Sobre isso, Succi (2003) afirma:

> *A questão da co-ocorrência de itens lexicais na linguagem vem sendo amplamente discutida e, conjuntamente com a intensidade das pesquisas, encontramos uma profusão de termos para denominar o fenômeno da co-ocorrência de palavras. Dentre estes termos, temos os seguintes, cujas traduções já foram consagradas em português: colocações (collocations),*

> *porções (chunks), multi-palavras (multi-word items), linguagem formulaica (formulaic language) e expressões fixas (fixed expressions). Sem uma tradução consagrada em português encontramos: automatic language, composites, conventionalised forms, formulae, gambits, holophrases, routine formulae, phrasemes, preassembled speech, prefabricated routines and patterns, ready-made utterances, sentence stems (para maiores detalhes sobre a diversidade de nomenclatura, o leitor deve consultar Hunston & Francis, 1999:7 e Wray, 1999:214).*

São exemplos de padronização a colocação, a coligação[15] e a prosódia semântica[16].

1.1.6.2.1 Colocação

Hunston (2002:68) define colocação como "a tendência de duas palavras em co-ocorrer ou como a tendência de uma palavra em atrair uma outra"[17]. Stubbs (1995) a define como "a relação de co-ocorrência habitual entre palavras".

Tagnin (2005:37) descreve a colocação de maneira simples e esclarecedora:

> *... certas palavras parecem combinar-se de forma natural, não havendo, via de regra, explicação para o fato. Em certos casos, as palavras se associam por terem uma ligação na vida real: cão e gato. Entretanto, porque não ocorre cachorro e gato?*

[15] Colocação e coligação são termos introduzidos por Firth (1957).
[16] Prosódia semântica é um termo introduzido por Louw (1993).
[17] Em inglês: the tendency of two words to co-occur, or as the tendency of one word to attract another.

As razões para a existência das colocações, segundo Krishnamurthy (1997:37) está na recorrência de situações similares na vida humana, na economia de esforço e na necessidade de agilizar a conversação. É mais fácil usar algo pronto, convencional, que a maioria das pessoas usa e conhece, do que criar enunciados inéditos a todo momento.

A colocação desempenha um papel importante no ensino de língua e na formação de sentido. Seguindo com os exemplos caninos, Lewis (1993:82) afirma que é quase impossível explicar o sentido de *latir* sem mencionar *cachorro*.

Existem colocações (Tagnin, 2005:38):

- Adjetivas: *Merry Christmas*, *close friend*, *foreign policy*;
- Nominais: *credit card*, *room service*, *phone book*;
- Verbais: *make an impression*, *take pride*, *come into force*;
- Adverbiais: *pay dearly*, *thank profusely*, *take seriously*.

São exemplos de colocações em português: *larga escala*, *redondamente enganado* e *pôr a mesa*.

Conhecer as colocações do inglês ajuda na construção da idiomaticidade dos aprendizes dessa língua. Por idiomático, entendemos como ”típico do modo natural no qual alguém fala ou escreve quando em uso de sua própria língua”[18] (Longman, 2003), ou seja, usar a língua inglesa de maneira mais próxima à efetivamente utilizada por seus falantes nativos em termos de combinações de palavras (Kjellmer, 1992:329; Medgyes, 1994:14). Tagnin (2005:14) refere-se a isso como a escolha das combinações de palavras ”aceita de comum acordo pela comunidade que fala determinada língua”. Em seu artigo sobre o ”falante ingênuo”, o lingüista americano Fillmore (1979:66) postula que

[18] Em inglês: Typical of the natural way in which someone speaks or writes when they are using their own language.

quanto maior o nível de conhecimento sobre os aspectos idiomáticos de uma língua, maior será a fluência daquele que a aprende.

Os dicionários monolíngües para aprendizes (*learner's dictionaries*), como o Longman Dictionary of Contemporary English (Longman, 2003), não mais simplesmente apresentam os significados das palavras, mas também mostram quais são seus colocados, isto é, os vocábulos que normalmente co-ocorrem com a palavra em questão.

Fizemos um pequeno experimento com a palavra *range*[19]. O Longman Dictionary of Contemporary English (Longman, 2003), dicionário monolíngüe baseado em corpus, apresenta as seguintes colocações adjetivas para *range*: *wide range*, *whole range*, *broad range* e *full range*. Não estão presentes colocações como *big range*, a qual poderia soar correta para o falante de português brasileiro ao verter *grande gama* ou *grande variedade* para o inglês. Verificamos as freqüências dessas colocações no British National Corpus (BNC) e comprovamos o baixo uso de *big range*, como apresentado no quadro 1.1 abaixo, confirmando a posição do Longman Dictionary of Contemporary English de não incluir *big range* como uma colocação relevante para o ensino do inglês como língua estrangeira[20].

Colocado	Freqüência no BNC
wide	2.743
whole	659
full	417
broad	159
big	4

Quadro 1.1 – Colocados de range e as freqüências no BNC.

Bahns (1993:108) concluiu em seu artigo *"Should we teach EFL students collocations?"*[21]:

[19] Em português: gama, variedade.
[20] Ver seção 1.4.4 sobre inglês como língua estrangeira.
[21] Em português: Deveríamos ensinar colocações a alunos de inglês como língua estrangeira?

> *It can be concluded from this study that learners are more than twice as likely to select an unacceptable collocate as they are to select an unacceptable general word, and that EFL learners' knowledge of general vocabulary far outstrips their knowledge of collocations".*[22]

Assim, contar com esse tipo de informação auxilia professores e alunos falantes de português brasileiro a conhecer a língua-alvo com mais profundidade, evitando assim colocações não-idiomáticas (Brezolin et al., 2001:5).

1.1.6.2.2 Coligação

O termo coligação refere-se à associação entre itens lexicais e itens gramaticais, como por exemplo a associação existente entre um verbo e uma preposição (*begin + to*): *He began to cry* (Berber Sardinha, 2004:40).

Conforme Tagnin (2005:31), existem os seguintes tipos de coligação: coligações de regência, *phrasal verbs* e coligações prepositivas.

Coligações de regência com:

- Verbos: *congratulate on*, *devote to*, *talk about*;
- Substantivos: *aptitude for*, *expert in*, *remorse for*;
- Adjetivos: *crazy about*, *good at*, *hard on*;
- Advérbios: *because of*, *instead of*, *together with*.

Phrasal verbs:

- *Give in*, *find out*, *bring about*.

[22] Em português: Pode-se concluir desse estudo que a probabilidade de os aprendizes escolherem um colocado inaceitável é duas vezes maior do que a probabilidade de eles escolherem um item lexical inaceitável, e que o conhecimento de vocabulário geral dos aprendizes de inglês como língua estrangeira supera de longe o conhecimento deles sobre colocações".

Coligações prepositivas:

- *At random*, *in accordance with*, *by appointment.*

1.1.6.2.3 Prosódia Semântica

Trata-se de mais um conceito que nos auxilia a aprofundar nosso conhecimento sobre as palavras e suas colocações: a associação entre itens lexicais e sua conotação (positiva, negativa ou neutra). Essa combinação recebe o nome de prosódia semântica (Louw, 1993) ou associação semântica (Hoey, 2003).

São três os tipos de prosódia semântica, de acordo com Berber Sardinha (2004:41):

- Negativa: como por exemplo a palavra *causar*, que quase sempre se associa a palavras negativas, tais como *causar um problema*, *causar um acidente*, *causar um dano*, *causar câncer* e *causar uma crise*;

- Positiva: como por exemplo a palavra *prover*, que normalmente tem colocados de natureza positiva, como *prover ajuda*, *prover assistência*, *prover auxílio* e *prover socorro*;

- Neutra: *prover* também pode apresentar prosódia semântica neutra, como em *prover treinamento*, onde a palavra *treinamento* não tem sentido nem positivo nem negativo.

Há ainda muita divergência entre os autores sobre este conceito e sua nomenclatura, porém *prosódia semântica* é o termo consagrado na Lingüística de Corpus. Para uma revisão da literatura mais pormenorizada sobre prosódia semântica, o leitor pode consultar Nelson (2005).

1.1.6.3 Concordâncias

Outro fruto da Lingüística de Corpus presente no ensino de língua estrangeira são as chamadas concordâncias em formato KWIC (Keyword in Context[23]). A figura 1.1 a seguir mostra uma concordância KWIC retirada de um corpus (Tagnin, 2001) composto por textos sobre mercado financeiro, tendo *price*[24] como nódulo (palavra pesquisada, em posição central na concordância):

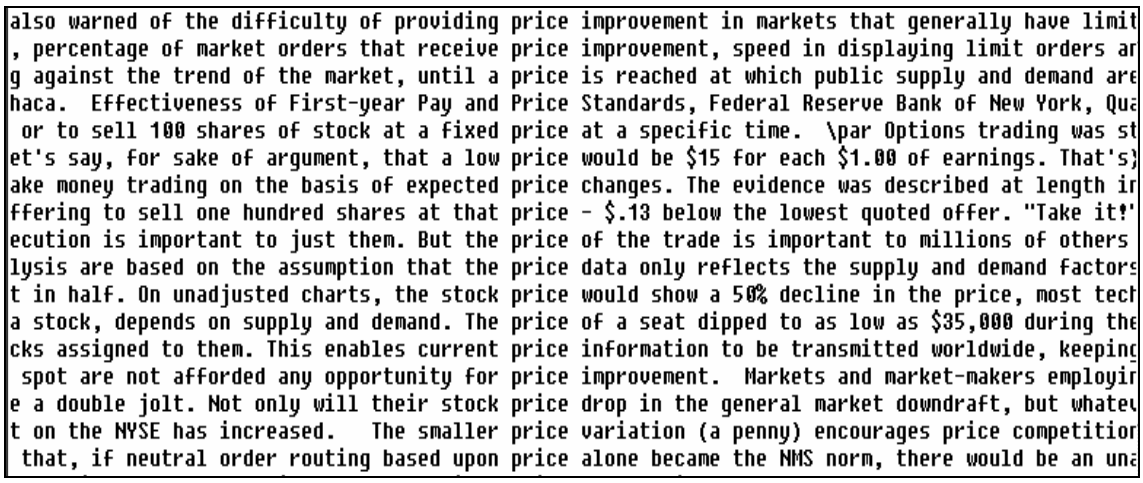

Figura 1.1 – Exemplo de concordância de *price*.

Há grandes vantagens em apresentar a língua-alvo através de concordâncias:

a) Possibilidade de apresentar a língua autêntica, ao invés de textos artificialmente elaborados para as atividades do curso, que poderiam não representar a língua-alvo com propriedade;

b) Ater-se mais ao registro[25] da linguagem que se quer apresentar aos alunos. Isso pode ser obtido através da utilização de corpora com textos de conteúdo acadêmico,

[23] Em português, Keyword in Context quer dizer Palavra-Chave em Contexto.
[24] Em inglês: *preço*.
[25] Na lexicografia, o registro de uma palavra indica em qual situação ela tem seu uso: formal, informal, literário ou técnico (Longman, 2003:xv).

jornalístico ou científico, para exemplificar o registro formal. Para o registro informal, pode-se, por exemplo, usar textos que sejam transcrições de conversações informais, ou de programas televisivos cômicos, ou ainda de livros e revistas que contenham este tipo de linguagem. Ao fazermos isso, adequamos o vocabulário e as estruturas ensinadas ao tipo de linguagem com que se deseja trabalhar sem alterar sua autenticidade;

c) Investigar os aspectos idiomáticos da linguagem (Berber Sardinha, 2004:273). Utilizando concordâncias, o aluno tem acesso não apenas a palavras isoladas, mas a uma amostra da língua em uso, com uma grande parcela de seus possíveis cotextos (palavras adjacentes), e os vários sentidos que a palavra pode assumir de acordo com tais cotextos. Em uma concordância, pode-se analisar o observável, isto é, o que está presente; e o esperado, porém ausente, ou seja, aquilo que se esperava encontrar, mas que por algum motivo não está presente na concordância;

d) Utilizar a léxico-gramática ao invés de vocabulário em um momento e gramática em outro (Berber Sardinha, 2004b). Nas concordâncias, o aprendiz pode entrar em contato com vocabulário novo e, ao mesmo tempo, aprender a gramática envolvida nesse novo vocabulário (preposições, por exemplo). ”The dichotomy grammar/vocabulary is invalid”[26] (Lewis, 1996:vi);

e) Possibilidade de treinar os alunos a observarem e descobrirem a padronização da língua-alvo ao estilo DDL, Data-Driven Learning, de Tim Johns (1994).

[26] Em português: A dicotomia gramática/vocabulário é inválida.

1.1.6.4 A Freqüência de Uso no Ensino de L2[27]

Outro fator oriundo das pesquisas baseadas em corpora é o estudo das freqüências de uso dos itens lexicais de uma língua. Com facilidade, as ferramentas computacionais da Lingüística de Corpus podem analisar um corpus e gerar uma *wordlist*, ou seja, uma lista com todas as palavras contidas no corpus e suas respectivas freqüências de uso. Por freqüência de uso, entende-se o número de vezes que a palavra apareceu no corpus. O analista de corpus então interpreta essa lista de palavras à luz de seus propósitos, verificando, por exemplo, quais são os itens lexicais mais freqüentes, os menos freqüentes e os esperados, porém, ausentes.

McEnery & Wilson (1996:12) vêem a freqüência de uso de uma palavra ou de um construto como um fator importante na descrição da linguagem e afirmam que os seres humanos têm uma vaga noção da freqüência, mas a observação natural dos dados por meio de corpora parece ser a única fonte confiável para a análise dessa característica da linguagem.

Existem dois tipos básicos de contagem de palavras em um corpus: a contagem de *types* e a contagem de *tokens*. O número de *types* (também chamado de *forma*, *palavra*, *vocábulo* ou *tipo*) de um corpus relaciona-se ao número de palavras diferentes nele contidas. O número de *tokens* (também chamado de *itens* ou *ocorrências*) refere-se ao número total de palavras de um corpus, ainda que repetidas. Por exemplo, na frase *"A menina comeu a torta"*, existem cinco *tokens*, porém apenas quatro *types*, visto que o artigo *a* aparece duas vezes (Berber Sardinha, 2004:165; Mona Baker, 1995:236). A freqüência de uso está relacionada à contagem de *tokens* de um corpus.

Sobre a freqüência de uso no ensino de língua estrangeira, Granger (2002:22) e Sökmen (1997:239-240) dizem:

[27] L2 significa uma outra língua que não seja a língua-mãe (L1) de um indivíduo.

> *In the field of vocabulary teaching, for instance, specialists are in agreement that both frequency and difficulty have to be taken into account. This comes out clearly in Sökmen's (1997:239-240) survey of current trends in vocabulary teaching: "Difficult words need attention as well. Because students will avoid words which are difficult in meaning, in pronunciation, or in use, preferring words which can be generalized (...), lessons must be designed to tackle the tricky, less frequent words along with the highly-frequent. Focusing on words which will cause confusion, e.g. false cognates, and presenting them with an eye to clearing up confusion is also time well-spent".*[28]

Fox (1998:26) mostra que a freqüência de uso não é o único critério para a seleção do que ensinar. Contudo, ela é uma variável de grande importância. Informação sobre freqüência de uso permite ao professor focalizar as palavras mais importantes, assegurando que os alunos saibam efetivamente como usá-las. Fox também aponta para a importância de observar as palavras que são infreqüentes, pois essas, em linhas gerais, merecem menos atenção no processo de ensino e aprendizagem. Palavras infreqüentes têm uso muito relacionado a um tópico específico e precisariam receber mais atenção e serem incluídas no processo de ensino e aprendizagem apenas quando forem necessárias para desenvolver alguma tarefa que envolva um vocabulário diferenciado ou técnico.

[28] Em português: No campo de ensino de vocabulário, por exemplo, os especialistas estão de acordo que tanto a freqüência quanto a dificuldade têm de ser levadas em conta. Isso aparece claramente na pesquisa de Sökmen (1997:239-240) sobre as tendências atuais no ensino de vocabulário: "Palavras difíceis precisam de atenção também, porque os alunos evitarão palavras que são difíceis em termos de sentido, pronúncia ou uso, preferindo palavras que possam ser generalizadas (...), as lições devem ser elaboradas para dar conta das palavras complicadas e menos freqüentes junto com as altamente freqüentes. Focalizando em palavras que causarão confusão, tais como os falsos cognatos, e apresentando-as com vistas a esclarecer a confusão é também tempo bem gasto."

1.2 RELAÇÃO ENTRE FALA E ESCRITA

A seguir, expomos mais alguns aspectos teóricos sobre os quais nos apoiamos para abordar as questões sobre a fala e a escrita abordadas neste trabalho.

1.2.1 Definições

Antes de prosseguirmos, é preciso definir alguns termos-chave. Normalmente, termos como escrita, sistema de escrita e ortografia são usados sem muita especificidade (Coulmas, 2000:37). Seguem as definições destes e outros termos.

a) Escrita: refere-se à gravação de marcas gráficas de relação convencional com a linguagem em uma superfície durável, com o propósito de comunicar algo (Coulmas, 2000:17), ou de fixar, imobilizar a linguagem articulada, por essência fugidia (Higounet, 2003:9);

b) Sistema de escrita: Coulmas (2000:17) o define como sendo um sistema que descreve as unidades lingüísticas de diferentes níveis estruturais (palavras, sílabas e fonemas). Morais (1995:75) aponta que o sistema de escrita se caracteriza pelo nível de estrutura de linguagem por ele representado. Assim, por exemplo, o sistema logográfico representa a linguagem no nível da palavra e o alfabético, no nível do fonema;

c) Tipo de escrita (em inglês, *script*): refere-se às instâncias gráficas do sistema de escrita. Comumente, generalizamos o *script* e o chamamos de ”alfabeto”. Podemos citar como tipos de escrita o alfabeto romano, o

alfabeto grego e o alfabeto cirílico, os quais são usados na escrita de diferentes línguas;

d) Grafema ≠ Letra: Morais (1995:76) define grafema como ”todos os grupos de letras que podem ser lidas como um único fonema.” É a unidade mínima da escrita. O grafema é uma unidade abstrata; a letra, por sua vez, é a materialização do grafema. O grafema <e>, em português, pode ser expresso por letras de diferentes formas, tamanhos, estilos, efeitos e cores: E, *E*, ***e***, ε, **e** etc. Scliar-Cabral (2003:27) diz que:

> *... deve-se entender o grafema como uma ou mais letras que representam um fonema (no sistema alfabético do português do Brasil, não mais que duas letras). Por exemplo, em ”nasce” temos cinco letras e quatro grafemas para representar /ˈnɑsi/. No caso, o grafema ”sc” é um dígrafo.”*

Neste trabalho, ao referir-nos a grafemas, usaremos os sinais de menor (<) e maior (>) e letra minúscula, conforme Crystal (1997:257). Utilizamos também o hífen, como usado no trabalho de Venezky (1970), para mostrar a posição dos grafemas dentro da palavra: <-or> significa grafemas <or> em posição final, <or->, em posição inicial, e <-or->, em qualquer posição;

e) Palavra: neste trabalho adotamos a noção ortográfica de palavra, isto é, uma unidade morfológica separada por dois espaços das outras unidades morfológicas quando escritas. Linell (1982:83) *apud* Coulmas (2000:40) afirma que palavra é uma ”seqüência de letras cercada

por espaços vazios sem conter espaços vazios internos"[29];

f) Fonema: menor unidade distintiva do sistema sonoro de uma língua. Assim, *name* se distingue de *fame* pelo fonema inicial, pela oposição /n/ x /f/. É uma abstração do conjunto de alofones[30], que são suas diferentes realizações ou pronúncias. Os alofones [t^h] de *tent*, [t] de *stay* e [ɾ] de *better* (na pronúncia americana) são alofones do fonema /t/.

g) Fonotática: estudo da distribuição dos fonemas em seqüências e grupos. É o que informalmente se denomina o "cevecê" (CVC) da língua, ou seja, como as consoantes (C) e as vogais (V) se combinam na formação de sílabas. A seqüência de fonemas /st/, por exemplo, pode ocorrer em qualquer posição no inglês: posição inicial (*state*), medial (*posture*) ou final (*latest*). Em português, apenas em posição medial com os fonemas em sílabas diferentes: *es-ta-do, bas-ti-dor* etc (Steinberg, 1985:74);

h) Ortografia: refere-se às regras aplicadas ao uso do *script*. A ortografia é específica à língua com a qual se relaciona. Coulmas (2000:37-39) assinala:

> *Orthographies are always language specific ... Every orthography makes a specific selection of the possibilities of a*

[29] Em inglês: a sequence of letters surrounded by empty spaces but containing no internal spaces.
[30] Para mais referências sobre esse assunto, consultar a seção 1.4.1 Fonética e Fonologia.

script for writing a particular language in a uniform and standardized way.[31]

Como as línguas, as ortografias estão sujeitas a mudanças históricas e geográficas. Pode-se, portanto, falar de diferentes ortografias dentro de uma mesma língua, como por exemplo, a ortografia do inglês britânico e a ortografia do inglês americano.

A ortografia também pode ser classificada entre profunda ou superficial. Uma língua tem uma ortografia profunda quanto mais se distancia do princípio alfabético, ou seja, quanto maior for a distância entre a forma sonora e a forma escrita das palavras, como é o caso do inglês e do francês. Por outro lado, o castelhano, o português, o alemão e o italiano apresentam uma ortografia superficial, visto que a forma oral e a escrita estão bem próximas (Llol, 1999:70).

1.2.2 Diferenças entre Fala e Escrita

Morais (1995:43) mostra a diferença de idade entre a fala e a escrita:

> *"Não se sabe exatamente desde quando os homens falam. Há 30 mil anos, pelo menos, sob uma forma bastante próxima da comunicação lingüística atual. Sob formas mais primitivas, certamente há muito mais tempo... Comparada à linguagem falada, a linguagem escrita é uma aquisição muito recente. Os primeiros traços de escrita têm apenas seis mil anos... Essa diferença de idade entre a linguagem escrita e a falada é uma das características pela qual esses dois modos de comunicação se opõem de maneira evidente."*

[31] Em português: As ortografias são sempre específicas à língua... Toda ortografia faz uma seleção específica das possibilidades oferecidas por um tipo de escrita para escrever uma língua em particular de modo uniforme e padronizado.

A escrita teve sua origem na contabilidade: escravos, empregados, cabeças de gado e sacos de grãos eram contados em placas de argila. Porém, ainda não se tratava de escrita que representasse a linguagem oracional, a qual estima-se que tenha surgido há três ou quatro mil anos.

Hoje, contudo, ainda existem comunidades que não possuem escrita, chamadas de comunidades ágrafas (Mori, 2004:150; Steinberg, 1985:61). Entretanto, não há registro algum, em parte alguma do planeta, de comunidades formadas por indivíduos que não falem. Isso nos mostra que se trata de dois sistemas de comunicação distintos, que não nasceram juntos.

Os dois sistemas têm origens diferentes no homem: a fala é espontânea, sua predisposição é inata, ou seja, o indivíduo, sem comprometimento perceptual ou neuromotor, pode desenvolver-se por si mesmo, com a condição de haver traços de humanização ao seu redor. O ser humano está biopsiquicamente programado para falar (Scliar-Cabral, 2003:53; Luria, 2001:169). Existe uma ”compulsão natural que cada bebê normal tem, desde que participante da interação lingüística, para adquirir a variedade oral de uma ou mais línguas” (Scliar-Cabral, 2003:20). A escrita, por sua vez, aparece como algo artificial, sua origem é completamente externa ao indivíduo e precisa ser adquirida (Vygotsky, 2000:119). Saussure (2001:33) chama a escrita de "estranha ao sistema interno". A escrita depende de treinamento artificial e específico, em outras palavras, de escola.

A língua falada usa imagens acústicas (sonoras) como signos, que Saussure chama de significante; e um conceito, uma idéia, por ele chamado de significado. A língua escrita, por sua vez, usa imagens gráficas como significante. Tanto os signos da fala como os da escrita, considerando a escrita alfabética, são arbitrários (sem semelhança física com o objeto), lineares (numa cadeia sucessiva) e institucionalizados

(”membros de uma mesma comunidade atribuem os mesmos valores às unidades que estão sendo processadas” – Scliar-Cabral, 2003:29).

A língua falada apresenta vantagens sobre a comunicação através da escrita. A principal é que ela permite a utilização de meios não-verbais no momento da comunicação, tais como gestos, mímica e expressão facial (Luria, 2001:169, 171). A língua escrita lança mão de sinais de pontuação, os quais conseguem substituir tais meios não-verbais apenas parcialmente (Olson, 1994:91).

A fala comporta-se de maneira volátil, desaparece no ar. A escrita, por sua vez, constitui um objeto permanente e sólido, passando sua mensagem ao longo do tempo, servindo como, não o único, mas o principal meio de transmissão cultural (Scliar-Cabral, 2003:33). Através dela desenvolvemos idéias, articulamos pensamentos e expomos opiniões. "É sempre possível reler aquilo que foi escrito, quer dizer, voltar voluntariamente a todos os elementos que estão incluídos no texto, o que é completamente impossível na linguagem oral" (Luria, 2001:169).

Podemos dizer, portanto, que existem duas línguas distintas: a língua falada e a língua escrita, ambas inseridas na linguagem verbal (Santaella, 1983:10) – ”verbal” em oposição a outros tipos de linguagens estudadas pela Semiótica, como por exemplo a linguagem corporal de um artista.

1.2.3 Sistemas de Escrita

A seguir, apresentamos os sistemas de escrita com base em Morais (1995:48):

a) Pictográfico: é o sistema mais primitivo no qual um objeto é representado por desenhos que buscam retratá-lo o mais fielmente possível. Este sistema de escrita representa diretamente o mundo:

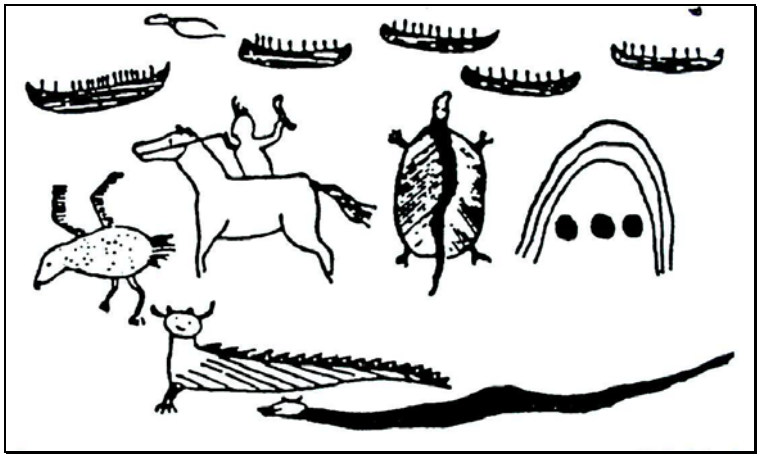

Figura 1.2 – Desenho indígena em rocha nos EUA (Schoolcraft, 1851).

b) Ideográfico: os ideogramas representam uma idéia, como os exemplos abaixo, representando a palavra ”não”, ”pinheiro”, ”madeira” e ”beleza”. Esse sistema representa um salto em direção à arbitrariedade.

Figura 1.3 – Ideograma chinês para a palavra "não" (Morais, 1995:53).

Figura 1.4 – Ideograma chinês para "pinheiro" (à esquerda), formado a partir dos elementos semânticos "madeira" (no meio) e "beleza" (à direita) (Morais, 1995:53).

A seguir vêm dois sistemas de escrita que comportam informação sobre a maneira como a palavra deve ser pronunciada. Trata-se dos sistemas de escrita fonográficos:

c) Silábico: cujos signos representam uma sílaba. São encontrados na escrita suméria ou nos silabários da escrita japonesa. Assim, por exemplo, numa escrita silábica, podem existir cinco símbolos para as sílabas que se iniciam com /m/: um símbolo para a sílaba *ma*, outro para a sílaba *me*, e assim por diante. A palavra *mimo* seria escrita com apenas dois símbolos.

d) Alfabético: sistema que busca representar a língua no nível fonemático. O sistema alfabético constitui um sistema altamente analítico, no qual seus signos gráficos representam a língua falada na segunda articulação: no nível do fonema (Scliar-Cabral, 2003:37)[32].

Uma condição *sine qua non*, portanto, para um indivíduo dominar o sistema alfabético é a capacidade de segmentar a fala em fonemas para poder representá-los através de grafemas (Scliar-Cabral, 2003:50). Morais (1995:88) afirma:

> *Sem receber uma instrução sobre o código alfabético, a criança não descobre os fonemas ... Aprender a utilizar o código alfabético é, ao mesmo tempo, aprender a encontrar os correspondentes fonêmicos das letras, o que implica poder analisar conscientemente a fala em fonemas, e aprender a fundir os fonemas sucessivos.*

"Fundir fonemas sucessivos" recebe o nome de coarticulação. Ao pronunciarmos uma palavra, não pronunciamos os fonemas de maneira isolada. Ao pronunciarmos "chave", temos 5 letras, 4 grafemas, representando 4 fonemas, coarticulados em 2 sílabas.

Essa relação grafema-fonema tende a complicar a aprendizagem, pois levar o indivíduo a alcançar a consciência dos fonemas e sua relação com os grafemas não constitui um processo natural. Os estudos têm mostrado que a consciência silábica, por sua vez, constitui algo mais natural. Um estudo de Morais (1995:89) com cantores poetas portugueses iletrados mostra claramente que eles possuem a consciência silábica. Ao serem testados em sua habilidade para segmentar palavras em sílabas, ficava provado que, mesmo sem terem ido a uma escola e aprendido a ler e a escrever, eles eram capazes de fazer as segmentações em sílabas propostas pelo teste. Na realidade,

[32] A primeira articulação refere-se aos morfemas e a segunda, aos fonemas, conforme a teoria da dupla articulação proposta por Martinet (1971)

essa capacidade já estava aparente nas rimas cantadas em suas poesias. Porém, ao serem testados em relação à segmentação de sílabas em fonemas, tal habilidade se reduziu drasticamente.

Ao tocarmos na questão do fonema, há alguns autores que tendem a dar uma visão mais inatista à consciência fonêmica. Em nossa opinião, o que é inato ao ser humano é apenas o potencial para segmentar a fala em unidades fonêmicas, o qual é ativado através da instrução. Não concordamos que a consciência fonêmica em si já esteja presente na mente dos indivíduos, como Morais (1995:78) deixa transparecer: ''o fonema é uma entidade bem escondida no nosso inconsciente cognitivo".

A conclusão a que chegamos em relação à língua falada e à escrita é a de que há mais pontos de divergência entre os dois sistemas que de convergência. Excetuando-se o fato de ambos serem um sistema de signos que podem representar os mesmos objetos – a palavra "cadeira", dita ou escrita, refere-se ao mesmo objeto – todas as outras variáveis, tais como origens, aplicações, aquisição, aprendizagem e prestígio, tendem a ser divergentes.

Podemos dizer que a escrita seria uma outra língua que um indivíduo aprende após ter adquirido a língua falada. São sistemas que partem do mesmo ponto (porque a escrita alfabética inicialmente se apóia na fala para estabelecer a relação grafema-fonema) e que, num segundo momento, rumam em direções diferentes, cada qual com suas especificidades. A escrita exige um léxico, uma estrutura sintática, uma coesão e coerência diferentes da fala. Contudo, é bom frisar que não se trata de sistemas antagônicos; eles interagem e se completam, possibilitando a comunicação entre os indivíduos.

1.3 CORRESPONDÊNCIA GRAFOFONÊMICA

Os avanços tecnológicos na área da computação também estão possibilitando maior aprofundamento nas pesquisas sobre a

correspondência entre grafema e fonema. Tal estudo recebe o nome em inglês de *graphophonics*, *graphophonemics* ou apenas *phonics*.

No Brasil, esta área é normalmente chamada de *decodificação grafofonêmica* (Capovilla et al., 2001) ou de *correspondência grafofonêmica* (Scliar-Cabral, 2003:20; Schirmer et al., 2004). Percebe-se a preferência por usar a palavra *grafofonêmica* como adjetivo, e não como substantivo. Nos países de língua hispânica, o termo mais utilizado é o substantivo *grafofonética* (Ferreiro, 1988).

Trata-se de uma área inserida na Psicologia, porém não apenas psicólogos, mas também lingüistas e educadores pesquisam a ortografia do inglês e sua complexa relação com a pronúncia.

Resolvemos adotar neste trabalho o termo *correspondência grafofonêmica* por soar mais ligado à Lingüística que o termo *decodificação grafofonêmica*, o qual traz à mente um enfoque maior nos processos mentais e desenvolvimento de estratégias de conversão adotados pelos usuários de determinada língua, não se alinhando com o enfoque lingüístico deste trabalho.

Existem duas direções no estudo da correspondência grafofonêmica (Kiran, Tuchtenhagen & Spelman, 2003):

1. Podemos partir da forma sonora (oral) e transpô-la para a forma escrita (visual). Em inglês, esse estudo recebe comumente nomes como *from sound to spelling*, *phoneme-grapheme correspondence*, *phoneme to grapheme conversion* ou simplesmente *spelling*.

2. Podemos ainda ir em sentido contrário, e partir da forma escrita, estudando suas correspondências sonoras. Esse estudo recebe nomes em inglês como *from spelling to sound*, *grapheme-phoneme correspondence*, *grapheme to phoneme conversion* ou simplesmente *reading*.

O presente trabalho insere-se nesta segunda modalidade de pesquisa.

1.3.1 Combinações Intrassílabicas

Ainda há uma grande indefinição sobre os limites da sílaba. Porém, já existe algum consenso sobre quais são seus constituintes: ataque (*onset*, em inglês*)*, núcleo (*nucleus*, em inglês) e coda. O núcleo constitui o elemento de maior sonoridade da sílaba, por isso, na maior parte das vezes é uma vogal. Entretanto, pode haver uma língua em que o núcleo seja uma consoante. No caso do inglês e do português, o núcleo é sempre uma vogal. Ataque refere-se à consoante que precede o núcleo; e coda, à consoante que o sucede. Ao conjunto núcleo + coda dá-se o nome de rima. Assim, se tomarmos a palavra *cap* (boné, em inglês) como exemplo, teremos a estrutura exibida na figura 1.5 (Kessler & Treiman, 1997:297).

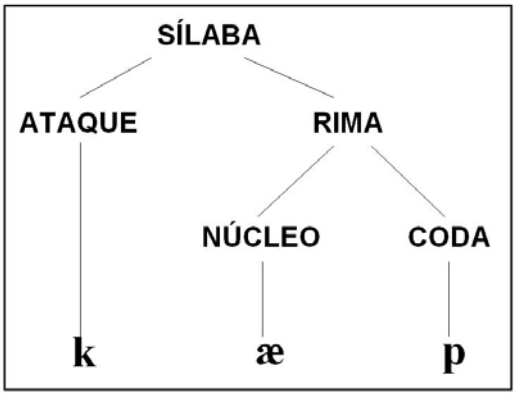

Figura 1.5 - Estrutura típica da sílaba em inglês.

Alguns autores, como Kessler & Treiman referem-se ao conjunto ataque + núcleo como *body*, porém ainda não se trata de uma nomenclatura consagrada.

Já desde os estudos de Venezky (1970), corroborados por outros trabalhos, tais como Kessler & Treiman (1997, 2001) e Connelly (2002), observa-se que, em se tratando de correspondência grafofonêmica, há uma associação mais forte entre os grafemas dentro da rima do que no

conjunto formado pelo ataque e núcleo. A coda determina a pronúncia do núcleo com freqüência muito maior do que o ataque. O núcleo é, portanto, freqüentemente desambiguado pela consoante posterior do que pela anterior.

O ataque não tem associação significativa com a coda em termos de estratégias de conversão grafofonêmica. Apenas partes adjacentes da sílaba influenciam umas as outras significativamente.

1.3.2 Consistência

Dizemos que uma seqüência de grafemas é consistente, se ela exibir regularidade grafofonêmica. A consistência diminui conforme aumenta o número de pronúncias diferentes para a mesma seqüência de grafemas. Ela também diminui quanto mais equiprováveis forem essas pronúncias (Kessler & Treiman, 2001:594).

Em inglês, a parte mais inconsistente da sílaba, e a que mais recebe influência das partes a ela adjacentes (ataque e coda), é o núcleo, ou seja, a vogal. Historicamente, a vogal sofreu muito mais mudanças de pronúncia do que as consoantes[33]. Das mudanças de pronúncia de vogal, condicionadas por uma consoante, listadas por Welna (1978) *apud* Kessler & Treiman (2001:612), 22 foram condicionadas apenas pela coda, 1 apenas pelo ataque e 2 pelos dois em conjunto.

Essa inconsistência geralmente conduz a erro nativos e não-nativos ao lerem uma palavra em inglês, usando apenas o conhecimento da correspondência grafofonêmica dos grafemas individuais da palavra.

Kessler & Treiman (2001:592) mostram algumas razões para a inconsistência da correspondência grafofonêmica no inglês:

[33] Ver seção 1.5.2

1. Manter a grafia de morfemas mesmo quando eles mudam de pronúncia, como por exemplo *heal* e *health*;
2. Diferenciar homófonos: *broach* e *brooch*;
3. Ecoar a ortografia da língua da qual a palavra foi tomada emprestada: *stein* do alemão e *nymph* do grego;
4. Manter a concordância com o uso passado, como em *write*, onde o <w> costumava ser pronunciado.

A maioria dos autores da área concorda com a necessidade de utilizar meios estatísticos para pesquisa de consistência da correspondência grafofonêmica. Segundo Kessler & Treiman (2001:594), a era dos estudos computadorizados em grande escala sobre vocabulário começou com o trabalho sobre correspondência fonema-grafema de Hanna et al. (1966). Todavia, nem o trabalho de Hanna nem os trabalhos que a seguiram – como, por exemplo, o de Venezky (1970), ou o de Brown (1988) sobre o Functional Load, que analisa o léxico através de pares mínimos[34] – envolveram freqüência de uso na língua, fato que teria dado um caráter empírico aos resultados.

1.3.3 Estratégias de Conversão Grafema-Fonema

Devido à alta irregularidade do inglês, teóricos tendem a concordar que existe um léxico mental que é acessado durante a leitura de um vocábulo, ao invés de fazer a conversão grafema a grafema. Todavia, estudos mais recentes (Kessler & Treiman, 2001; Treiman et al., 2002) têm revelado que tais indivíduos também são sensíveis a certa padronização entre o núcleo e a coda, isto é, a rima.

Wimmer & Goswami (1994) compararam as estratégias de conversão grafema-fonema usadas por crianças falantes de inglês, uma língua de ortografia profunda, com as usadas por crianças falantes de

[34] Pares mínimos (*minimal pairs*, em inglês) são pares de palavras que diferem em apenas um fonema: *ship* e *sheep*, *bat* e *bet*, *fit* e *feet* (Laver, 1995:36, Kreidler, 1999:10).

alemão, língua de ortografia superficial. *Grosso modo*, as crianças alemãs pareciam construir a pronúncia convertendo grafema a grafema; enquanto as inglesas pareciam lançar mão de uma estratégia de reconhecimento mais direta, envolvendo memorização de palavras inteiras. Isso ficou patente ao testá-las com logatomas[35] (palavras que não têm sentido, *nonwords* ou *nonsense words*, em inglês), onde as crianças que usavam estratégia de conversão grafema a grafema apresentaram habilidade muito maior para ler tais logatomas do que aquelas que se valiam da abordagem direta.

Vários autores (Prator & Robinett, 1985:219; Laver, 1995:37; Kessler & Treiman, 2001:592; Scliar-Cabral, 2003:53;), entretanto, sugerem um meio termo entre estas duas estratégias: a correspondência grafofonêmica seria condicionada pelo contexto (*context*). Contexto refere-se ao grafema (ou grafemas) à direita e/ou à esquerda do núcleo, na grande maioria dos casos, dentro da mesma sílaba (intrassilábico). Treiman et al. (2002:465) afirmam:

> *Good spellers at the college level are more sensitive than poor spellers to the contextual factors influencing vowel representation.*[36]

Isso quer dizer que as decisões de pronúncia não seriam tomadas isoladamente nem no nível do fonema nem no nível da palavra, mas no nível da sílaba, envolvendo o contexto.

1.4 ENSINO DA PRONÚNCIA DO INGLÊS COMO LÍNGUA ESTRANGEIRA

A seguir, descrevemos os conceitos teóricos relacionados à área de ensino da pronúncia do inglês como língua estrangeira.

---

[35] Para saber mais sobre logatomas ver Gama-Rossi (2004).

[36] Em português: Pessoas de nível universitário que tem boa ortografia são mais sensíveis aos fatores contextuais que influenciam a representação da vogal que pessoas que têm uma ortografia ruim.

1.4.1 Fonética e Fonologia

A Fonética e a Fonologia são as áreas que estudam os sons da fala. Por terem o mesmo objeto de estudo são ciências relacionadas. Contudo, esse mesmo objeto é observado de pontos de vista diferentes em cada caso (Massini-Cagliari & Cagliari, 2004:105).

O termo *Fonética* é usado desde o século XIX para designar o estudo dos sons da voz humana, examinando as suas propriedades físicas independentemente de seu papel lingüístico de construir formas da língua.

A Fonética divide-se em três áreas:

a) Fonética Articulatória: descreve os sons da língua estudando a produção dos signos pelo aparelho fonador do remetente;

b) Fonética Auditiva: descreve os sons da língua observando os efeitos que eles produzem no ouvido do destinatário dos signos;

c) Fonética Acústica: descreve os sons da língua estudando as propriedades físicas das ondas sonoras que se propagam do remetente ao destinatário.

A Fonologia, por sua vez, busca interpretar os resultados obtidos por meio da descrição fonética dos sons da fala, em função dos sistemas de sons das línguas e dos modelos teóricos disponíveis. Faz parte do trabalho fonológico, por exemplo, explicar porque os falantes brasileiros de algumas variantes do português do Brasil consideram como sendo ”o mesmo som” as consoantes iniciais das palavras *tapa* e *tia* ([t] e [tʃ], respectivamente), embora elas sejam bastante diferentes articulatória e perceptualmente. A Fonologia, também chamada de Fonêmica pelos americanos, foi estabelecida a partir da segunda década do século XX, na Europa com o Círculo Lingüístico de Praga e, nos

Estados Unidos com a obra de Leonard Bloomfield e Edward Sapir (Lopes, 1987:97).

Assim, a Fonética é uma ciência de caráter mais descritivo, analisando os sons da fala do ponto de vista de sua produção, percepção e transmissão, ao passo que a Fonologia tem um caráter mais explicativo, interpretativo, buscando o valor dos sons na língua (Massini-Cagliari & Cagliari, 2004:106).

A Fonética pode ser considerada como a ciência do aspecto material dos sons da linguagem humana, estudando seus aspectos físicos, as bases acústicas relacionadas à percepção e bases fisiológicas relacionadas à produção. A Fonologia busca relacionar seus estudos à função que os sons cumprem numa língua específica.

Os sons da fala podem ser descritos, tomando como base três aspectos:

a) Composição

b) Distribuição

c) Função

A Fonética ocupa-se do item a) e a Fonologia, dos itens b) e c) (Lopes 1987:97).

A unidade de estudo da Fonética é o fone, que é transcrito entre colchetes: [p], [t], [k] etc. A unidade de estudo da Fonologia é o fonema, transcrito entre barras inclinadas para a direita: /p/, /t/, /k/ etc (Mori, 2004:145).

A divisão entre Fonética e Fonologia, contudo, não é um consenso dentro da Lingüística. Lopes (1987:98) já chamava a atenção para falta de total acordo sobre a área coberta por ambas as disciplinas. Ainda

hoje, Picasso (2005:25) afirma que ”muitos defendem que ambas as áreas deveriam ser tratadas como uma só”.

Mori (2004:150) comenta sobre a divisão Fonética-Fonologia:

> *Por exemplo, pretender descrever a fonologia de uma língua indígena falada no Brasil sem considerar o aspecto fonético seria absurdo. Do mesmo modo, o estudo da fonética de uma língua, qualquer que seja, resulta pouco proveitoso, de alcance limitado, se não se considera a função que os segmentos fônicos desempenham no sistema dessa língua.*

Mori (2004:150) também aponta para a proximidade entre a Fonologia e o sistema ortográfico de uma língua, ressaltando a importância de o professor conhecer o sistema fonológico da língua para poder explicar as questões oriundas da ortografia.

Nosso trabalho está no campo da Fonologia, estudando as relações entre os grafemas e os fonemas da língua inglesa.

1.4.2 A Pronúncia do Inglês e os Professores Não-Nativos

Se observarmos quais são as pessoas que necessitam comunicar-se em inglês (como língua estrangeira), encontraremos homens de negócios (Celce-Murcia & Goodwin, 1991:137), cientistas, tecnólogos, professores universitários e membros da comunidade acadêmica, entre outros (Morley, 1991:492) e, também o foco de nosso interesse, professores não-nativos de inglês que desejam servir de modelo para seus alunos (Celce-Murcia, Brinton & Goodwin, 1996:8).

Para a maioria dos professores não-nativos, ter um domínio deficitário do inglês pode ser motivo de constante desânimo e complexo

de inferioridade. O autor húngaro Peter Medgyes (1994:15) descreve essa situação:

> *... compared to native speakers, they do less well in every aspect of language performance, as a rule. This feeling of underachievement is particularly excruciating when their performance is compared to that of native speakers with similar variables in terms of age, sex, education, intelligence and especially profession. Let me mention in passing that we non-native English speaking teachers go through this painful experience day in, day out.*[37]

Medgyes (1994:36) conduziu um estudo envolvendo 216 professores de inglês não-nativos de 10 nacionalidades, incluindo 21 brasileiros. Os professores responderam a perguntas do tipo ”Quais são suas principais dificuldades ao usar inglês?” ou ”Suas dificuldades o atrapalham em seu trabalho?”. Os resultados desse estudo colocaram a área de pronúncia como a terceira área que mais afeta negativamente o desempenho dos professores, atrás apenas de vocabulário e fluência, primeiro e segundo lugares, respectivamente. Em contrapartida, a pronúncia também ficou em último lugar como a área na qual os professores percebem o menor progresso.

Quando erros de pronúncia ocorrem, abre-se espaço para um sentimento de incompetência de minha parte, o professor, para com meus alunos e uma sensação de que eu, como professor, não estou provendo um bom modelo nem provendo informações corretas sobre a língua-alvo (Agard, 1969:5). Estamos falhando em auxiliar os alunos a atingir as metas que Morley (1991:500) *apud* Schmitz (2003) apresenta

[37] Em português: ... comparado com falantes nativos, eles não se saem tão bem em cada aspecto de desempenho lingüístico, de modo geral. Esse sentimento de insucesso é especialmente excruciante ao comparar seu desempenho com o de falantes nativos com variáveis similares em termos de idade, sexo, escolaridade, inteligência e, principalmente, profissão. Deixe-me dizer, de passagem, que nós, professores não-nativos de inglês, passamos por essa experiência dolorosa todo santo dia.

como sendo quatro metas razoáveis e desejáveis para os aprendizes de inglês como língua estrangeira:

a) Inteligibilidade funcional: a intenção é auxiliar os aprendizes a desenvolver um inglês oral que seja (pelo menos) razoavelmente fácil de entender e que não desvie a atenção do ouvinte da mensagem.

b) Comunicabilidade funcional: o objetivo é ajudar o aprendiz a desenvolver um inglês oral que preencha completamente as necessidades do aprendiz de ter um sentimento de competência comunicativa.

c) Autoconfiança crescente: a intenção é auxiliar o aprendiz a sentir-se confortável e confiante ao usar o inglês oral, e ajudá-lo a desenvolver uma auto-imagem positiva como falante não-nativo de inglês e a ter um sentimento crescente de apropriação (*empowerment*) da língua na comunicação oral.

d) Habilidades de monitoração da fala e estratégias de modificação da fala para uso além da sala de aula: o objetivo é dar suporte aos aprendizes para desenvolverem uma consciência da fala (*speech awareness*), habilidades de monitoração da fala e estratégias de ajuste da fala que os capacitarão a desenvolver a comunicabilidade e confiança tanto dentro da sala de aula como fora.

O professor não necessita ter como alvo fazer seus alunos terem uma pronúncia de falante nativo. Com exceção de alguns indivíduos com grandes dons lingüísticos, esse alvo não é real. Uma meta mais modesta e realista seria a de ajudar os alunos terem boa inteligibilidade e que a pronúncia não seja impedimento para sua comunicação (Celce-Murcia, Brinton & Goodwin, 1996:9; Morley, 1991:498). É fundamental

ter em mente, todavia, que o ensino de inglês como língua estrangeira em geral se dá por meio de um texto: livro do aluno, caderno de exercícios, artigos de revista, notícias de jornal, material impresso da Internet etc. O professor na maior parte do tempo estará pronunciando a partir da forma escrita, carregando, portanto, a responsabilidade de ser um modelo de pronúncia para seus alunos, os quais, em termos gerais no Brasil, não têm muito acesso a outras fontes de informação nesse campo, como TV a cabo, DVD e cursos em CD-ROM.

Certamente, existem vários fatores individuais que interferem no aperfeiçoamento da pronúncia de cada falante não-nativo, tais como idade, sexo, grau de instrução, extroversão, aptidão individual para imitar sons, tempo de exposição à língua-alvo, motivação e preocupação individual por ter uma boa pronúncia. Além do mais, muitas vezes o falante não-nativo de inglês consciente ou inconscientemente mantém traços de sua língua-mãe para marcar sua etnia, identidade cultural, nacional ou social. (Pennington & Richards, 1986:215; Kenworthy, 1987:4 *apud* Celce-Murcia, 1991:137; Jenkins, 2003:125; Laver, 1995:69).

Cremos que o professor não necessita ter a pronúncia igual a de um nativo, porém é necessário ter a avidez para progredir e aprimorar, evitando que os erros passem de geração para geração (Medgyes, 1994:37). Este trabalho visa a dar um passo na direção de suprir essa carência.

1.4.3 Inteligibilidade

A frase de Morley (1991:488), a seguir, resume bem nossa visão de pronúncia: *"Intelligible pronunciation is an essential component of communicative competence"*.[38]

[38] Em português: Pronúncia inteligível é um componente essencial da competência comunicativa.

Faz-se necessário, porém, definir *inteligibilidade*. Jenkins (2000:69) faz uma revisão da literatura sobre inteligibilidade e chega à conclusão de que não há ainda total consenso sobre o que vem a ser inteligibilidade: *"... for there is as yet no broad agreement on a definition of the term 'intelligibility' "*.[39]

Uma visão de inteligibilidade que estava fortemente presente no ensino de inglês como língua estrangeira, relatada por Bamgbose (1998:10) *apud* Atechi (2004:61), era a seguinte:

> *Such intelligibility was a one-way process in which non-native speakers are striving to make themselves understood by native speakers whose prerogative was to decide what is intelligible and what is not.*[40]

No mesmo artigo, Bamgbose (1998:11) *apud* Jenkins (2000:69) define então inteligibilidade como:

> *A complex of factors comprising recognizing an expression, knowing its meaning, and knowing what that meaning signifies in the sociocultural context.*[41]

Essa definição de Bamgbose envolve fatores que Smith & Nelson (1985:334) *apud* Jenkins (2000:70) dividem em:

a) Inteligibilidade (*intelligibility*): relacionada ao reconhecimento de uma palavra ou enunciado;

b) Compreensibilidade (*comprehensibility*): relacionada à compreensão do sentido da palavra ou enunciado;

---

[39] Em português: ... pois até agora não há muito acordo sobre uma definição para o termo *inteligibilidade*.
[40] Em português: Tal inteligibilidade era um processo de mão única no qual falantes não-nativos estão se esforçando para se fazer entendidos por falantes nativos, cuja prerrogativa era decidir o que é inteligível e o que não é.
[41] Em português: Um complexo de fatores compreendendo o reconhecimento de uma expressão, o conhecimento de seu sentido e o conhecimento do que esse sentido significa no contexto sociocultural.

c) Interpretabilidade (*interpretability*): relacionada à compreensão da intenção do falante ao produzir o enunciado.

Outros autores apresentam nomenclaturas diferentes, como *identificação* (*identification*) de Brown (1995:10) em relação à *inteligibilidade* de Smith & Nelson. James (1998:212) chama de *inteligibilidade* o que Smith & Nelson chamam de *compreensibilidade*, e ainda apresenta o conceito de *comunicatividade* (*communicativity*), a qual ele descreve como ”uma noção mais ambiciosa, envolvendo acesso a forças pragmáticas, implicaturas e conotações”.

Em nosso trabalho, assumimos a visão de Smith & Nelson, e assumimos também que a pronúncia ruim pode interferir na inteligibilidade, ou seja, no reconhecimento das palavras, na compreensibilidade, impossibilitando a compreensão do sentido da palavra e na interpretabilidade, atrapalhando a compreensão da intenção do falante.

1.4.4 EFL, EIL ou ELF?

Existem várias nomenclaturas para o uso do idioma inglês por parte de falantes não-nativos:

a) Inglês como Língua Estrangeira (English as a Foreign Language - EFL): o falante não-nativo não mora numa localidade onde o inglês desempenha funções no governo, legislação, educação etc. Por exemplo, um brasileiro aprendendo inglês em São Paulo.

b) Inglês como Língua Internacional (English as an International Language - EIL): trata-se de usar o inglês sem um prévio alinhamento com as pronúncias britânica, americana ou qualquer outra. Refere-se a uma variedade de inglês mais universal.

c) Inglês como Língua Franca (English as a Lingua Franca - ELF): fundamenta-se no fato de que a língua inglesa é mais utilizada na comunicação entre não-nativos do que entre não-nativos e nativos. Esse termo visa a dar uma idéia de comunidade, e não de estranheza (*alienness*), diminuindo a dicotomia nativo/não-nativo, focalizando no fato de o inglês ser uma língua que liga os povos, comum a todos (Jenkins, 2004:33; Laver, 1995:80).

Optamos por utilizar o termo *inglês como língua estrangeira*, pelo fato de, como Jenkins (2003:126) põe em xeque, não haver ainda uma visão clara do que seria o inglês internacional, quais seriam suas características reais, e se seria possível manter a inteligibilidade entre todas as variantes de inglês.

Inglês como Língua Franca também não é ainda um termo consagrado na literatura acadêmica, conforme diz sua maior incentivadora, Jennifer Jenkins (2004:33) e também não há muito consenso sobre suas características.

Assim, em nosso trabalho assumimos o termo *inglês como língua estrangeira*.

Não há muita discórdia sobre o termo *inglês como segunda língua* (English as a Second Language - ESL): quando o falante não-nativo está inserido numa localidade onde o inglês é a língua através da qual os falantes dessa localidade desempenham suas funções (L1), como nos EUA, ou onde o inglês é uma segunda língua institucionalizada presente na educação, legislação, no governo etc., como em Camarões, onde o inglês divide o *status* de língua oficial com o francês (Jenkins, 2003:2), tal falante estaria aprendendo inglês como segunda língua. Um brasileiro que mora nos EUA ou em Camarões, portanto, estaria aprendendo inglês como segunda língua.

1.4.5 Breve Histórico do Ensino da Pronúncia do Inglês

O ensino da pronúncia teve altos e baixos em termos de *status* dentro do ensino do inglês como língua estrangeira. Porém, sua evolução e amadurecimento são inegáveis, deixando para trás a atitude autoritária, na qual qualquer desvio da regra imposta pela pronúncia tida como padrão era sumariamente condenado. Esse excerto de Stevick, 1976:93 apud Medgyes, 1944:49 dá uma boa visão da área nos anos 20, onde os alunos eram vistos como ''pacientes sofrendo de defeitos de dialeto estrangeiro":

> *If the patient stubbornly persists in substituting T as in "town" for TH as in "thin" ... hold the blade of his tongue forcibly down in its proper position by means of a wire form [called] a "fricator", if he persists ... push his tongue back into its proper position with a forked metal brace.*[42]

Dos anos 40 até o início dos 60, a pronúncia ocupava uma posição central no ensino de inglês, em métodos como o audiolingual, onde a gramática correta e a pronúncia precisa eram metas de alta prioridade O foco estava na produção repetitiva de sons isolados e palavras, usando pares mínimos, sem muita atenção ao acento, ritmo e entonação (Pennington & Richards, 1986:207).

Nos anos 60, a importância do ensino de pronúncia começa a ser questionada. Questões sobre se ela deveria ser o foco central no ensino ou uma área acessória, se ela deveria ser ensinada de maneira direta ou diluída em todas as outras áreas do ensino do inglês, questões até mesmo sobre se ensinar pronúncia é algo factível ou não. Todas essas

[42] Em português: Se o paciente teimosamente persistir em substituir o T de *town* por TH de *thin* ... segure a língua dele para baixo com força na posição correta por meio de um fio [chamado] "fricador". Se ele persistir ... empurre a língua dele para trás na posição correta com um grampo de metal bifurcado.

interrogações estavam sob influência direta do paradigma chomskyano vigente na época. Como conseqüência, houve perda crescente de espaço para o ensino da gramática e de vocabulário – alguns programas chegaram até mesmo a banir por completo o ensino da pronúncia – e diminuição do volume de publicações sobre o assunto (Morley, 1991:485; Celce-Murcia, Brinton & Goodwin, 1996:5).

Abordagens da época, passam a ver o erro de pronúncia como parte do processo natural de aprendizagem, e assumiam que tais erros desapareceriam conforme o aprendiz fosse aumentando seu nível de proficiência. Portanto, não precisavam receber muita atenção em sala de aula.

Durante os anos 70, houve algumas indicações de mudança, porém a pronúncia passou a ser vista de uma perspectiva diferente. Passa-se a questionar as práticas em sala de aula em relação a como corrigir o aluno, ao papel do aluno no processo de aprendizagem e a seu aspecto emocional etc. Surgem métodos como o Silent Way e o Community Approach, que valorizam o aspecto da pronúncia no ensino. Dá-se também mais espaço ao estudo das relações da ortografia com a pronúncia, como os trabalhos de Kriedler (1972) e Dickerson (1975).

A partir dos anos 80 até os dias de hoje, a Abordagem Comunicativa (Communicative Approach) figura como a abordagem dominante no ensino de línguas, valorizando, como o próprio nome já explicita, a comunicação como o propósito central da linguagem, trazendo uma nova urgência ao ensino de pronúncia: por maior que seja o domínio do falante não-nativo de inglês sobre a gramática e o vocabulário, se ele estiver abaixo de um limite mínimo em termos de qualidade de pronúncia, ele terá problemas de comunicação oral (Celce-Murcia, Brinton & Goodwin, 1996:7).

1.5 A ORTOGRAFIA DO INGLÊS

Nesta seção, apresentamos alguns aspectos teóricos sobre a ortografia do inglês.

1.5.1 Um Breve Histórico

Segundo Katsiavriades & Qureshi (2002), estima-se que nos dias atuais haja mais de 300 milhões de falantes nativos e outros 300 milhões que usam o inglês como segunda língua. O inglês é a língua da ciência, da computação, da diplomacia, do turismo e da aviação. Figura como língua oficial ou co-oficial em mais de 45 países e é falada extensivamente em outros países onde não tem *status* oficial. É a segunda língua mais falada no mundo, perdendo apenas para o mandarim:

> *Whether you like it or not, English has become the primary language of international communication, the lingua franca of the world, and it is rolling ahead like a juggernaut. More people speak English today than have ever spoken any single language in the recorded history of the world* (Medgyes, 1994:1).[43]

O inglês é classificado como uma língua germânica, da família das línguas Indo-Européias e sua história divide-se em três períodos (Schütz, 2005), como exposto a seguir no quadro 1.2.

Através dos séculos, povos de língua celta, germânica (anglo-saxões), latina (romanos) e normanda (da região ao norte da França) disputaram o domínio das ilhas britânicas. O inglês que usamos hoje, século XXI, é o resultado de centenas de guerras e invasões travadas

[43] Em português: Quer você goste ou não, o inglês tornou-se a principal língua de comunicação internacional, a língua franca do mundo e está avançando como uma locomotiva. Mais pessoas falam inglês hoje do que já falaram qualquer outra língua de que se tenha registro na história do mundo.

em solo britânico; é a mistura de milhares de vocábulos, resultando em uma ortografia heterogênea.

De 500 D.C. a 1100 D.C.	*Old English* - Inglês Antigo
De 1100 D.C. a 1500 D.C.	*Middle English* - Inglês Médio
De 1500 D.C. até hoje	*Modern English* - Inglês Moderno

Quadro 1.2 – Períodos da história do inglês.

Dentre os povos acima citados, os celtas são o povo que menos marca presença no inglês usado hoje pelo fato de terem sido praticamente dizimados pelos anglo-saxões no século V. E também porque, com a introdução do cristianismo no final do século VI, a cultura celta, estigmatizada pela bruxaria, sofreu fortíssima rejeição.

Sobre o Old English, Schütz (2005) comenta:

> *Old English, às vezes também denominado Anglo-Saxon, comparado ao inglês moderno, é uma língua quase irreconhecível, tanto na pronúncia, quanto no vocabulário e na gramática. Para um falante nativo de inglês hoje, das 54 palavras do Pai Nosso em Old English, menos de 15% são reconhecíveis na escrita, e provavelmente nada seria reconhecido ao ser pronunciado. A correlação entre pronúncia e ortografia, entretanto, era muito mais próxima do que no inglês moderno. No plano gramatical, as diferenças também são substanciais. Em Old English, os substantivos declinam, têm gênero (masculino, feminino e neutro) e os verbos são conjugados.*

Em 1066, a Batalha de Hastings foi um marco histórico para a Inglaterra. Representou não só uma drástica reorganização política,

mas também alterou os rumos da língua inglesa, marcando o início de uma nova era. William the Conqueror, Duque da Normandia (norte da França), comandou a invasão das ilhas britânicas, conquistando assim um território com mais de um milhão e meio de habitantes e, provavelmente, o mais rico da Europa na época. Durante os 300 anos que se seguiram (Middle English), principalmente nos 150 anos iniciais, a língua usada pela aristocracia na Inglaterra foi o francês, tornando-se a língua do poder. Falar francês tornou-se então uma condição para aqueles de origem anglo-saxônica em busca de ascensão social através da simpatia e dos favores da classe dominante.

O leitor poderá ver toda a evolução da língua inglesa através do histórico de Schütz (2005) na Internet[44], que descreve os principais fatos que a influenciaram.

1.5.2 The Great Vowel Shift

Uma acentuada mudança na pronúncia das vogais do inglês ocorreu entre 1450 (final do Middle English) e 1700 (Modern English) e foi amplamente generalizada por volta de 1750. Praticamente todos os sons vocálicos, inclusive ditongos, sofreram alterações e algumas consoantes deixaram de ser pronunciadas. O quadro 1.3 a seguir exemplifica essa grande mudança.

Sobre a Great Vowel Shift, Schütz (2005) afirma:

> *O sistema de sons das vogais da língua inglesa antes do século 15 era bastante semelhante ao das demais línguas da Europa ocidental, inclusive do português de hoje. Portanto, a atual falta de correlação entre ortografia e pronúncia do inglês moderno, que se observa principalmente nas vogais, é, em grande parte, conseqüência desta mudança ocorrida no século 15.*

[44] Endereço na Internet: http://www.sk.com.br/sk-enhis.html

Durante o Middle English, também ocorreu a gradual perda das declinações e neutralização dos substantivos.

Vocábulo	Pronúncia das vogais antes da Great Vowel Shift	Pronúncia moderna
Fine	/fine/	/faɪn/
House	/hus/	/haʊs/
Deed	/ded/, semelhante à pronúncia de *dedo* em português	/did/
Fame	/fɑme/, semelhante à atual pronúncia de *father* (em relação ao <a>)	/feɪm/
So	/sɔ/, semelhante à atual pronúncia de *saw*	/soʊ/
To	/toʊ/, semelhante à atual pronúncia de *toe*	/tu/

Quadro 1.3 – Exemplos de mudanças nas vogais ocasionadas pela Great Vowel Shift (Schütz, 2005).

O período que se seguiu, Modern English, caracterizou-se pela padronização e unificação da língua inglesa, após o advento da imprensa em 1475 e do serviço postal criado por Henrique VIII, disseminando assim o dialeto de Londres, que já possuía o *status* de centro político, social e econômico da Inglaterra. A disponibilidade de materiais impressos também impulsionou a educação, trazendo a alfabetização ao alcance da classe média.

Tal disseminação do inglês coincidiu com a Great Vowel Shift iniciada no período anterior, Middle English.

D'Eugenio (1982:319) assim explica o que ocorreu:

> *O processo de padronização da língua inglesa iniciou em princípios do século 16 com o advento da litografia, e acabou fixando-se nas presentes formas ao longo do século 18, com a publicação dos dicionários de Samuel*

> *Johnson em 1755, Thomas Sheridan em 1780 e John Walker em 1791. Desde então, a ortografia do inglês mudou em apenas pequenos detalhes, enquanto que a sua pronúncia sofreu grandes transformações. O resultado disto é que hoje em dia temos um sistema ortográfico baseado na língua como ela era falada no século 18, sendo usada para representar a pronúncia da língua no século 20* (tradução de Schütz).

Portanto, as mudanças ocorridas na pronúncia não se traduziram em reformas ortográficas.

Sampson (1996:214), entretanto, discorda da posição de que a ortografia se distanciou da pronúncia devido a uma simples postura avessa a reformas. Para ele, o principal fator foi a introdução de grafias estrangeiras, especialmente o francês, e sua influência sobre os copistas nativos ingleses. Para o autor, não fora o domínio normando de três séculos e meio sobre as ilhas britânicas, o inglês de hoje seria tão fonêmico quanto o alemão ou as línguas escandinavas. Sampson relata que o próprio francês ainda não havia adotado convenções ortográficas convincentes e definidas. Além do mais, os copistas, ainda que falantes nativos de inglês, passavam boa parte de seu tempo escrevendo em francês e acabavam naturalmente por transferir convenções do francês para o inglês. Isso trouxe inconsistências, como <ee> em *deed* e *heel*, mas <ie> em *thief*, alinhando-se à ortografia francesa.

No final do século XV, William Caxton introduziu a técnica da impressão na Inglaterra, após ter vivido trinta anos nos Países Baixos. Tal fato o impediu de estar a par das convenções ortográficas britânicas no momento e possibilitou a influência das convenções ortográficas do holandês, como o <gh> em *ghost*.

Para aumentar a distância entre a ortografia e a pronúncia do inglês, havia ainda a influência do latim através do princípio fonético. Tal princípio primava por manter a origem latina na grafia das palavras.

O inglês medieval, porém, adotou esse princípio de maneira inconsistente: o <h> latino está presente em *honour* e *hour*, mas não em *ability*, por exemplo.

Sampson, portanto, expõe que a grafia do inglês moderno resulta de uma variedade de causas, e não de uma simples postura conservadora de não alterar a escrita a despeito das mudanças na língua falada.

1.5.3 Reformas

Reformas ortográficas começam a ser consideradas sempre que o uso prolongado de um sistema ortográfico apresenta corrupções nas relações fundamentais entre seus signos gráficos e as unidades lingüísticas que eles representam (Coulmas, 2000:248)[45].

Crystal (1997:276) relata alguns tipos de abordagens para reformar a língua:

a) Abordagem de padronização: usa letras já conhecidas, de maneira mais regular, normalmente adicionando novos dígrafos, porém sem introduzir novos símbolos.

b) Abordagem de aumento: adiciona novos símbolos, letras e diacríticos[46].

c) Abordagem de suplantação: substitui toda a ortografia tradicional por novos símbolos[47].

d) Abordagem de regularização: aplica as regras já existentes de maneira mais consistente, retirando letras mudas, letras redundantes etc.

---

[45] Para saber sobre as reformas ortográficas no Brasil, ver Scliar-Cabral, (2003:71).
[46] Diacríticos são sinais que se apõem a uma letra para dar-lhe novo valor, como a cedilha, o til, o trema e os acentos (Steinberg, 1985:62; Scliar-Cabral, 2003:28).
[47] Há exatos 40 anos, Wijk (1966:150) chamava a atenção de seus leitores para o custo de uma abordagem de aumento ou de suplantação incorridos por causa da substituição de máquinas de escrever e equipamento de impressão. Hoje, entretanto, faz-se quase tudo digitalmente, o que tornaria os custos da reforma infinitamente mais baixos e tal argumento bem mais fraco.

O fato de a relação ortografia-pronúncia no inglês não ser transparente há muito vem gerando debates e alimentado vários movimentos em prol de uma reforma ortográfica.

Venezky (1970:8), na introdução de seu famoso The Structure of the English Orthography, deixa bem claro que não concorda com os educadores, filólogos e reformistas que tacham a língua inglesa de antiquada, inconsistente, ilógica, degenerada e fraca em termos de adaptabilidade, clamando por uma condenação rápida e uma execução sumária da ortografia vigente. Venezky, citando o uso de computadores, revela ter encontrado um alto grau de padronização jamais verificado antes. O autor defende alguns ajustes, como a eliminação de letras mudas (o <b> em *doubt*, por exemplo), mas não defende a adoção de um sistema do tipo um grafema para cada fonema, porque isso alteraria a padronização morfológica básica da ortografia, como em *sane* /seɪn/ e *sanity* /ˈsænɪti/, onde os atuais grafemas <a> seriam escritos com letras diferentes, ocultando a raiz morfológica que liga essas duas palavras.

Sampson (1996:224) também se posiciona contra a reforma ortográfica. Porém o autor crê que ela deveria ser uma exigência popular, e não uma imposição do governo, que por sinal, na maioria dos países onde o inglês é falado, tem por tradição a não-intervenção em assuntos culturais. Além do mais, um único país não poderia implantar as mudanças unilateralmente. A demanda popular deveria ocorrer em várias nações ao mesmo tempo para dar um caráter universal à reforma.

O autor continua argumentando contra a realização da reforma ortográfica, dizendo que as razões mais fortes que farão com que ela nunca ocorra são mais de caráter subjetivo que objetivo. As pessoas crêem que, em termos estéticos, por exemplo, uma ortografia reformada seria pouco atraente.

Por outro lado, nos dias atuais, pela Internet, proliferam os sítios que levantam a bandeira da reforma-já e de uma ortografia simplificada, sugerindo novos alfabetos, novas convenções etc.

Alinhamo-nos ao pensamento de Venezky. Cremos que as abordagens de padronização, aumento e, especialmente, a de suplantação não preservariam as raízes morfológicas do inglês e atrapalharia a correspondência grafofonêmica ainda mais, posto que a maioria dos leitores está familiarizada com os processos de derivação e flexão.

1.5.4 Reformistas

A história cita grandes defensores da reforma ortográfica, como Mark Twain (1835-1910), famoso escritor americano, autor de *As Aventuras de Tom Sawyer*, em 1881, e *Huckleberry Finn*, em 1884, e George Bernard Shaw (1856-1950), dramaturgo e crítico literário irlandês, ganhador do prêmio Nobel de literatura em 1925.

Obteve também destaque como grande reformista o advogado de formação, porém professor de profissão, Noah Webster, responsável por alterações na ortografia americana para deliberadamente torná-la diferente do modelo britânico, tais como:

a) Queda do <u> em palavras com final <our>: *color*, *favor* e não mais *colour*, *favour*;

b) Queda de consoantes redundantes: *traveled,* e não mais *travelled*;

c) Queda do <k> final: *frolic*, *almanac*, *traffic* e não mais *frolick*, *almanack*, *traffick*;

d) Transposição do <e> e do <r>: *center*, *theater, fiber* e não mais *centre, theatre*, *fibre*.

Neste capítulo, buscamos apresentar ao leitor um breve resumo dos princípios teóricos da Lingüística de Corpus, da relação entre a fala

e a escrita, da correspondência grafofonêmica, do ensino da pronúncia do inglês como língua estrangeira e da ortografia do inglês. Princípios estes que nortearam nosso trabalho.

No capítulo seguinte, apresentamos a metodologia de pesquisa utilizada em nossa investigação.

Capítulo 2 – Metodologia de Pesquisa

Breaking new ground requires a lot of wrestling with the language to make it say just what you want it to say and not what the generally accepted opinion says.

Monaghan (1979:3)

A seguir, descrevemos a metodologia de pesquisa utilizada na investigação relatada nesta dissertação. Primeiramente, apresentamos os objetivos de nosso trabalho; em seguida, as ferramentas eletrônicas criadas para esta pesquisa. Por fim, os procedimentos metodológicos empregados na coleta de dados e análise dos resultados.

2.1 Objetivos e Questões de Pesquisa

O objetivo desta pesquisa é saber quais vocábulos da língua inglesa exibem uma correspondência grafofonêmica inconsistente, ou seja, uma relação atípica entre a ortografia e a pronúncia, mas que também exibem freqüência de uso relevante, mostrada através de um corpus de inglês geral. Baseamo-nos em grafemas e seqüências de grafemas extraídos do trabalho de Lessa (1985), que tendem a levar brasileiros falantes de inglês como língua estrangeira a cometerem erros de pronúncia. Como exemplos, podemos citar algumas palavras presentes no trabalho de Lessa: *sew* /soʊ/, *gnarled* /narld/, *canoe* /kəˈnu/, *bury* /ˈberi/ e *butcher* /ˈbʊtʃər/.

Com base em nossos dados, buscaremos também hierarquizar os grafemas em termos de complexidade. Isso auxiliará os professores e elaboradores de material didático a focalizarem mais nos casos que mais tendem a causar confusão em termos de correspondência grafofonêmica.

Desejamos que este seja o primeiro passo no processo de aprimoramento da formação de professores brasileiros de inglês na área de pronúncia a partir da escrita.

Duas questões de pesquisa nortearam-nos em nosso trabalho:

a) Com base nos grafemas extraídos do trabalho de Lessa (1985), quais são os vocábulos que exibem uma relação

atípica entre a ortografia e a pronúncia e que apresentam freqüência de uso relevante na língua inglesa?

b) Quais são os grafemas que exibem maior atipicidade grafofonêmica do ponto de vista léxico-freqüencial?

2.2 Delimitação do Escopo da Pesquisa e Definição de Erro

A área de pronúncia de uma língua estrangeira envolve várias subáreas comumente abordadas em materiais didáticos. Por exemplo, o livro Pronunciation Plus da Cambridge University Press (Hewings & Goldstein, 1998) traz as seguintes seções: ritmo, entonação, acento, as vogais, as consoantes, fala corrente (*connected speech*) e uma seção dedicada à pronúncia de palavras a partir da forma escrita (*Part 8 - Pronouncing Written Words*). Nosso enfoque recai exatamente sobre esta última subárea: pronúncia a partir da forma escrita, comumente chamada em inglês de *from spelling to sound* ou *phonics*.

Trata-se de palavras cuja forma ortográfica conduz à escolha de uma pronúncia destoante da forma convencionalizada pela sociedade. Temos basicamente duas causas:

a) Transferência: refere-se à influência da língua-materna do falante de inglês como língua estrangeira. Um exemplo seria pronunciar grafemas mudos, (*gnome*, *leopard, salmon* etc.), que são pronunciados em português.

b) Generalização dentro da língua-alvo: refere-se ao uso da correspondência grafofonêmica mais comum de uma dada seqüência de grafemas para todas as seqüências iguais ou semelhantes. Por exemplo: usar a pronúncia da seqüência <-uce> de *reduce* /rɪˈdus/, *produce* /prəˈdus/, *deduce* /dɪˈdus/ etc. para pronunciar *lettuce* /ˈletəs/, incorrendo, portanto, em erro.

Novamente, gostaríamos de frisar que questões relacionadas às outras áreas acima mencionadas (ritmo, entonação, acento, fala corrente, articulação etc.) não foram abordadas neste trabalho.

Tomamos como padrão em nossa pesquisa a pronúncia americana presente no dicionário fonêmico CMU - Carnegie Mellon University (http://www.speech.cs.cmu.edu/cgi-bin/cmudict), descrito na seção 2.5. O tipo de inglês americano presente no dicionário eletrônico CMU é o GA – General American: uma variedade de inglês americano que revela pouco ou nada sobre a origem geográfica do falante, com poucas peculiaridades regionais, não apresentando traços nem do leste nem do sul dos EUA. É o inglês usado pela maioria dos apresentadores de programas de rádio e TV voltados ao público americano (Laver, 1995:58).

2.3 Procedimentos de Pesquisa

Para que o leitor tenha uma visão geral dos procedimentos usados em nossa investigação, descrevemos aqui todas as fases da pesquisa, as quais serão detalhadas nas seções a seguir:

1. Seleção no trabalho de Lessa (1985) dos grafemas que causam dificuldades a falantes de português brasileiro ao falar inglês;
2. Coleta no dicionário eletrônico CMU das palavras que contêm os grafemas mencionados no item acima;
3. Coleta no corpus de inglês geral BNC das freqüências de uso de cada uma das palavras coletadas no CMU;
4. Análise e identificação das palavras que apresentam correspondência grafofonêmica inconsistente, porém com freqüência de uso relevante, respondendo a pergunta de pesquisa a).

5. Análise e identificação dos grafemas mais atípicos. Os grafemas que apresentaram maior número de realizações fonêmicas e maior soma de freqüência de uso foram considerados os mais atípicos, respondendo a pergunta de pesquisa b);

2.4 Coleta e Seleção dos Grafemas

A pesquisa iniciou-se com a coleta dos grafemas, que causam mais dificuldades para o falante brasileiro de inglês como língua estrangeira. Poderíamos ter estudado todas as combinações de grafemas da língua inglesa, contudo optamos por utilizar esse recurso para tornar nosso trabalho mais específico em relação ao caso do falante de português brasileiro.

Para esse recorte, baseamo-nos no trabalho de Lessa (1985). A autora não define explicitamente quais seriam os grafemas que mais causam confusão para a pronúncia dos brasileiros, porém analisamos seu trabalho e coletamos 90 palavras usadas nos testes aplicados a alunos brasileiros participantes de sua pesquisa, consideradas pela autora como de difícil pronúncia devido à relação grafema-fonema atípica. Dessas palavras, extraímos os grafemas que causam tal dificuldade, à luz do exposto por Treiman et al. (2002) no tocante à consideração do contexto grafêmico. Por essa razão, buscamos não analisar grafemas isoladamente, mas sempre incluir um contexto. A exceção a essa regra foram os grafemas iniciais mudos <h> em *heir*, <k> em *knapsack*, e <p> em *psychology*.

O quadro 2.1 apresenta os vocábulos extraídos do trabalho de Lessa (1985), suas transcrições fonológicas extraídas do dicionário eletrônico da Carnegie Mellon University, os erros típicos baseados nos trabalhos de Shepherd (1987) e Lieff & Nunes (1993), envolvendo alunos brasileiros de níveis básicos e avançados. E na última coluna, os grafemas extraídos dos vocábulos de Lessa (1985).

Lessa (1985) considerou as pronúncias dos vocábulos *almond* e *herb* como sendo apenas /ˈæmənd/ e /ɜrb/. Contudo, o CMU apresenta estas pronúncias e também as variantes /ˈælmənd/ e /hɜrb/.

Lessa (1985) também considerou a pronúncia de *thyme* unicamente como /tɑɪm/, pronúncia esta corroborada pelo Longman Dictionary of Contemporary English (2003). O dicionário eletrônico CMU, todavia, traz apenas /θɑɪm/.

	Vocábulos	Transcrição	Erro Típico	Grafemas
1	abbey	ˈæbi	ˈæbeɪ	<ey> final
2	allegiance	əˈlidʒəns	əˈledʒəns	<e> interconsonantal
3	almond	ˈæmənd ou ˈælmənd	-	<l> mudo em qualquer posição
4	arch	ɑrtʃ	ɑrk	<ch> final
5	athlete	ˈæθlit	ˈæθlet	<e> interconsonantal
6	baked	beɪkt	ˈbeɪkɪd	<ed> final
7	barley	ˈbɑrli	ˈbɑrleɪ	<ey> final
8	blood	blʌd	blʊd	<oo> em qualquer posição
9	breakfast	ˈbrekfəst	ˈbreɪkfəst	<ea> em qualquer posição
10	bribery	ˈbrɑɪbəri	ˈbrɪbəri	<i> interconsonantal
11	brooch	broʊtʃ	brʊtʃ	<oo> qualquer posição
12	bury	ˈberi	ˈbʌri	<ury> final
13	butcher	ˈbʊtʃər	ˈbʌtʃər	<u> interconsonantal
14	butter	ˈbʌtər	ˈbʌdər	<t> intervocálico
15	cabs	kæbz	kæbs	<s> final
16	canoe	kəˈnu	kəˈnoʊ	<oe> final
17	chocolate	ˈtʃɔklət	ˈtʃɔkleɪt	<ate> final
18	color	ˈkʌlər	ˈkɔlər	<o> interconsonantal
19	comb	koʊm	koʊmb	<omb> final
20	cough	kɑf	kʌf	<ough> final
21	country	ˈkʌntri	ˈkaʊntri	<ount> qualquer posição
22	cover	ˈkʌvər	ˈkoʊvər	<o> interconsonantal
23	cushion	ˈkʊʃən	ˈkʌʃən	<u> interconsonantal
24	dim	dɪm	din	<m> final
25	doubtful	ˈdaʊtfəl	ˈdaʊbtfəl	<bt> em qualquer posição

Quadro 2.1 – Vocábulos com correspondência grafofonêmica atípica segundo Lessa (1985).

	Vocábulos	Transcrição	Erro Típico	Grafemas
26	draught	dræft	drɔgt	<aught> final
27	exact	ɪgˈzækt	ɪkˈzækt	<ex> inicial
28	famous	ˈfeɪməs	ˈfeɪmoʊs	<ous> final
29	finite	ˈfɑɪnɑɪt	ˈfɪnɪt	<i> interconsonantal
30	flood	flʌd	flʊd	<oo> em qualquer posição
31	freight	freɪt	frɑɪt	<ei> em qualquer posição
32	fruit	frut	frʊɪt	<ui> em qualquer posição
33	furlough	ˈfɜrloʊ	ˈfɜrləf	<ough> final
34	gaol	dʒeɪl	geɪl	<aol> final
35	gauge	geɪdʒ	gɔdʒ	<auge> em qualquer posição
36	gem	dʒem	dʒeɪm	<m> final
37	gnarled	nɑrld	gnɑrld	<gn> inicial
38	gnome	noʊm	gnoʊm	<gn> inicial
39	guinea	ˈgini	ˈgʊɪni	<ui> em qualquer posição
40	half	hæf	hɑlf	<l> mudo
41	heart	hɑrt	hʌrt	<ear> em qualquer posição
42	heifer	ˈhaɪfər *ou* ˈheɪfər	ˈhaɪfər *ou* ˈheɪfər	<ei> em qualquer posição
43	heir	er	her	<h> inicial mudo
44	helmet	ˈhelmət	ˈhelmeti	<t> final
45	herb	hɜrb *ou* ɜrb	-	<h> inicial mudo
46	heritage	ˈherɪtɪdʒ	ˈherɪteɪdʒ	<age> final
47	journal	ˈdʒərnəl	ˈdʒɔrnəl	<our> em qualquer posição
48	juice	dʒus	dʒʊɪs	<ui> em qualquer posição
49	knapsack	ˈnæpsæk	ˈknæpsæk	<kn> inicial
50	leisure	ˈliʒər	ˈleɪʒər	<ei> em qualquer posição
51	leopard	ˈlepərd	ˈleoʊpɑrd	<leo> inicial
52	lettuce	ˈletəs	ˈletus	<uce> final
53	linen	ˈlɪnən	ˈlɑɪnən	<i> interconsonantal
54	loathed	loʊðd	ˈloʊθɪd	<ed> final
55	love	lʌv	lɔv	<o> interconsonantal
56	method	ˈmeθəd	ˈmesəd	<th> em qualquer posição
57	milk	mɪlk	ˈmɪlki	<k> final
58	minute	ˈmɪnət	ˈmɪnʊt	<ute> final
59	money	ˈmʌni	ˈmʌneɪ	<ey> final
60	museum	mjʊˈziəm	mjʊˈziən	<m> final

Quadro 2.1 – Vocábulos com correspondência grafofonêmica atípica, segundo Lessa (1985) (continuação).

	Vocábulos	Transcrição	Erro Típico	Grafemas
61	nothing	ˈnʌθɪŋ	ˈnʌsɪŋ	<th> em qualquer posição
62	nourish	ˈnɜrɪʃ	ˈnurɪʃ	<our> em qualquer posição
63	nuisance	ˈnusəns	ˈnʊɪsəns	<ui> em qualquer posição
64	orange	ˈɔrəndʒ	ˈɔreɪndʒ	<ange> final
65	original	əˈrɪdʒɪnəl	oʊˈrɪdʒɪnəl	<or> inicial
66	paradigm	ˈpærədɑɪm	ˈpæradɪgm	<igm> final
67	patriotism	ˈpeɪtriətɪzəm	ˈpeɪtriətɪsm	<ism> final
68	pear	per	pir	<ear> em qualquer posição
69	pearl	pɜrl	perl	<ear> em qualquer posição
70	pencil	ˈpensəl	ˈpeɪnsəl	<en> em qualquer posição
71	plaid	pled	pleɪd	<aid> final
72	psalm	sɑm *ou* sɑlm	psɑlm	<p> inicial mudo
73	realm	relm	ˈreɑlm	<ea> em qualquer posição
74	reign	reɪn	regn	<reign> em qualquer posição
75	sandage	ˈsændɪdʒ	ˈsændeɪdʒ	<age> final
76	sew	soʊ	sjʊ	<ew> final
77	sewage	ˈsuɪdʒ	ˈsueɪdʒ	<age> final
78	sling	slɪŋ	ˈslɪngi	<g> final
79	slough	sləf	sloʊg	<ough> final
80	social	ˈsoʊʃəl	ˈsoʊsiəl	<cial> final
81	soup	sʊp	soʊp	<oup> final
82	steak	steɪk	stik	<ea> em qualquer posição
83	stopped	stɑpt	ˈstɔped	<ed> final
84	subtle	ˈsʌtəl	ˈsʌbtəl	<bt> em qualquer posição
85	theory	ˈθiəri	ˈsiəri	<th> inicial
86	Thomas	ˈtɑməs	ˈθɔməs	<th> inicial com som de /t/
87	thyme	θaɪm	-	<th> inicial com som de /t/
88	unclean	ʌnˈclin	ʌnˈclɪn	<ea> em qualquer posição
89	vegetable	ˈvedʒtəbəl	vedʒˈteɪbəl	acento
90	washed	wɑʃt	ˈwɑʃed	<ed> final

Quadro 2.1 – Vocábulos com correspondência grafofonêmica atípica, segundo Lessa (1985) (continuação).

Nem todos os casos foram estudados. Agrupamos no quadro 2.2 as palavras em 32 seqüências de grafemas a serem estudadas, cobrindo 45 das 90 palavras coletadas do trabalho de Lessa.

Usamos 3 critérios de exclusão: ”questão muito ampla”, ”questão articulatória” e ”acento”. Os grafemas não estudados estão no quadro 2.3

		Grafemas	Vocábulo
1	1	<age> final	heritage
2			sandage
3			sewage
4	2	<aid> final	plaid
5	3	<aol> final	gaol
6	4	<ange> final	orange
7	5	<auge> final	gauge
8	6	<aught> final	draught
9	7	<bt> em qualquer posição	doubtful
10			subtle
11	8	<cial> final	social
12	9	<ear> em qualquer posição	heart
13			pear
14			pearl
15	10	<ew> final	sew
16	11	<ex> inicial	exact
17	12	<ey> final	abbey
18			barley
19			money
20	13	<gn> inicial	gnarled
21			gnome
22	14	<h> inicial mudo	heir
23			herb
24	15	<igm> final	paradigm
25	16	<ism> final	patriotism
26	17	<kn> inicial	knapsack
27	18	<leo> inicial	leopard
28	19	<omb> final	comb
29	20	<or> inicial	original
30	21	<oe> final	canoe
31	22	<ough> final	cough
32			furlough
33			slough
34	23	<ount> em qualquer posição	country
35	24	<oup> final	soup
36	25	<our> em qualquer posição	journal
37			nourish
38	26	<ous> final	famous
39	27	<p> inicial mudo	psalm
40	28	<reign> em qualquer posição	reign
41	29	<th> inicial com som de /t/	Thomas
42			thyme
43	30	<uce> final	lettuce
44	31	<ury> final	bury
45	32	<ute> final	minute

Quadro 2.2 – Grafemas pesquisados em ordem alfabética.

Questão Muito Ampla			
1	1	<ate> final	chocolate
2	2	<ch> final	arch
3	3	<e> interconsonantal	allegiance
4			athlete
5	4	<ea> em qualquer posição	breakfast
6			realm
7			steak
8	5	<ei> em qualquer posição	freight
9			heifer
10			leisure
11	6	<i> interconsonantal	bribery
12			finite
13			linen
14	7	<l> mudo	almond
15			half
16	8	<o> interconsonantal	color
17			cover
18			love
19	9	<oo> em qualquer posição	blood
20			brooch
21			flood
22	10	<u> interconsonantal	butcher
23			cushion
24	11	<ui> em qualquer posição	fruit
25			guinea
26			juice
27			nuisance
Questão Articulatória			
28	12	consoante final	helmet
29			milk
30			sling
31	13	<ea> em qualquer posição	unclean
32	14	<ed> final	baked
33			loathed
34			stopped
35			washed
36	15	<en> em qualquer posição	pencil
37	16	<m> final	dim
38			gem
39			museum
40	17	<s> final	cabs
41			ways
42	18	<th> em qualquer posição	method
43			nothing
44			theory
Acento			
45	19	vegetable	

Quadro 2.3 – Grafemas não pesquisados.

2.4.1 Exclusão dos Casos Considerados como ”Questão Muito Ampla” e ”Questão Articulatória”

Excluímos alguns casos de nosso plano de pesquisa pelo fato de demandarem a análise de milhares de ocorrências específicas, requerendo ferramentas computacionais mais poderosas do que as que estão a nossa disposição, o que também fugiria ao escopo de uma dissertação de mestrado. Por isso, rotulamos esses grafemas de ”questão muito ampla”, como por exemplo *c<u>o</u>ver, l<u>o</u>ve* e *c<u>o</u>lor*. Para estudar a questão do grafema interconsonantal <o>, teríamos que levantar em toda a língua inglesa todas as palavras que contêm tal grafema, incluindo todos os contextos, descobrir todas as realizações fonêmicas possíveis, separá-las e ainda pesquisar, somar e analisar milhares de freqüências de uso. Seria um trabalho hercúleo, não compatível com nosso cronograma nem com nossas condições técnicas.

Outros grafemas, como <m> em *dim* e *gem* e <th> em *nothing*, referem-se a questões articulatórias, ou seja, à maneira correta de usar o aparelho fonador para produzir sons condizentes com um padrão preestabelecido. A confusão não está necessariamente na escolha errada do fonema devido à forma escrita, mas sim na aproximação excessiva ao modo de articular do português, como mostra o quadro 2.4 a seguir.

Shepherd (1987:115) não considera esses casos acima como erros influenciados pela escrita, mas sim questões de articulação.

Ele aponta que o falante brasileiro de inglês tende a:

1. nasalizar a vogal anterior aos fonemas /m/, /n/ e /ŋ/ finais, praticamente eliminando o som da consoante. Dos vocábulos abaixo, *dim*, *gem*, *museum*, *pencil* e *sling* estão nessa categoria.

2. realizar o fonema /θ/ muitas vezes como /s/ ou /t/, e o fonema /ð/, como /z/ ou /d/. É o caso de *method*, *nothing* e *theory*.

3. adicionar uma vogal ao final de uma palavra que termina em consoante, construindo uma nova sílaba, como em *helmet* e *milk*.

4. Confundir a pronúncia de /s/ e /z/, como no caso de *cabs*.

Vocábulos de Lessa (1985)	Transcrição IPA	Erro[48]
cabs	/kæbz/	/kæbs/
dim	/dɪm/	/din/
gem	/dʒem/	/ʒeɪn/
helmet	/ˈhelmət/	/ˈhelməti/
method	/ˈmeθəd/	/ˈmesəd/
milk	/mɪlk/	/ˈmɪlki/
museum	/mjʊˈziəm/	/mjʊˈziən/
nothing	/ˈnʌθɪŋ/	/ˈnʌsɪŋ/
pencil	/ˈpensəl/	/ˈpeɪnsəl/
sling	/slɪŋ/	/ˈslingi/
theory	/ˈθiəri/	/ˈsiəri/

Quadro 2.4 – Vocábulos classificados como questão articulatória.

2.5 Descrição do Dicionário Fonêmico CMU

O CMU é um dicionário de pronúncia de inglês de acesso gratuito pela Internet para consulta *online* e para *download*. Ele contém 127.041 palavras grafadas segundo a ortografia americana e suas respectivas transcrições fonêmicas.

Trata-se do único dicionário fonêmico eletrônico disponível de que temos conhecimento. Ser eletrônico era um pré-requisito, visto que

[48] Existem outros erros possíveis relacionados a estes vocábulos, porém decidimos ater-nos àqueles pertinentes a este trabalho. Erros segundo Shepherd (1987) e Lieff e Nunes (1993).

trabalharíamos com um grande número de palavras. Analisá-las uma a uma manualmente consumiria muito tempo e deixaria a pesquisa altamente vulnerável a erros.

Nossa escolha não se deveu, entretanto, exclusivamente a razões técnicas. Consideramos a pronúncia americana como a mais influente na ciência, na literatura, nas artes, no mundo dos negócios, em suma, em quase todas as áreas de atividade do homem contemporâneo urbano (Crystal, 1997:111).O sistema de transcrição usado, contudo, não é o tradicional IPA (International Phonetic Alphabet), que é a base do sistema de transcrição da linha de dicionários para aprendizes de inglês como língua estrangeira e de publicações sobre pronúncia de editoras como MacMillan, Longman, Cambridge University Press e Oxford University Press. O CMU usa um sistema próprio, criado com base no sistema ASCII (American Standard Code for Information Interchange). Surgido em 1961, tendo Robert W. Bemer como um de seus inventores, ASCII é um conjunto básico de códigos usado pelo computador para representar números, letras, pontuação e outros caracteres, e que está presente em todos os computadores do mundo, não importando o sistema operacional utilizado – Windows, Mac OS, Linux, Unix etc. – (Wikipédia, 2005).

A razão que levou os criadores do dicionário eletrônico CMU a utilizar esse sistema foi de caráter puramente técnico. O objetivo é tornar o dicionário acessível através de qualquer computador, em qualquer lugar do mundo, utilizando qualquer sistema ou tipo de escrita (alfabeto cirílico, ideogramas etc). O sistema ASCII representa o que há de mais básico em termos de caracteres digitais.

2.5.1 Como consultar pronúncias através do CMU

Para realizar uma consulta no dicionário eletrônico de pronúncia CMU via Internet, o usuário digita a palavra cuja pronúncia deseja conhecer e o dicionário lhe apresenta a transcrição.

Na figura 2.1, o usuário pesquisou a pronúncia da palavra *about*, escrevendo-a no campo indicado e clicando à direita em *Look Up*. Recebeu como resposta a transcrição AH0 B AW1 T.

Para interpretar a transcrição, o usuário necessita consultar a tabela de referência com 39 fonemas, exposta no quadro 2.5, onde mostramos também a equivalência dos símbolos usados no CMU com os símbolos do IPA presentes no dicionário MacMillan for Advanced American English Learners (Rundell, 2002).

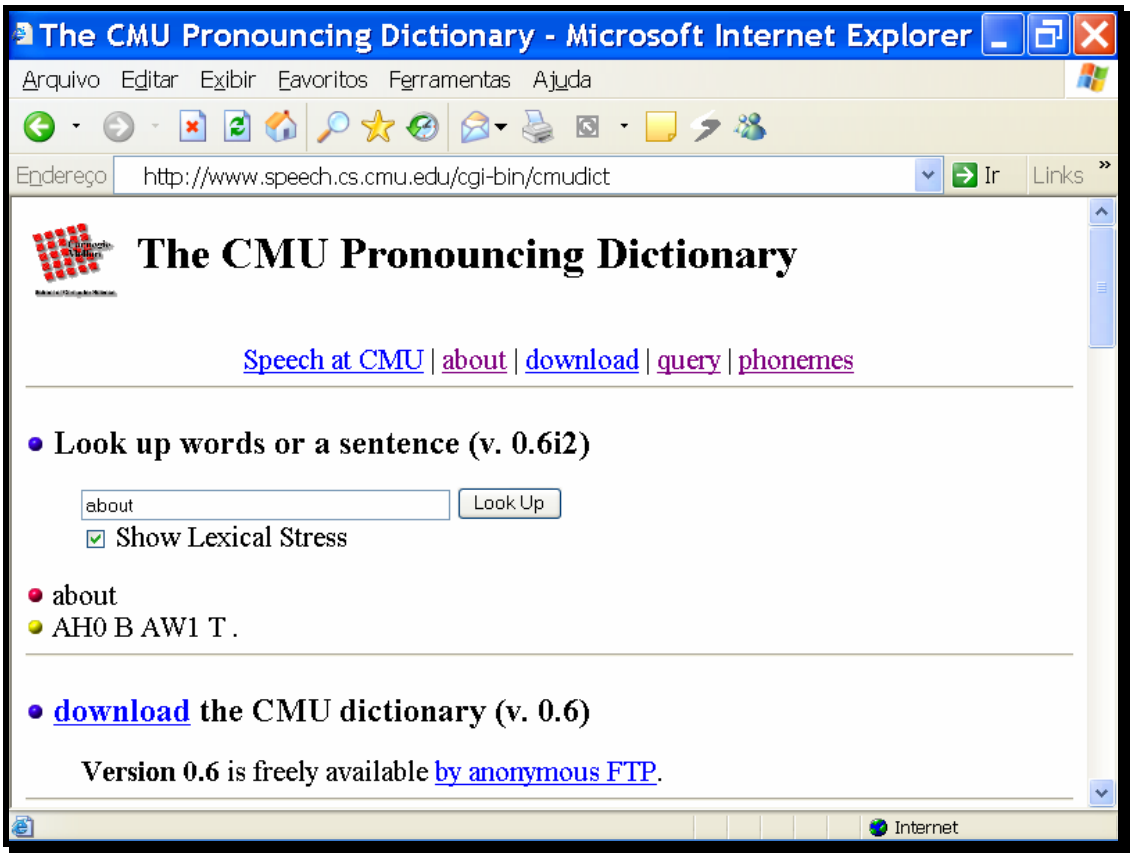

Figura 2.1 – Aspecto do sítio de busca do dicionário eletrônico CMU.

Se desejar, o usuário pode configurar o CMU para exibir também o acento, clicando em *Show Lexical Stress*. O acento lhe será mostrado através de números colocados ao lado direito das vogais. O número zero significa que o fonema que o precede é átono e o número 1, que o fonema é tônico.

O dicionário eletrônico CMU não utiliza nenhum recurso de som, apenas transcrições fonêmicas.

	Símbolo CMU	Símbolo IPA	Transcrição CMU	Example
Vogais - curtas	IH	ɪ	IH T	it
	EH	e	EH D	Ed
	AE	æ	AE T	at
	AH	ə/ʌ	HH AH T	hut
	UH	ʊ	HH UH D	hood
Vogais - longas	IY	i	IY T	eat
	AA	ɑ	AA D	odd
	AO	ɔ	AO T	ought
	UW	u	T UW	two
	ER	ɜ	HH ER T	hurt
Ditongos	EY	eɪ	EY T	ate
	AY	ɑɪ	HH AY D	hide
	OY	oɪ	T OY	toy
	OW	oʊ	OW T	oat
	AW	ɑʊ	K AW	cow
Consoantes	B	b	B IY	be
	CH	tʃ	CH IY Z	cheese
	D	d	D IY	dee
	DH	ð	DH IY	thee
	F	f	F IY	fee
	G	g	G R IY N	green
	HH	h	HH IY	he
	JH	dʒ	JH IY	gee
	K	k	K IY	key
	L	l	L IY	lee
	M	m	M IY	me
	N	n	N IY	knee
	NG	ŋ	P IH NG	ping
	P	p	P IY	pee
	R	r	R IY D	read
	S	s	S IY	sea
	SH	ʃ	SH IY	she
	T	t	T IY	tea
	TH	θ	TH EY T AH	theta
	V	v	V IY	vee
	W	w	W IY	we
	Y	j	Y IY L D	yield
	Z	z	Z IY	zee
	ZH	ʒ	S IY ZH ER	seizure

Quadro 2.5 – Símbolos usados no dicionário eletrônico de pronúncia CMU.

2.6 Descrição do Buscador do CMU Pronouncing Dictionary – PUC/SP, LAEL, CEPRIL

Como o leitor provavelmente já deve ter percebido, os recursos de busca oferecidos pelo sítio da Carnegie Mellon University não seriam suficientes para pesquisar os grafemas coletados no trabalho de Lessa (1985). Como encontraríamos no CMU, por exemplo, todas as palavras que terminam com os grafemas <ough>, usando uma ferramenta de busca tão simples? Portanto, tivemos que desenvolver uma ferramenta de busca mais poderosa, que fosse capaz de pesquisar a correspondência grafofonêmica a partir de grafemas e não exclusivamente de palavras inteiras. Fazia-se necessário também controlar a posição desses grafemas na palavra (no início, no final, em qualquer posição etc.) e viabilizar buscas não somente pela presença de um grafema, mas também por sua ausência, por exemplo: palavras que contenham os grafemas finais <ew>, mas que não contenham o fonema final /u/ em sua transcrição, como em *sew* /sou/.

Assim, foi desenvolvida pelo Prof. Dr. Tony Berber Sardinha e por mim uma ferramenta de busca para que pudéssemos realizar o trabalho proposto nesta dissertação. Trata-se do Buscador do CMU Pronouncing Dictionary – PUC/SP, LAEL, CEPRIL, que nos permitiu transformar o dicionário eletrônico CMU em uma ferramenta capaz de manipular grandes quantidades de vocábulos com velocidade e precisão. Em realidade, trata-se da primeira ferramenta para estudo da relação grafema-fonema do inglês no mundo e pode ser acessada por qualquer pesquisador, pois seu acesso é livre e gratuito[49].

Abaixo segue a figura 2.2, que ilustra o funcionamento do Buscador do CMU (os números são ilustrativos). As opções apresentadas na figura demonstram a busca por ocorrências que

[49] Endereço na Internet: http://www2.lael.pucsp.br/corpora/cmu/index.html

primeiramente contenham <ough> no final das palavras e, entre essas, por ocorrências que contenham /f/ como fonema final.

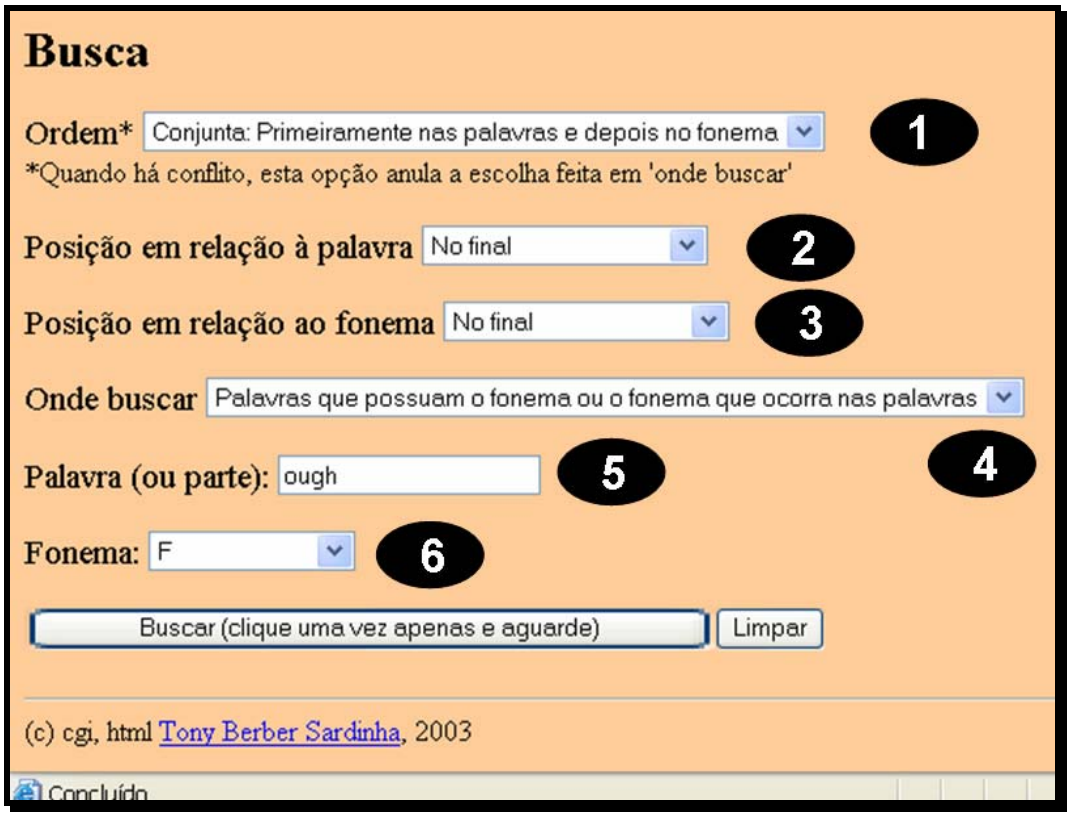

Figura 2.2 – Aspecto do Buscador do CMU – CEPRIL, LAEL, PUC/SP.

1. Ordem: o usuário define que tipo de pesquisa deseja fazer. Existem quatro opções:

   - ”Independente: somente nas palavras”
   - ”Independente: somente nos fonemas”
   - ”Conjunta: primeiramente nas palavras e depois nos fonemas”
   - ”Conjunta: primeiramente no fonema e depois nas palavras”

2. Posição em relação à palavra: o usuário define a posição dos grafemas (em nosso exemplo, <ough>) dentro da palavra:

   - ”Em qualquer posição”
   - ”No início”
   - ”Na segunda posição”
   - ”Antepenúltima”

- ”Penúltima”
- ”No final”

3. Posição em relação ao fonema: o usuário define a posição dos fonemas (/f/, em nosso exemplo), com as mesmas opções do item anterior.

4. Onde buscar: o usuário define que tipo de palavras será o foco da busca:

   - ”Palavras que possuam o fonema ou o fonema que ocorra nas palavras”
   - ”Ocorrências da palavra que não possuam o fonema”
   - ”Ocorrências do fonema que não apareçam na palavra”

5. Palavra (ou parte): introduzem-se uma palavra ou apenas grafemas. Em nosso exemplo, introduzimos os grafemas <ough>.

6. Fonema: escolhe-se da lista dos 39 fonemas aquele que fará parte da busca. Ou então, se for o caso, escolhe-se ”nenhum” para trabalhar apenas com os grafemas introduzidos no campo n.º 5.

Após clicar uma única vez em ”Buscar”, surge a tela de resultados exibida na figura 2.3, mostrando quantas palavras foram encontradas. A seguir, clica-se em ’’Resultados” para obter as palavras em ordem alfabética e suas transcrições, como exibido na figura 2.4.

2.7 Coleta das Freqüências de Uso no BNC

Seria um trabalho árduo tomar cada palavra resultante da pesquisa com o Buscador do CMU e encontrar suas respectivas freqüências de uso na lista de palavras do BNC escrito. Sem mencionar o fato de a possibilidade de ocorrerem inúmeros erros durante a cópia manual dos números ser realmente alta.

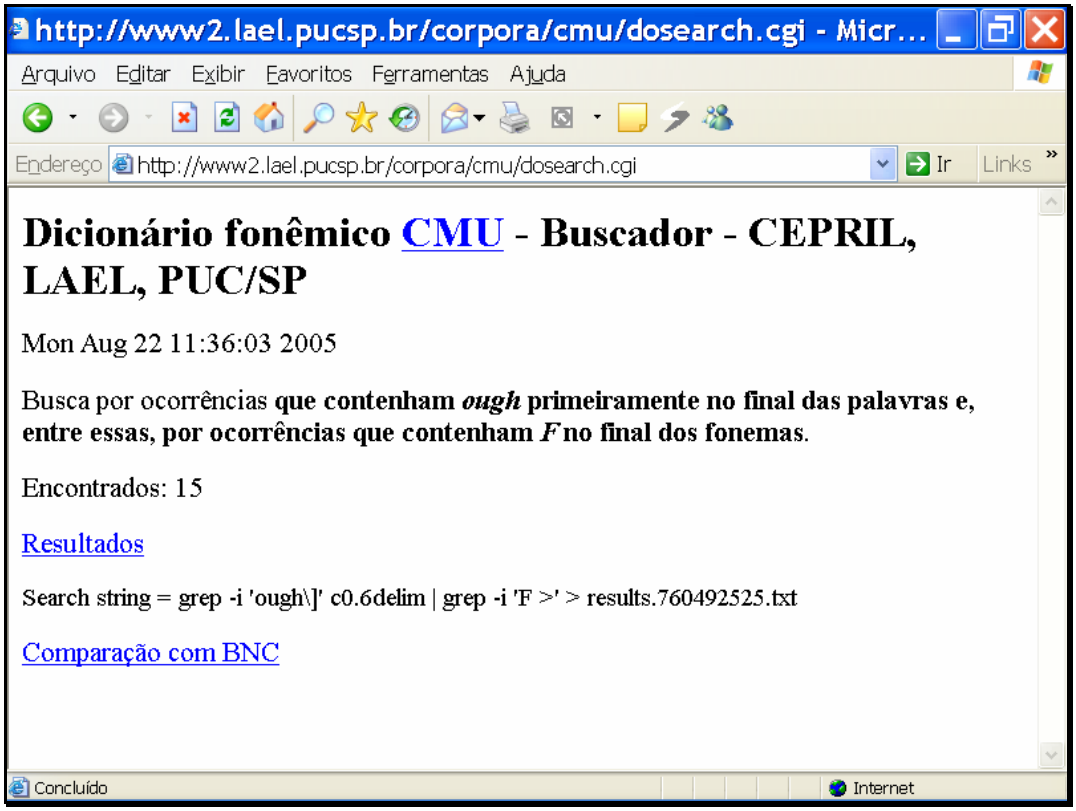

Figura 2.3 – Tela de resultados do Buscador CMU.

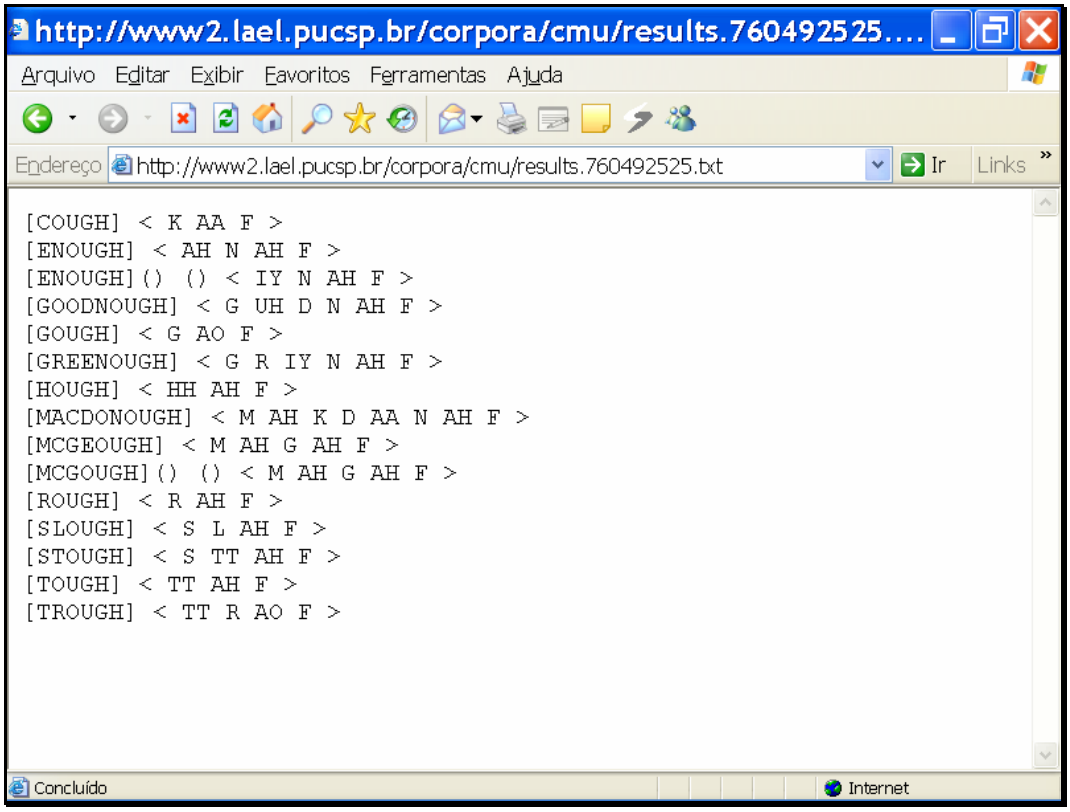

Figura 2.4 – Tela de vocábulos resultantes da pesquisa com o Buscador CMU.

Desenvolvemos, então, mais uma função para o Buscador do CMU: a de trazer os resultados da pesquisa com a freqüência de uso obtida da lista de palavras do BNC. Assim, como mostra a figura 2.3, o usuário pode, ao invés de clicar em ”Resultados” e obter as palavras com suas respectivas transcrições em ordem alfabética, clicar em ”Comparação com BNC” e obter as palavras com suas respectivas freqüências de uso extraídas do BNC, dispostas em ordem decrescente de freqüência, como mostra a figura 2.5. Observando a parte inferior da mesma figura, pode-se observar que o buscador fornece ainda a soma total das freqüências de uso e a média aritmética (soma das freqüências de uso dividida pelo número de palavras).

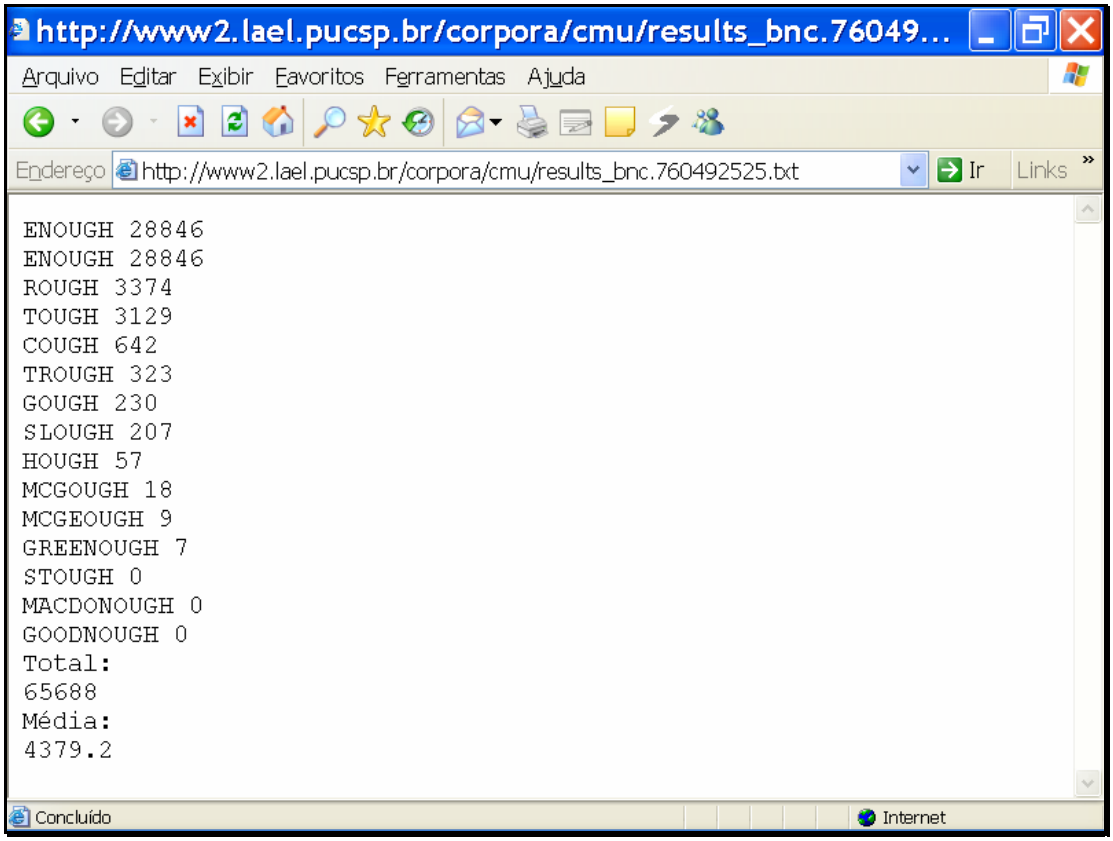

Figura 2.5 – Vocábulos do CMU e suas freqüências no BNC.

Caso a palavra tenha duas ou mais pronúncias, como no caso de *enough*[50], que tem duas pronúncias, o dicionário eletrônico CMU atribui

[50] As duas pronúncias de *enough* conforme o dicionário eletrônico CMU são /əˈnʌf/ e /ɪˈnʌf/.

a mesma freqüência de uso para todas. Pode-se, porém, identificar as pronúncias secundárias (menos comuns) no CMU através da marcação "( )", como apresentado na figura 2.4. Em nossa pesquisa, identificamos manualmente as pronúncias secundárias através das marcações e lhes atribuímos valor de freqüência de uso zero. Portanto, a freqüência do BNC atribuída somente à pronúncia principal. Futuramente, essa tarefa será automática.

2.8 Descrição do BNC (British National Corpus)

Lançado em 1995, o BNC é o resultado do esforço conjunto da Longman, Oxford University Press, Lancaster University e British Library. Possui em seus arquivos 100.106.008 de palavras, sendo 90% composto de inglês britânico de origem escrita, extraído de jornais de cobertura regional e nacional, periódicos especializados e publicações para todas as idades e interesses, livros acadêmicos e de ficção, cartas publicadas e não publicadas, memorandos, redações escolares e universitárias, dentre outros tipos de textos. Os 10% restantes são de transcrições de conversas informais não roteirizadas, gravadas por voluntários selecionados de diferentes idades, classes sociais e regiões do Reino Unido, de modo demograficamente balanceado. Há também transcrições de linguagem falada coletada em diferentes contextos, variando de reuniões formais de negócios ou com membros do governo a programas de rádio.

Nossa escolha baseou-se no fato de o BNC incluir em seus arquivos mais de 100 milhões de palavras. Levando em conta também seus critérios de coleta de textos, podemos considerá-lo como representativo da língua inglesa britânica. Outro ponto está no fato de as listas de palavras do BNC escrito e oral estarem disponíveis para *download* na Internet[51].

[51] Disponível na Internet no endereço: http://www.lexically.net/wordsmith/index.html.

Em nossa pesquisa, utilizamos a lista de palavras do BNC escrito. Apesar de estarmos trabalhando com o aspecto oral do inglês, não usamos o corpus falado, haja vista que nosso ponto de partida é a palavra escrita. Além disso, o corpus falado representa apenas 10% do BNC e certamente não mede as freqüências de uso das palavras mais comuns do inglês escrito com precisão. Usar ambos teria sido trabalhoso sem trazer grandes acréscimos à pesquisa.

2.9 Inglês Americano (CMU) e Inglês Britânico (BNC)

O leitor pode estar agora se questionando se não houve incoerência em usar em nosso trabalho o inglês americano proveniente do dicionário eletrônico CMU e o inglês britânico do BNC. A resposta é não. Estamos analisando a língua inglesa em geral, e não apenas o inglês americano ou o inglês britânico ou o inglês de qualquer outra origem. Portanto, não precisamos nos ater apenas ao inglês britânico ou inglês americano.

Lessa (1985) incluiu duas palavras que seguem a ortografia britânica: *draught* e *gaol*, que na ortografia americana são grafadas como *draft* e *jail* (Longman, 2003). *Draught* está presente no CMU, e os grafemas <aught> foram plenamente pesquisáveis. *Gaol*, entretanto, não figura no CMU. Pode-se dizer, portanto, que praticamente não houve incompatibilidade entre os grafemas coletados do trabalho de Lessa, o dicionário eletrônico americano CMU e o British National Corpus. Todas as palavras e grafemas interagiram em perfeita harmonia, ou seja, o que resultava do trabalho de Lessa era pesquisável no Buscador do CMU, o que resultava do Buscador do CMU era pesquisável no BNC. Sendo assim, a questão inglês britânico x inglês americano não trouxe entrave ou inconsistência à pesquisa.

2.10 Análise das Correspondências

Nosso primeiro passo, após coletar todas as realizações fonêmicas e freqüências de uso de uma seqüência de grafemas, foi de identificar

as correspondências que continham os vocábulos oriundos do trabalho de Lessa (1985). Partimos do princípio de que todas palavras contidas no trabalho de Lessa têm uma correspondência grafofonêmica atípica. Na figura 2.6, que mostra os resultados em relação aos grafemas <-ount->, a correspondência de Lessa é identificada pela letra L.

A grande maioria das palavras tem uma correspondência grafofonêmica mais freqüente, típica ou menos marcada, que chamamos em nossa investigação de correspondência-padrão. Consideramos como correspondência-padrão aquela de maior soma de freqüência de uso no BNC, excluindo a correspondência de Lessa. No exemplo abaixo, AW N T é a correspondência-padrão identificada com a letra P.

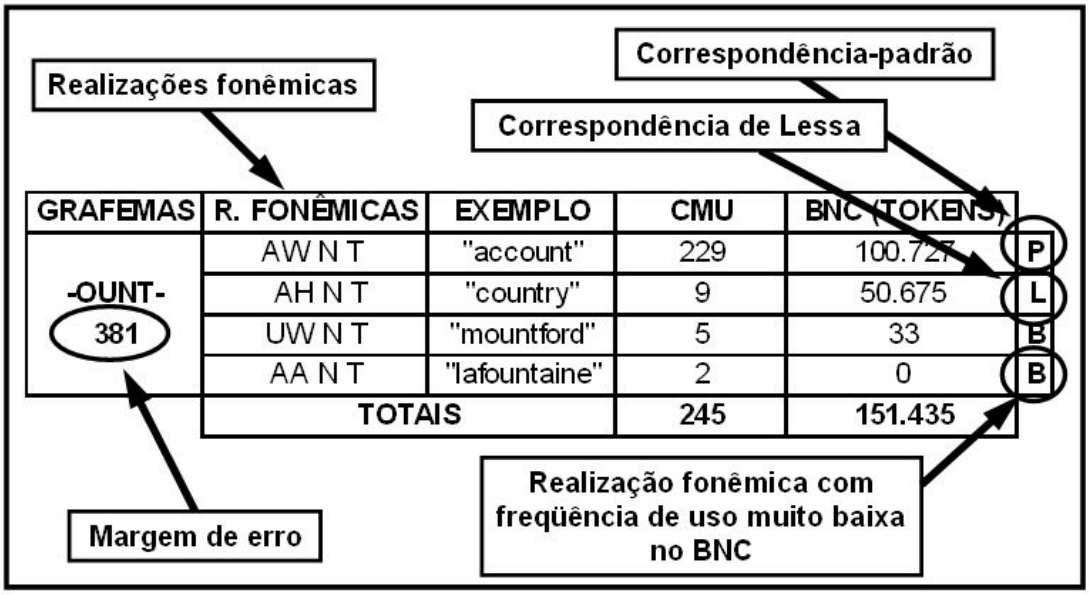

Figura 2.6 – Modelo da apresentação dos resultados.

Como era de se esperar, além de ser a correspondência que acumula maior freqüência de uso (*tokens*), a correspondência-padrão é, na maioria dos casos, a que também acumula maior número de palavras no léxico (CMU).

Nosso trabalho, entretanto, buscou focalizar-se mais no que está fora do padrão, no que é atípico e pode causar dificuldade. Em nossa análise, eliminamos a correspondência-padrão para focalizarmos nossa

análise no que é atípico. Essa eliminação, porém, não quer dizer que assumimos que a correspondência-padrão nunca cause confusão na conversão grafofonêmica; na grande maioria dos grafemas estudados nesta dissertação, ela realmente não causa. Todavia, há casos, como o de <leo-> [52] em que a correspondência-padrão também requer maior atenção. Quando for esse o caso, incluímos comentários na análise.

Um grafema poderia parecer mais complexo do que realmente é, se incluíssemos em nossa análise realizações fonêmicas que na verdade não têm freqüência de uso relevante na língua. Para evitar isso, calculamos a margem de erro para a soma das freqüências de uso do grafema, por meio da Calculadora de Erro Amostral PUC/SP, LAEL, CEPRIL, calculadora residente na Internet, de acesso gratuito no endereço http://www2.lael.pucsp.br/corpora/ea/index.html. Assim as realizações que ficaram abaixo da margem de erro foram identificadas com a letra B e também foram eliminadas.

Ainda sobre o modelo da figura 2.6, o vocábulo presente no campo ”exemplo” é o representante da correspondência grafofonêmica com maior freqüência de uso no BNC. Assim, em nosso exemplo acima, *account* é a palavra de maior freqüência de uso no BNC, que contém os grafemas <ount> em qualquer posição com realização fonêmica AW N T, *country* é a palavra de maior freqüência de uso no BNC com a realização fonêmica AH N T e assim por diante.

2.11 Identificação dos Vocábulos e dos Grafemas mais Atípicos

Para responder a outra pergunta de pesquisa, fazia-se necessário saber até qual freqüência de uso no BNC deveríamos classificar um vocábulo como relevante. Qual a freqüência de uso limite para um vocábulo ser considerado freqüente ou infreqüente? Um vocábulo com freqüência 20.000 deve ser considerado freqüente ou não? E outro com freqüência 1.000? Para responder tais questões, aplicamos o cálculo da

[52] Ver seção 3.2.8

margem de erro em relação à soma total das freqüências de uso do grafema. Os vocábulos com freqüência inferior à margem de erro não entraram na relação de palavras com correspondência grafofonêmica atípica.

Após termos identificado a correspondência de Lessa, eliminado a correspondência-padrão e eliminado as realizações fonêmicas com freqüência muito baixa, colocamos os resultados na tabela 3.71. Multiplicamos o número de realizações fonêmicas pelo respectivo número de *tokens* no BNC e dividimos por 1.000 para termos uma medida normalizada (Biber, Conrad & Reppen 1998:263). Essa medida privilegia os grafemas que têm mais realizações fonêmicas relevantes (acima da margem de erro) e maior freqüência de uso na língua, revelando, portanto, qual deles tem maior complexidade em termos de correspondência grafofonêmica à luz da freqüência de uso, respondendo, assim, a pergunta de pesquisa b), quais são os grafemas que exibem maior atipicidade grafofonêmica do ponto de vista léxico-freqüencial?

Após relacionarmos todas as palavras, limpamos os dados, ocultando palavras derivadas. Por exemplo, ao invés de citar *original* e *originally*, citamos apenas *original*. Fizemos isso com o intuito de compactar os resultados.

Tendo, portanto, apresentado a metodologia de pesquisa usada em nosso trabalho, passamos ao capítulo seguinte, o qual mostrará os resultados obtidos e suas respectivas análises.

Capítulo 3 – Apresentação e Análise dos Resultados

What learner has met all the words he or she will ever need to pronounce?

Dickerson (1985:303)

Apresentamos neste capítulo os resultados de nossa pesquisa, bem como suas respectivas análises. O leitor tem a sua disposição, no CD-ROM que acompanha esta dissertação, as listagens completas com os vocábulos e suas respectivas freqüências de uso retiradas do BNC.

3.1 Resultados que não exibem inconsistência.

Iniciamos com os resultados menos complexos, de grafemas que não apresentaram inconsistência na relação grafema-fonema. Os resultados mostram apenas uma correspondência grafofonêmica relevante.

3.1.1 <-aol>

Para os grafemas finais <aol>, oriundos da palavra *gaol*, que segue a ortografia britânica, não encontramos nenhuma palavra no CMU, haja vista que a grafia correspondente em inglês americano é *jail*.

3.1.2 <-cial>

Abaixo apresentamos os resultados referentes à seqüência de grafemas finais <cial>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-CIAL	SH AH L	"social"	31	112.165	L
328	S IY AA L	"marcial"	1	1	B
	TOTAIS		32	112.166	

Tabela 3.1- Resultados referentes aos grafemas finais <cial>.

Encontramos apenas um único vocábulo (*marcial*) com realização fonêmica diferente da realização fonêmica de Lessa, porém com a baixíssima freqüência de uso de 1.

3.1.3 <-igm>

Abaixo apresentamos os resultados referentes à seqüência de grafemas finais <igm>:

GRAFEMAS	R. FONÊMICA	EXEMPLO	CMU	BNC (TOKENS)	
-IGM	AY M	"paradigm"	1	675	L
	TOTAIS		1	675	

Tabela 3.2- Resultados referentes aos grafemas finais <igm>.

Nossa pesquisa revelou apenas uma realização fonêmica e apenas um vocábulo com a seqüência de grafemas <igm> em posição final de palavra. Não havendo, por conseguinte, inconsistência.

3.1.4 <-ism>

A seguir, apresentamos os resultados referentes à seqüência de grafemas finais <ism>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-ISM	Z AH M	"criticism"	277	41.176	L
199	Z M	"athleticism"	5	114	B
	TOTAIS		282	41.290	

Tabela 3.3- Resultados referentes aos grafemas finais <ism>.

Além da correspondência de Lessa, encontramos apenas mais uma realização fonêmica, com freqüência de uso inferior à margem de erro, sendo, portanto, eliminada.

3.1.5 <-ous>

A seguir, apresentamos os resultados referentes à seqüência de grafemas finais <ous>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-OUS 392	AH S	"various"	383	159.940	L
	UW	"rendezvous"	1	287	B
	UW Z	"sous"	2	57	B
	AW S	"milhous"	1	4	B
	AW Z	"thous"	1	2	B
	AO S	"chavous"	2	0	B
	UW S	"lajous"	1	0	B
	IY S	"brocious"	1	0	B
	TOTAIS		392	160.290	

Tabela 3.4- Resultados referentes aos grafemas finais <ous>.

Os resultados mostram que os grafemas acima têm um número elevado de realizações fonêmicas (oito). Contudo, sete delas apresentam freqüência de uso muito baixa, abaixo da margem de erro, sendo, portanto, eliminadas.

3.1.6 <gn->

A seguir, apresentamos os resultados referentes à seqüência de grafemas iniciais <gn>:

GRAFEMAS	R. FONÊMICA	EXEMPLO	CMU	BNC (TOKENS)	
GN-	N	"gnarled"	24	664	L
	TOTAIS		24	664	

Tabela 3.5- Resultados referentes aos grafemas iniciais <gn>.

Apenas uma realização fonêmica foi encontrada para esses grafemas, não havendo, portanto, inconsistência na correspondência grafofonêmica.

3.1.7 <kn->

Apresentamos os resultados referentes à seqüência de grafemas iniciais <kn>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
KN-	N	"know"	209	169.561	L
404	K N	"Knutson"	3	4	B
	K AH N	"Knievel"	2	3	B
	TOTAIS		214	169.568	

Tabela 3.6- Resultados referentes aos grafemas iniciais <kn>.

Além da correspondência de Lessa, encontramos mais duas realizações fonêmicas, porém ambas com freqüência de uso abaixo da margem de erro, sendo, portanto, eliminadas.

Os resultados para os grafemas iniciais <gn> e <kn> comprovam a regra pedagógica de Kriedler (1972) *apud* Celce-Murcia, Brinton & Goodwin (1996:280), que diz que quando houver uma consoante inicial precedendo imediatamente o grafema <n>, deve-se ignorar essa consoante inicial e simplesmente pronunciar /n/, como em *gnaw* /nɔ/, *knapsack* /ˈnæpsæk/, *mnemonic* /nəˈmɑnɪk/ e *pneumonia* /nuˈmoʊniə/. Existem, contudo, nomes próprios que não seguem este padrão (*Knievel* /kəˈnivəl/, por exemplo), porém com freqüência de uso muito baixa de acordo com o BNC.

3.2 Resultados com Seleção de Vocábulos.

3.2.1 <-aid>

A seguir, apresentamos os resultados referentes à seqüência de grafemas finais <aid>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-AID	EH D	"said"	4	181.622	L
460	EY D	"paid"	36	38.870	P
	TOTAIS		40	220.492	

Tabela 3.7- Resultados referentes aos grafemas finais <aid>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	SAID	S EH D	181.340

Tabela 3.8 - Vocábulos selecionados referentes aos grafemas finais <aid>.

Os resultados oriundos do BNC foram totalmente contrários à nossa expectativa. O léxico leva-nos a crer que a realização fonêmica EY D de *paid* seria a mais relevante que EH D, de *said* e *plaid* por haver mais palavras com EY D (36) do que com EH D (4), de acordo com o CMU. Porém, o BNC mostra que EH D tem maior freqüência de uso, basicamente devido a *said* (181.340). Lessa (1985) deixa transparecer essa mesma visão ao incluir *plaid* em seus testes. Todavia, *plaid*, apresentou freqüência de uso de apenas 113, abaixo da margem de erro (460), não sendo, portanto, incluída em nossa relação final de vocábulos de correspondência ortografia-pronúncia atípica.

3.2.2 <-ange>

Os resultados a seguir referem-se à seqüência de grafemas <ange> posicionados em final de palavra.

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-ANGE	EY N JH	"change"	24	68.127	P
261	AH N JH	"orange"	1	2.511	L
	AE NG	"lange"	2	78	B
	AE N JH	"flange"	6	52	B
	AA N JH	"delagrange"	1	0	B
	TOTAIS		34	70.768	

Tabela 3.9 - Resultados referentes aos grafemas finais <ange>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	ORANGE	AO R AH N JH	2.511

Tabela 3.10 - Vocábulos selecionados referentes aos grafemas finais <ange>.

Na realidade, a própria correspondência-padrão deste grafema pode causar confusão, por ser comumente substituída por AE N JH[53], como em *flange*.

3.2.3 <-auge->

Seguem abaixo os resultados referentes à seqüência de grafemas <auge> posicionados em qualquer ponto da palavra:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-AUGE- 34	EY JH	"gauge"	4	1.125	L
	AO G	"auger"	4	56	P
	AO JH	"hauge"	6	19	B
	AW G	"haugen"	7	5	B
	TOTAIS		**21**	**1.205**	

Tabela 3.11 - Resultados referentes aos grafemas <auge> em qualquer posição.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	GAUGE	G EY JH	908
2	GAUGES	G EY JH AH Z	110
3	GAUGED	G EY JH D	104

Tabela 3.12 - Vocábulos selecionados referentes aos grafemas em qualquer posição <auge>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	GAUGE	G EY JH	908

Tabela 3.13 - Seleção final dos vocábulos com grafemas <auge> em qualquer posição.

3.2.4 <-bt->

A seguir, apresentamos os resultados referentes à seqüência de grafemas <bt> posicionados em qualquer ponto da palavra.

[53] Equivalente em IPA: /ændʒ/

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-BT-	MUDO	"doubt"	36	31.006	L
211	B	"obtained"	32	15.159	P
	TOTAIS		68	46.165	

Tabela 3.14 - Resultados referentes aos grafemas <bt> em qualquer posição.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	DOUBT	D AW T	11.550
2	DEBT	D EH T	5.447
3	UNDOUBTEDLY	AH N D AW T AH D L IY	2.335
4	DOUBTS	D AW T S	2.053
5	DEBTS	D EH T S	1.821
6	SUBTLE	S AH T AH L	1.763
7	DOUBTFUL	D AW T F AH L	1.229
8	DOUBTLESS	D AW T L AH S	866
9	DEBTOR	D EH T ER	781
10	DOUBTED	D AW T AH D	695
11	DEBTORS	D EH T ER Z	379
12	SUBTLY	S AH T AH L IY	354
13	UNDOUBTED	AH N D AW T AH D	287
14	SUBTLETY	S AH T AH L T IY	260
15	DOUBTING	D AW T IH NG	244
16	INDEBTED	IH N D EH T AH D	234

Tabela 3.15 - Vocábulos selecionados referentes aos grafemas <bt> em qualquer posição.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	DEBT	D EH T	5.447
2	DOUBT	D AW T	11.550
3	SUBTLE	S AH T AH L	1.763

Tabela 3.16 - Seleção final dos vocábulos com os grafemas <bt> em qualquer posição.

3.2.5 <-ear->

Seguem abaixo os resultados referentes à seqüência de grafemas <ear> posicionados em qualquer ponto da palavra:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-EAR-	IH R	"years"	262	360.571	P
748	ER	"early"	183	164.369	L
	EH R	"bear"	63	24.817	L
	AA R	"heart"	45	19.206	L
	IY ER	"nuclear"	13	10.179	
	IY R	"Shakespeare"	61	2.426	
	EYR	"menswear"	1	75	B
	AO R	"tearle"	2	1	B
	TOTAIS		630	581.644	

Tabela 3.17 - Resultados referentes aos grafemas <ear> em qualquer posição.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	EARLY	ER L IY	32.815
2	RESEARCH	R IY S ER CH	26.533
3	HEARD	HH ER D	17.803
4	EARLIER	ER L IY ER	15.590
5	HEART	HH AA R T	13.699
6	LEARNING	L ER N IH NG	8.937
7	EARTH	ER TH	8.762
8	NUCLEAR	N UW K L IY ER	8.393
9	LEARN	L ER N	7.465
10	SEARCH	S ER CH	7.190
11	BEAR	B EH R	5.281
12	LEARNED	L ER N D	5.259
13	WEARING	W EH R IH NG	4.798
14	WEAR	W EH R	4.366
15	EARNINGS	ER N IH NG Z	3.174
16	BEARING	B EH R IH NG	2.887
17	RESEARCHERS	R IY S ER CH ER Z	2.541
18	SEARCHING	S ER CH IH NG	2.205
19	EARL	ER L	2.146
20	EARNED	ER N D	2.049
21	EARLIEST	ER L IY AH S T	1.920
22	LEARNT	L ER N T	1.892
23	EARN	ER N	1.786
24	HEARTS	HH AA R T S	1.465
25	LINEAR	L IH N IY ER	1.397
26	SHAKESPEARE	SH EY K S P IY R	1.323
27	BEARS	B EH R Z	1.315
28	EARNING	ER N IH NG	1.118
29	SEARCHED	S ER CH T	1.094
30	RESEARCHER	R IY S ER CH ER	990
31	EARTH'S	ER TH S	838
32	HEARTED	HH AA R T AH D	760
33	PEARL	P ER L	754

Tabela 3.18 - Vocábulos selecionados referentes aos grafemas <ear> em qualquer posição.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	BEAR	B EH R	5.281
2	EARL	ER L	2.146
3	EARLY	ER L IY	32.815
4	EARN	ER N	1.786
5	EARTH	ER TH	8.762
6	HEARD	HH ER D	17.803
7	HEART	HH AA R T	13.699
8	LEARN	L ER N	7.465
9	LINEAR	L IH N IY ER	1.397
10	NUCLEAR	N UW K L IY ER	8.393
11	PEARL	P ER L	754
12	SEARCH	S ER CH	7.190
13	SHAKESPEARE	SH EY K S P IY R	1.323
14	WEAR	W EH R	4.366

Tabela 3.19 - Seleção final dos vocábulos com os grafemas <ear> em qualquer posição.

Esta é a combinação de grafemas com segundo maior número de realizações fonêmicas (9)[54], estando 6 delas acima da margem de erro.

Cremos, contudo, que a realização fonêmica EY R relativa à palavra *menswear* seja uma falha de digitação do dicionário eletrônico de pronúncia da Carnegie Mellon University, haja vista que em nenhum outro dicionário figura tal transcrição. Ao invés de Y, provavelmente deveria ter sido digitado H, uma letra que, no teclado, encontra-se exatamente abaixo da letra Y.

No CD-ROM anexo, apresentamos uma lista com 19 vocábulos (*rearmament, firearm, prearrange* etc.) que não foram incluídos na pesquisa por não fazerem parte de mesma sílaba. Neles, a combinação grafêmica <-ear-> ocorre por derivação ou composição.

3.2.6 <-ey>

Abaixo apresentamos os resultados referentes à seqüência de grafemas finais <ey>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-EY	EY	"they"	67	336.110	**P**
661	IY	"money"	1.753	119.117	**L**
	TOTAIS		**1.820**	**455.227**	

Tabela 3.20 - Resultados referentes aos grafemas finais <ey>.

Os vocábulos selecionados a partir dos dados acima estão na tabela 3.21 a seguir.

Percebe-se que a seqüência de grafemas acima é particularmente importante para a pronúncia correta de antropônimos (nomes de pessoas), como Ashley, Bailey, Shelley etc. e topônimos (nomes de lugares), tais como Wembley, New Jersey, Sydney etc.

A correspondência-padrão destes grafemas também pode causar confusão por causa da generalização. Os que conhecem a pronúncia de

[54] A seqüência <-our-> na seção 3.2.14 tem 10 realizações fonêmicas, o maior número encontrado em nossa pesquisa.

money, *journey* e *valley* podem generalizá-la em relação a *survey* /ˈsɜrveɪ/, por exemplo, pronunciando erroneamente */ˈsɜrvi/.

	Vocábulo	Transcrição CMU	Freq. BNC
1	MONEY	M AH N IY	31.442
2	KEY	K IY	12.190
3	JOURNEY	JH ER N IY	4.609
4	VALLEY	V AE L IY	4.550
5	TURKEY	T ER K IY	1.948
6	ABBEY	AE B IY	1.783
7	SURREY	S ER IY	1.599
8	GEOFFREY	JH EH F R IY	1.512
9	HONEY	HH AH N IY	1.466
10	STANLEY	S T AE N L IY	1.245
11	HARVEY	HH AA R V IY	1.187
12	ATTORNEY	AH T ER N IY	1.137
13	JERSEY	JH ER Z IY	1.054
14	SHELLEY	SH EH L IY	1.013
15	SYDNEY	S IH D N IY	981
16	WEMBLEY	W EH M B L IY	945
17	BAILEY	B EY L IY	941
18	ASHLEY	AE SH L IY	929
19	STOREY	S T AO R IY	872
20	JOCKEY	JH AA K IY	739
21	CHIMNEY	CH IH M N IY	682
22	WESLEY	W EH S L IY	661

Tabela 3.21 - Vocábulos selecionados referentes aos grafemas finais <ey>.

3.2.7 <h->

A seguir, os resultados referentes ao grafema inicial <h>:

GRAFEMA	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
H-	pronunciado	"he"	6.031	4.160.644	**P**
210	mudo	"hours"	150	45.938	**L**
	TOTAIS		**6.181**	**4.206.582**	

Tabela 3.22 - Resultados referentes ao grafema inicial <h>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	HOURS	AW ER Z	17.069
2	HOUR	AW ER	11.142
3	HONEST	AA N AH S T	2.359
4	HEIR	EH R	1.018
5	HONESTLY	AA N AH S T L IY	992
6	HERBS	ER B Z	849
7	HONORARY	AA N ER EH R IY	741
8	HONESTY	AA N AH S T IY	689
9	HOMAGE	AA M AH JH	448
10	HERB	ER B	400
11	HEIRS	EH R Z	373
12	HOURLY	AW R L IY	340
13	HERBAL	ER B AH L	236

Tabela 3.23 - Vocábulos selecionados referentes ao grafema inicial <h>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	HEIR	EH R	1.018
2	HERB	ER B	400
3	HOMAGE	AA M AH JH	448
4	HONEST	AA N AH S T	2.359
5	HONORARY	AA N ER EH R IY	741
6	HOUR	AW ER	11.142

Tabela 3.24 - Seleção final dos vocábulos com o grafema inicial <h>.

3.2.8 <LEO->

Abaixo estão os resultados referentes à seqüência de grafemas iniciais <leo>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
LEO-66	L EH	"leonard"	22	1.511	L
	L IY AH	"leonora"	9	1.332	P
	L IY OW	"leo"	9	1.189	
	L IY AA	"leon"	4	531	
	L EH OW	"leoni"	1	18	B
	L IH OW	"leo"	1	2	B
	L IH OY	"leoine"	1	0	B
	TOTAIS		**47**	**4.583**	

Tabela 3.25 - Resultados referentes aos grafemas iniciais <leo>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	LEO	L IY OW	821
2	LEONARD	L EH N ER D	806
3	LEON	L IY AA N	496
4	LEOPARD	L EH P ER D	244
5	LEONE	L IY OW N	210
6	LEONARD'S	L EH N ER D Z	186
7	LEONIE	L EH N IY	161
8	LEOPARDS	L EH P ER D Z	95
9	LEOMINSTER	L IY OW M IH N S T ER	71
10	LEO'S	L IY OW Z	50

Tabela 3.26 - Vocábulos selecionados referentes aos grafemas iniciais <leo>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	LEO	L IY OW	821
2	LEONARD	L EH N ER D	806
3	LEON	L IY AA N	496
4	LEOPARD	L EH P ER D	244
5	LEONE	L IY OW N	210
6	LEONIE	L EH N IY	161
7	LEOMINSTER	L IY OW M IH N S T ER	71

Tabela 3.27 - Seleção final dos vocábulos com os grafemas iniciais <h>.

Percebe-se que a seqüência de grafemas acima é particularmente importante para a pronúncia correta de antropônimos, como *Leonard*, *Leonora*, *Leopold*, *Leone*, *Leo* etc.

Na realidade, a própria correspondência-padrão deste grafema pode causar confusão, por ser comumente substituída por L IY OW[55], como em *Leo*.

3.2.9 <-oe>

Abaixo seguem os resultados referentes à seqüência de grafemas finais <oe>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-OE	OW	"Joe"	90	8.299	P
99	UW	"shoe"	8	1.712	L
	OW IY	"Zoe"	2	270	
	TOTAIS		100	10.281	

Tabela 3.28 - Resultados referentes aos grafemas finais <oe>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	SHOE	SH UW	1.149
2	CANOE	K AH N UW	374
3	ZOE	Z OW IY	197
4	HORSESHOE	HH AO R S SH UW	163

Tabela 3.29 - Vocábulos selecionados referentes aos grafemas finais <oe>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	CANOE	K AH N UW	374
2	SHOE	SH UW	1.149
3	ZOE	Z OW IY	197

Tabela 3.30 - Seleção final dos vocábulos com os grafemas finais <oe>.

3.2.10 <-omb>

Os resultados referentes à seqüência de grafemas finais <omb> são apresentados a seguir.

[55] Equivalente a /ˈlioʊ/ em IPA.

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-OMB 68	AA M	"bomb"	8	3.242	P
	UW M	"tomb"	3	1.043	
	OW M	"comb"	4	475	L
	AH M	"titcomb"	21	44	B
	AO M	"edgecomb"	2	0	B
	TOTAIS		**38**	**4.804**	

Tabela 3.31 - Resultados referentes aos grafemas finais <omb>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	COMB	K OW M	413
2	TOMB	T UW M	637
3	WOMB	W UW M	402

Tabela 3.32 - Vocábulos selecionados referentes aos grafemas finais <omb>.

A correspondência-padrão deste grafema, AA M, também pode causar confusão, por ser comumente substituída por AO M B[56].

3.2.11 <or->

Os resultados a seguir referem-se à seqüência de grafemas iniciais <or>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
OR- 673	AO R	"or"	351	454.549	P
	ER	"original"	18	16.673	L
	OW R	"Orion"	3	185	B
	AA R	"oratorio"	1	41	B
	TOTAIS		**373**	**471.448**	

Tabela 3.33 - Resultados referentes aos grafemas iniciais <or>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	ORIGINAL	ER IH JH AH N AH L	10.914
2	ORIGINALLY	ER IH JH AH N AH L IY	4.179

Tabela 3.34 - Vocábulos selecionados referentes aos grafemas iniciais <or>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	ORIGINAL	ER IH JH AH N AH L	10.914

Tabela 3.35 - Seleção final dos vocábulos com os grafemas iniciais <or>.

[56] Equivalente em IPA: /ɔmb/

3.2.12 <-ough>

Abaixo seguem os resultados referentes à seqüência de grafemas finais <ough>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-OUGH 443	OW	"although"	29	88.077	L
	UW	"through"	3	77.178	P
	AH F	"enough"	12	35.704	L
	AW	"clough"	38	1.893	
	AO F	"trough"	2	737	
	AA F	"cough"	1	724	L
	AH	"McCollough"	3	1	B
	AWG	"keough"	1	0	B
	TOTAIS		**89**	**204.314**	

Tabela 3.36 - Resultados referentes aos grafemas finais <ough>.

Os vocábulos selecionados a partir dos dados acima estão na tabela a seguir:

	Vocábulo	Transcrição CMU	Freq. BNC
1	ALTHOUGH	AO L DH OW	42.032
2	THOUGH	DH OW	40.633
3	ENOUGH	AH N AH F	28.856
4	ROUGH	R AH F	3.414
5	TOUGH	T AH F	3.142
6	BOROUGH	B ER OW	1.753
7	THOROUGH	TH ER OW	1.122
8	COUGH	K AA F	724
9	CLOUGH	K L AW	589
10	PLOUGH	P L AW	560
11	SCARBOROUGH	S K AA R B ER OW	536
12	PETERBOROUGH	P IY T ER B ER OW	531
13	TROUGH	T R AO F	484

Tabela 3.37 - Vocábulos selecionados referentes aos grafemas finais <ough>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	ALTHOUGH	AO L DH OW	42.032
2	BOROUGH	B ER OW	1.753
3	CLOUGH	K L AW	589
4	COUGH	K AA F	724
5	ENOUGH	AH N AH F	28.856
6	PLOUGH	P L AW	560
7	ROUGH	R AH F	3.414
8	THOROUGH	TH ER OW	1.122
9	THOUGH	DH OW	40.633
10	TOUGH	T AH F	3.142
11	TROUGH	T R AO F	484

Tabela 3.38 - Seleção final dos vocábulos com os grafemas finais <ough>.

Esta seqüência de grafemas tem um número elevado de realizações fonêmicas (8), estando seis delas acima da margem de erro, com palavras gramaticais de alta freqüência, tais como *though* e *although*. Por outro lado, não esperávamos a inclusão de *clough* nem de *trough*.

3.2.13 <-ount->

Apresentamos abaixo os resultados referentes à seqüência de grafemas <ount> em qualquer posição:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-OUNT- 381	AW N T	"account"	229	100.727	P
	AH N T	"country"	9	50.675	L
	UW N T	"mountford"	5	33	B
	AA N T	"lafountaine"	2	0	B
	TOTAIS		245	151.435	

Tabela 3.39 - Resultados referentes aos grafemas <ount> em qualquer posição.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	COUNTRY	K AH N T R IY	26.936
2	COUNTRIES	K AH N T R IY Z	16.230
3	COUNTRYSIDE	K AH N T R IY S AY D	3.596
4	COUNTRY'S	K AH N T R IY Z	3.489

Tabela 3.40 - Vocábulos selecionados referentes aos grafemas <ount> em qualquer posição.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	COUNTRY	K AH N T R IY	26.936

Tabela 3.41 - Seleção final dos vocábulos com os grafemas <ount> em qualquer posição.

3.2.14 <-our->

Apresentamos os resultados referentes à seqüência de grafemas <our> em qualquer posição na palavra:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-OUR- 698	AO R	"your"	172	285.346	P
	AW ER	"our"	23	118.220	
	ER	"yourself"	118	56.497	L
	AW R	"hourly"	38	24.683	
	UH R	"tour"	101	22.382	
	AH R	"cherbourg"	2	72	B
	UW R	"kourou"	12	20	B
	W AA R	"jouret"	2	0	B
	AA R	"our" (2)	4	0	B
	OW UH R	"Kouri"	2	0	B
	TOTAIS		474	507.220	

Tabela 3.42 - Resultados referentes aos grafemas <our> em qualquer posição.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	OUR	AW ER	82.024
2	LABOUR	L EY B AW R	24.295
3	HOURS	AW ER Z	17.069
4	HOUR	AW ER	11.142
5	YOURSELF	Y ER S EH L F	9.142
6	TOUR	T UH R	6.385
7	ENCOURAGE	EH N K ER IH JH	4.840
8	JOURNEY	JH ER N IY	4.609
9	ENCOURAGED	EH N K ER IH JH D	4.496
10	OURSELVES	AW ER S EH L V Z	3.823
11	COLOURED	K AH L ER D	3.392
12	YOURS	Y UH R Z	3.232
13	ENCOURAGING	EH N K ER IH JH IH NG	2.732
14	JOURNAL	JH ER N AH L	2.335
15	COURAGE	K ER AH JH	2.012
16	HARBOUR	HH AA R B ER	1.985
17	TOURIST	T UH R AH S T	1.938
18	JOURNALISTS	JH ER N AH L AH S T S	1.761
19	TOURNAMENT	T UH R N AH M AH N T	1.634
20	TOURISTS	T UH R IH S T S	1.453
21	TOURISM	T UH R IH Z AH M	1.434
22	ENCOURAGEMENT	EH N K ER IH JH M AH N T	1.427
23	JOURNALIST	JH ER N AH L AH S T	1.356
24	RUMOURS	R UW M ER Z	1.292
25	BOURGEOIS	B UH R ZH W AA	1.105
26	OURS	AW ER Z	1.062
27	TOURS	T UH R Z	1.055
28	COURTESY	K ER T AH S IY	1.036
29	JOURNALS	JH ER N AH L Z	1.022
30	FLOUR	F L AW ER	999
31	ARMOUR	AA R M ER	973
32	TOURING	T UH R IH NG	821
33	ENCOURAGES	EH N K ER IH JH AH Z	790
34	JOURNEYS	JH ER N IY Z	732

Tabela 3.43 - Vocábulos selecionados referentes aos grafemas <our> em qualquer posição.

Após a exclusão dos vocábulos derivados da tabela acima, temos os vocábulos apresentados na tabela 3.44 a seguir.

Esta é a seqüência de grafemas com maior número de realizações fonêmicas, 10, estando 5 acima da margem de erro. Contudo, gostaríamos de chamar a atenção do leitor para a palavra *labour*, na segunda posição na tabela 3.43. O dicionário eletrônico CMU provê duas transcrições para essa palavra: L EY B AW R, como a pronúncia principal

e L EY B ER como secundária. Cremos que L EY B ER deveria ser tratada como a pronúncia principal: tanto o dicionário MacMillan (Rundell, 2002) quanto o Longman Dictionary of Contemporary English (Longman, 2003) referendam nossa posição. Certamente, houve uma falha na descrição dessa palavra no dicionário CMU.

	Vocábulo	Transcrição CMU	Freq. BNC
1	ARMOUR	AA R M ER	973
2	BOURGEOIS	B UH R ZH W AA	1.105
3	COLOURED	K AH L ER D	3.392
4	COURAGE	K ER AH JH	2.012
5	COURTESY	K ER T AH S IY	1.036
6	FLOUR	F L AW ER	999
7	HARBOUR	HH AA R B ER	1.985
8	HOUR	AW ER	11.142
9	JOURNAL	JH ER N AH L	2.335
10	JOURNEY	JH ER N IY	4.609
11	LABOUR	L EY B AW R	24.295
12	OUR	AW ER	82.024
13	RUMOURS	R UW M ER Z	1.292
14	TOUR	T UH R	6.385
15	YOURS	Y UH R Z	3.232
16	YOURSELF	Y ER S EH L F	9.142

Tabela 3.44 - Seleção final dos vocábulos com os grafemas <our> em qualquer posição (continuação).

3.2.15 <p->

Abaixo seguem os resultados referentes ao grafema inicial <p>:

GRAFEMA	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
P-	pronunciado	"people"	7.369	3.511.060	**P**
1.840	mudo	"psychological"	119	13.409	**L**
	TOTAIS		**7.488**	**3.524.469**	

Tabela 3.45 - Resultados referentes ao grafema inicial <p>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	PSYCHOLOGICAL	S AY K AH L AA JH IH K AH L	2.747
2	PSYCHOLOGY	S AY K AA L AH JH IY	2.393
3	PSYCHIATRIC	S AY K IY AE T R IH K	1.078
4	PSYCHOLOGISTS	S AY K AA L AH JH AH S T S	854
5	PSYCHIC	S AY K IH K	484
6	PSYCHOLOGIST	S AY K AA L AH JH AH S T	474
7	PSEUDO	S UW D OW	464
8	PNEUMONIA	N UW M OW N Y AH	431
9	PSYCHOANALYSIS	S AY K OW AH N AE L AH S AH S	368
10	PSYCHIATRIST	S AH K AY AH T R AH S T	353
11	PSYCHIATRISTS	S AH K AY AH T R AH S T S	322

Tabela 3.46 - Vocábulos selecionados referentes ao grafema inicial <p>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	PNEUMONIA	N UW M OW N Y AH	431
2	PSALM	S AA L M	228
3	PSEUDO	S UW D OW	464
4	PSI	S AY	149
5	PSYCHE	S AY K IY	242
6	PSYCHIATRY	S AY K AY AH T R IY	209
7	PSYCHIC	S AY K IH K	484
8	PSYCHO	S AY K OW	228

Tabela 3.47 - Seleção final dos vocábulos com o grafema inicial <p>.

3.2.16 <-reign->

A seguir, apresentamos os resultados referentes à seqüência de grafemas <reign> posicionados em qualquer ponto da palavra:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-REIGN-	R AH N	"foreign"	7	19.863	**P**
146	EY N	"reign"	4	2.342	**L**
	R N	"foreigner"(2)	3	0	**B**
	TOTAIS		**14**	**22.205**	

Tabela 3.48 - Resultados referentes aos grafemas <reign> em qualquer posição.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	REIGN	R EY N	1.856
2	REIGNING	R EY N IH NG	184
3	REIGNED	R EY N D	151
4	REIGNS	R EY N Z	151

Tabela 3.49 - Vocábulos selecionados referentes aos grafemas <reign> em qualquer posição.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	REIGN	R EY N	1.856

Tabela 3.50 - Seleção final dos vocábulos com os grafemas <reign> em qualquer posição.

3.2.17 <-uce>

Abaixo seguem os resultados referentes à seqüência de grafemas finais <uce>.

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
	UW S	"produce"	22	25.534	**P**
-UCE	AO S	"sauce"	3	1.351	
162	AH S	"lettuce"	1	365	**L**
	OW S IY	"beauce"	1	3	**B**
	UW CH IY	"bonaduce"	1	0	**B**
	TOTAIS		**28**	**27.253**	

Tabela 3.51 - Resultados referentes aos grafemas finais <uce>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	SAUCE	S AO S	1.350
2	LETTUCE	L EH T AH S	365

Tabela 3.52 - Vocábulos selecionados referentes aos grafemas finais <uce>.

3.2.18 <-ury>

Abaixo seguem os resultados referentes à seqüência de grafemas finais <ury>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-URY 204	ER IY	"century"	20	35.073	P
	EH R IY	"Canterbury"	48	4.642	L
	UH R IY	"jury"	7	3.395	
	UW R IY	"drury"	7	137	B
	AO R IY	"maury"	5	24	B
	AW R IY	"Khoury"	2	7	B
	AH R IY	"beury"	1	0	B
	TOTAIS		**90**	**43.278**	

Tabela 3.53 - Resultados referentes aos grafemas finais <ury>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	JURY	JH UH R IY	2.066
2	CANTERBURY	K AE N TT ER B EH R IY	1.147
3	FURY	F Y UH R IY	1.120
4	BURY	B EH R IY	837
5	SHREWSBURY	SH R UW Z B EH R IY	521
6	BANBURY	B AE N B EH R IY	468
7	NEWBURY	N UW B EH R IY	328
8	CADBURY	K AE D B EH R IY	303
9	SAINSBURY	S EY N S B EH R IY	297

Tabela 3.54 - Vocábulos selecionados referentes aos grafemas finais <ury>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	BURY	B EH R IY	837
2	FURY	F Y UH R IY	1.120
3	JURY	JH UH R IY	2.066

Tabela 3.55 - Seleção final dos vocábulos com os grafemas finais <ury>.

3.2.19 <-ute>

Os resultados referentes à seqüência de grafemas finais <ute> são apresentados a seguir.

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-UTE 190	UW T	"institute"	63	29.113	P
	AH T	"minute"	3	8.121	
	OW T	"haute"	2	161	B
	AW T	"stoute"	2	32	B
	AO T	"saute"	2	22	B
	TOTAIS		**72**	**37.449**	

Tabela 3.56 - Resultados referentes aos grafemas finais <ute>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	MINUTE	M IH N AH T	8.121

Tabela 3.57 - Vocábulo selecionado referente aos grafemas finais <ute>.

3.3 Resultados que requereram ajustes.

Abaixo seguem seis seqüências de grafemas, cujos resultados necessitaram de algum tipo de ajuste para tornar a análise mais precisa e focalizada nas questões de natureza grafofonêmica, e menos sensível à questões articulatórias.

3.3.1 <-age>

A seguir, apresentamos os resultados referentes à seqüência de grafemas finais <age>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-AGE 437	AH JH	"language"	65	73.192	L
	IH JH	"village"	132	63.732	L
	EY JH	"age"	38	57.323	P
	AA ZH	"garage"	13	4.299	
	AA JH	"fuselage"	4	519	
	TOTAIS		**252**	**199.065**	

Tabela 3.58 - Resultados referentes aos grafemas finais <age>.

Os vocábulos selecionados a partir dos dados acima estão apresentados na tabela a seguir.

	Vocábulo	Transcrição CMU	Freq. BNC
1	LANGUAGE	L AE NG G W AH JH	18.406
2	VILLAGE	V IH L IH JH	10.687
3	AVERAGE	AE V ER IH JH	9.453
4	DAMAGE	D AE M AH JH	8.098
5	MARRIAGE	M EH R IH JH	7.695
6	IMAGE	IH M AH JH	7.682
7	ADVANTAGE	AE D V AE N T IH JH	7.018
8	MESSAGE	M EH S AH JH	6.561
9	PACKAGE	P AE K IH JH	5.707
10	ENCOURAGE	EH N K ER IH JH	4.840
11	PASSAGE	P AE S AH JH	3.928
12	MANAGE	M AE N AH JH	3.588
13	COTTAGE	K AA T AH JH	3.022
14	STORAGE	S T AO R AH JH	2.890
15	PERCENTAGE	P ER S EH N T IH JH	2.609
16	MORTGAGE	M AO R G IH JH	2.534
17	COVERAGE	K AH V ER AH JH	2.132
18	COURAGE	K ER AH JH	2.012
19	CARRIAGE	K AE R IH JH	1.914
20	HERITAGE	HH EH R AH T AH JH	1.899
21	GARAGE	G ER AA ZH	1.625
22	SHORTAGE	SH AO R T AH JH	1.417
23	USAGE	Y UW S AH JH	1.134
24	SAVAGE	S AE V IH JH	1.130
25	DISADVANTAGE	D IH S AH D V AE N T IH JH	1.105
26	VOLTAGE	V OW L T AH JH	1.001
27	DRAINAGE	D R EY N AH JH	957
28	PATRONAGE	P AE T R AH N IH JH	887
29	VINTAGE	V IH N T IH JH	745
30	SEWAGE	S UW IH JH	721
31	FOLIAGE	F OW L IH JH	718
32	VOYAGE	V OY AH JH	703
33	LUGGAGE	L AH G AH JH	676
34	MASSAGE	M AH S AA ZH	619
35	PILGRIMAGE	P IH L G R AH M AH JH	481
36	DISCOURAGE	D IH S K ER IH JH	479
37	BAGGAGE	B AE G AH JH	473
38	HOSTAGE	HH AA S T IH JH	463
39	HOMAGE	AA M AH JH	448
40	SAUSAGE	S AO S AH JH	446
41	BARRAGE	B ER AA ZH	438

Tabela 3.59 - Vocábulos selecionados referentes aos grafemas finais <age>.

Para processar os dados destes grafemas, consideramos as realizações fonêmicas com /ə/ (AH JH) e /ɪ/ (IH JH)[57] como um só padrão por serem muito semelhantes e a distinção entre elas não ser exatamente um erro de influenciado pela ortografia. O Longman Dictionary of

[57] Equivalentes a /ədʒ/ e /ɪdʒ/, em IPA.

Contemporary English (2003:contracapa) adota um sinal especial, / ɪ̯ /, que indica que alguns falantes usam /ə/, enquanto outros usam /ɪ/.

Cabe dizer que nos surpreende a correspondência mais freqüente serem as correspondências de Lessa (AH JH e IH JH), e não EY JH. As correspondências de Lessa supostamente deveriam ter menos *tokens* e menos *types*, trazendo, portanto, maior dificuldade na determinação da correspondência grafofonêmica por parte do indivíduo brasileiro que pronuncia a palavra. Elas, entretanto, têm cinco vezes mais *types* que EY JH, e uma freqüência de uso quase duas vezes e meia maior. Porém, é EY JH que, provavelmente, está mais na mente dos brasileiros. Talvez, isso se deva ao fato de *age*, *stage* e *page*[58] serem palavras muito freqüentes no BNC, respondendo sozinhas por quase um quarto (23,92%) do total de *tokens* deste grafema. Elas são não apenas freqüentes em termos de uso na língua, mas especialmente *page* é muito freqüente no discurso de sala de aula. Por isso Lessa (1985) incluiu três vocábulos com IH JH/AH JH (*heritage, sandage e sewage*) em sua pesquisa, e nenhum com EY JH.

3.3.2 <-aught>

A seguir, apresentamos os resultados referentes à seqüência de grafemas finais <aught>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-AUGHT 114	AA T	"caught"	1	8.234	P
	AO T	"taught"	10	4.749	P
	AE F T	"draught"	1	482	L
	TOTAIS		12	13.465	

Tabela 3.60 - Resultados referentes aos grafemas finais <aught>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	DRAUGHT	D R AE F T	482

Tabela 3.61 - Vocábulo selecionado referente aos grafemas finais <aught>.

[58] *Age*, *stage* e *page* ocupam o primeiro, terceiro e quinto lugares, respectivamente, como as palavras mais freqüentes com <age> finais.

Para processar os dados destes grafemas, consideramos as realizações fonêmicas AA T e AO T[59] como um só padrão por serem muito semelhantes e a distinção entre elas não ser exatamente um erro de influenciado pela ortografia.

3.3.3 <-ew>

Abaixo estão os resultados referentes à seqüência de grafemas finais <ew>.

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-EW 498	UW	"new"	79	172.736	P
	Y UW	"few"	57	85.030	P
	OW	"sew"	2	192	L/B
	TOTAIS		**138**	**257.958**	

Tabela 3.62 - Resultados referentes aos grafemas finais <ew>.

Consideramos as realizações fonêmicas UW e Y UW, como a correspondência-padrão, haja vista que a língua inglesa permite certa variação na pronúncia, dependendo da origem de inglês que se adota (britânico ou americano, por exemplo). Em inglês americano, a pronúncia de *new* é /nu/, enquanto que em inglês britânico, a pronúncia contém o *invisible Y*[60], tema do estudo de Dickerson (1985) e também presente em Celce-Murcia (1996:278), e é pronunciada /nju/.

O BNC surpreende-nos, revelando que a realização fonêmica OW, como em *sew*, presente no trabalho de Lessa (1985), não é tão freqüente quanto nossa intuição de professor não-nativo pode nos levar a crer.

3.3.4 <ex->

Abaixo seguem os resultados referentes à seqüência de grafemas iniciais <ex>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
EX- 689	IH K S	"experience"	256	274.856	P
	IH G Z	"example"	116	116.665	L
	EH K S	"extra"	232	91.460	P
	EH G Z	"existence"	21	10.664	L
	TOTAIS		**625**	**493.645**	

Tabela 3.63 - Resultados referentes aos grafemas iniciais <ex>.

[59] Equivalentes a /ɑt/ e /ɔt/, em IPA.

[60] Representado por /j/ em IPA.

Os vocábulos selecionados a partir dos dados acima encontram-se a seguir:

	Vocábulo	Transcrição CMU	Freq. BNC
1	EXAMPLE	IH G Z AE M P AH L	34.600
2	EXISTING	IH G Z IH S T IH NG	9.442
3	EXACTLY	IH G Z AE K T L IY	8.633
4	EXECUTIVE	IH G Z EH K Y AH T IH V	7.921
5	EXAMPLES	IH G Z AE M P AH L Z	6.848
6	EXISTENCE	EH G Z IH S T AH N S	6.577
7	EXIST	IH G Z IH S T	5.310
8	EXAMINATION	IH G Z AE M AH N EY SH AH N	4.997
9	EXAMINE	IH G Z AE M AH N	3.755
10	EXAMINED	IH G Z AE M AH N D	3.667
11	EXISTS	IH G Z IH S T S	3.082
12	EXISTED	IH G Z IH S T AH D	2.511
13	EXACT	IH G Z AE K T	2.163
14	EXAMINING	IH G Z AE M AH N IH NG	1.628
15	EXHAUSTED	IH G Z AO S T AH D	1.497
16	EXAMINATIONS	IH G Z AE M AH N EY SH AH N Z	1.407
17	EXECUTIVES	IH G Z EH K Y AH T IH V Z	1.328
18	EXIT	EH G Z AH T	1.210
19	EXOTIC	IH G Z AA T IH K	1.145
20	EXILE	EH G Z AY L	1.046
21	EXEMPTION	IH G Z EH M P SH AH N	925
22	EXAGGERATED	IH G Z AE JH ER EY T AH D	917
23	EXHIBIT	IH G Z IH B AH T	790
24	EXHIBITED	IH G Z IH B AH T AH D	729
25	EXEMPT	IH G Z EH M P T	707
26	EXAMINES	IH G Z AE M AH N Z	704

Tabela 3.64 - Vocábulos selecionados referentes aos grafemas iniciais <ex>.

Após a exclusão dos vocábulos derivados, temos:

	Vocábulo	Transcrição CMU	Freq. BNC
1	EXAMPLE	IH G Z AE M P AH L	34.600
2	EXACT	IH G Z AE K T	2.163
3	EXAGGERATED	IH G Z AE JH ER EY T AH D	917
4	EXAMINE	IH G Z AE M AH N	3.755
5	EXECUTIVE	IH G Z EH K Y AH T IH V	7.921
6	EXEMPT	IH G Z EH M P T	707
7	EXHAUSTED	IH G Z AO S T AH D	1.497
8	EXHIBIT	IH G Z IH B AH T	790
9	EXILE	EH G Z AY L	1.046
10	EXIST	IH G Z IH S T	5.310
11	EXIT	EH G Z AH T	1.210
12	EXOTIC	IH G Z AA T IH K	1.145

Tabela 3.65 - Seleção final dos vocábulos com os grafemas iniciais <ex>.

Nosso enfoque em relação a este grafema está na pronúncia do <x> como /g/ ou /k/. Por isso consideramos as duas realizações fonêmicas com /k/ como a correspondência-padrão.

3.3.5 <th->

Exibimos a seguir os resultados referentes à seqüência de grafemas iniciais <th>:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
TH-	DH	"the"	65	8.664.562	**P**
2.971	TH	"through"	611	516.489	**P**
	T	"Thomas"	32	13.162	**L**
	TOTAIS		**708**	**9.194.213**	

Tabela 3.66 - Resultados referentes aos grafemas iniciais <th>.

Os vocábulos selecionados a partir dos dados acima foram:

	Vocábulo	Transcrição CMU	Freq. BNC
1	THOMAS	T AA M AH S	6.345

Tabela 3.67 - Vocábulo selecionado referente aos grafemas finais <th>.

Consideramos as correspondências DH e TH como padrão pelo fato de não estarmos analisando a pronúncia do <th> em termos de /θ/ (como em *through*) e /ð/ (como em *that*) devido ao caráter articulatório dessa questão[61].

3.3.6 <-oup>

Os resultados referentes à seqüência de grafemas finais <oup> estão apresentados a seguir:

GRAFEMAS	R. FONÊMICAS	EXEMPLO	CMU	BNC (TOKENS)	
-OUP	UW P	"group"	20	40.057	**L**
200	UW	"coup"	1	1.792	
	TOTAIS		**21**	**41.849**	

Tabela 3.68 - Resultados referentes aos grafemas finais <oup>.

Os vocábulos selecionados a partir dos dados acima foram vêm a seguir.

[61] Ver quadro 2.3

	Vocábulo	Transcrição CMU	Freq. BNC
1	COUP	K UW	1.792
2	GROUP	G R UW P	38.286
3	SOUP	S UW P	1.230

Tabela 3.69 - Vocábulos selecionados referentes aos grafemas finais <oup>.

É quase seguro que Lessa escolheu o vocábulo *soup* pelo fato de freqüentemente haver confusão ao pronunciar *soup* e *soap*. O problema está mais na palavra *soup* do que na seqüência de grafemas como um todo. Assim, não podemos dizer que a realização fonêmica exibida em *coup* é a correspondência-padrão, devido a sua baixa freqüência, tanto no CMU quanto no BNC. Portanto, não designamos uma correspondência-padrão para essa seqüência de grafemas e incluímos *coup* em nossa relação de vocábulos, haja vista que essa palavra está acima da margem de erro, sendo inclusive mais freqüente que *soup*.

3.4 Relação Final de Vocábulos com Correspondência Grafofonêmica Atípica

A seguir, apresentamos a relação final de vocábulos com correspondência grafofonêmica atípica, ordenada por ordem decrescente de número de palavras por grafema.

		Grafemas	n.º	Vocábulo	Transcrição CMU	Freq. BNC
1	1	-AGE	1	LANGUAGE	L AE NG G W AH JH	18.406
2			2	VILLAGE	V IH L IH JH	10.687
3			3	AVERAGE	AE V ER IH JH	9.453
4			4	DAMAGE	D AE M AH JH	8.098
5			5	MARRIAGE	M EH R IH JH	7.695
6			6	IMAGE	IH M AH JH	7.682
7			7	ADVANTAGE	AE D V AE N T IH JH	7.018
8			8	MESSAGE	M EH S AH JH	6.561
9			9	PACKAGE	P AE K IH JH	5.707
10			10	ENCOURAGE	EH N K ER IH JH	4.840
11			11	PASSAGE	P AE S AH JH	3.928
12			12	MANAGE	M AE N AH JH	3.588
13			13	COTTAGE	K AA T AH JH	3.022
14			14	STORAGE	S T AO R AH JH	2.890
15			15	PERCENTAGE	P ER S EH N T IH JH	2.609
16			16	MORTGAGE	M AO R G IH JH	2.534
17			17	COVERAGE	K AH V ER AH JH	2.132
18			18	COURAGE	K ER AH JH	2.012
19			19	CARRIAGE	K AE R IH JH	1.914
20			20	HERITAGE	HH EH R AH T AH JH	1.899
21			21	GARAGE	G ER AA ZH	1.625
22			22	SHORTAGE	SH AO R T AH JH	1.417
23			23	USAGE	Y UW S AH JH	1.134
24			24	SAVAGE	S AE V IH JH	1.130

Tabela 3.70 – Relação de vocábulos com correspondência grafofonêmica atípica.

		Grafemas	n.º	Vocábulo	Transcrição CMU	Freq. BNC
25			25	DISADVANTAGE	D IH S AH D V AE N T IH JH	1.105
26			26	VOLTAGE	V OW L T AH JH	1.001
27			27	DRAINAGE	D R EY N AH JH	957
28			28	PATRONAGE	P AE T R AH N IH JH	887
29			29	VINTAGE	V IH N T IH JH	745
30			30	SEWAGE	S UW IH JH	721
31			31	FOLIAGE	F OW L IH JH	718
32	1	-AGE	32	VOYAGE	V OY AH JH	703
33			33	LUGGAGE	L AH G AH JH	676
34			34	MASSAGE	M AH S AA ZH	619
35			35	PILGRIMAGE	P IH L G R AH M AH JH	481
36			36	DISCOURAGE	D IH S K ER IH JH	479
37			37	BAGGAGE	B AE G AH JH	473
38			38	HOSTAGE	HH AA S T IH JH	463
39			39	HOMAGE	AA M AH JH	448
40			40	SAUSAGE	S AO S AH JH	446
41			41	BARRAGE	B ER AA ZH	438
42			1	MONEY	M AH N IY	31.442
43			2	KEY	K IY	12.190
44			3	JOURNEY	JH ER N IY	4.609
45			4	VALLEY	V AE L IY	4.550
46			5	TURKEY	T ER K IY	1.948
47			6	ABBEY	AE B IY	1.783
48			7	SURREY	S ER IY	1.599
49			8	GEOFFREY	JH EH F R IY	1.512
50			9	HONEY	HH AH N IY	1.466
51			10	STANLEY	S T AE N L IY	1.245
52	2	-EY	11	HARVEY	HH AA R V IY	1.187
53			12	ATTORNEY	AH T ER N IY	1.137
54			13	JERSEY	JH ER Z IY	1.054
55			14	SHELLEY	SH EH L IY	1.013
56			15	SYDNEY	S IH D N IY	981
57			16	WEMBLEY	W EH M B L IY	945
58			17	BAILEY	B EY L IY	941
59			18	ASHLEY	AE SH L IY	929
60			19	STOREY	S T AO R IY	872
61			20	JOCKEY	JH AA K IY	739
62			21	CHIMNEY	CH IH M N IY	682
63			22	WESLEY	W EH S L IY	661
64			1	ARMOUR	AA R M ER	973
65			2	BOURGEOIS	B UH R ZH W AA	1.105
66			3	COLOURED	K AH L ER D	3.392
67			4	COURAGE	K ER AH JH	2.012
68			5	COURTESY	K ER T AH S IY	1.036
69			6	FLOUR	F L AW ER	999
70			7	HARBOUR	HH AA R B ER	1.985
71	3	-OUR-	8	HOUR	AW ER	11.142
72			9	JOURNAL	JH ER N AH L	2.335
73			10	JOURNEY	JH ER N IY	4.609
74			11	LABOUR	L EY B AW R	24.295
75			12	OUR	AW ER	82.024
76			13	RUMOURS	R UW M ER Z	1.292
77			14	TOUR	T UH R	6.385
78			15	YOURS	Y UH R Z	3.232
79			16	YOURSELF	Y ER S EH L F	9.142

Tabela 3.70 – Relação de vocábulos com correspondência grafofonêmica atípica (continuação).

		Grafemas	n.º	Vocábulo	Transcrição CMU	Freq. BNC
80	4	-EAR-	1	BEAR	B EH R	5.281
81			2	EARL	ER L	2.146
82			3	EARLY	ER L IY	32.815
83			4	EARN	ER N	1.786
84			5	EARTH	ER TH	8.762
85			6	HEARD	HH ER D	17.803
86			7	HEART	HH AA R T	13.699
87			8	LEARN	L ER N	7.465
88			9	LINEAR	L IH N IY ER	1.397
89			10	NUCLEAR	N UW K L IY ER	8.393
90			11	PEARL	P ER L	754
91			12	SEARCH	S ER CH	7.190
92			13	SHAKESPEARE	SH EY K S P IY R	1.323
93			14	WEAR	W EH R	4.366
94	5	P-	1	PNEUMONIA	N UW M OW N Y AH	431
95			2	PSALM	S AA L M	228
96			3	PSEUDO	S UW D OW	464
97			4	PSI	S AY	149
98			5	PSYCHE	S AY K IY	242
99			6	PSYCHIATRY	S AY K AY AH T R IY	209
100			7	PSYCHIC	S AY K IH K	484
101			8	PSYCHO	S AY K OW	228
102	6	EX-	1	EXAMPLE	IH G Z AE M P AH L	34.600
103			2	EXACT	IH G Z AE K T	2.163
104			3	EXAGGERATED	IH G Z AE JH ER EY T AH D	917
105			4	EXAMINE	IH G Z AE M AH N	3.755
106			5	EXECUTIVE	IH G Z EH K Y AH T IH V	7.921
107			6	EXEMPT	IH G Z EH M P T	707
108			7	EXHAUSTED	IH G Z AO S T AH D	1.497
109			8	EXHIBIT	IH G Z IH B AH T	790
110			9	EXILE	EH G Z AY L	1.046
111			10	EXIST	IH G Z IH S T	5.310
112			11	EXIT	EH G Z AH T	1.210
113			12	EXOTIC	IH G Z AA T IH K	1.145
114	7	-OUGH	1	ALTHOUGH	AO L DH OW	42.032
115			2	THOUGH	DH OW	40.633
116			3	ENOUGH	AH N AH F	28.856
117			4	ROUGH	R AH F	3.414
118			5	TOUGH	T AH F	3.142
119			6	BOROUGH	B ER OW	1.753
120			7	THOROUGH	TH ER OW	1.122
121			8	COUGH	K AA F	724
122			9	CLOUGH	K L AW	589
123			10	PLOUGH	P L AW	560
124			11	TROUGH	T R AO F	484
125	8	LEO-	1	LEO	L IY OW	821
126			2	LEONARD	L EH N ER D	806
127			3	LEON	L IY AA N	496
128			4	LEOPARD	L EH P ER D	244
129			5	LEONE	L IY OW N	210
130			6	LEONIE	L EH N IY	161
131			7	LEOMINSTER	L IY OW M IH N S T ER	71
132	9	H-	1	HOUR	AW ER	11.142
133			2	HONEST	AA N AH S T	2.359
134			3	HEIR	EH R	1.018
135			4	HONORARY	AA N ER EH R IY	741
136			5	HOMAGE	AA M AH JH	448
137			6	HERB	ER B	400
138	10	-BT-	1	DEBT	D EH T	5.447
139			2	DOUBT	D AW T	11.550
140			3	SUBTLE	S AH T AH L	1.763

Tabela 3.70 – Relação de vocábulos com correspondência grafofonêmica atípica (continuação).

		Grafemas	n.º	Vocábulo	Transcrição CMU	Freq. BNC
141	11	-OE	1	SHOE	SH UW	1.149
142			2	CANOE	K AH N UW	374
143			3	ZOE	Z OW IY	197
144	12	-OMB	1	TOMB	T UW M	637
145			2	COMB	K OW M	413
146			3	WOMB	W UW M	402
147	13	-URY	1	JURY	JH UH R IY	2.066
148			2	FURY	F Y UH R IY	1.120
149			3	BURY	B EH R IY	837
150	14	-OUP	1	GROUP	G R UW P	38.286
151			2	COUP	K UW	1.792
152			3	SOUP	S UW P	1.230
153	15	-UCE	1	SAUCE	S AO S	1.350
154			2	LETTUCE	L EH T AH S	365
155	16	-AID	1	SAID	S EH D	181.340
156	17	-ANGE	1	ORANGE	AO R AH N JH	2.511
157	18	-AUGE	1	GAUGE	G EY JH	908
158	19	-AUGHT	1	DRAUGHT	D R AE F T	482
159	20	OR-	1	ORIGINAL	ER IH JH AH N AH L	10.914
160	21	-OUNT-	1	COUNTRY	K AH N T R IY	26.936
161	22	-REIGN-	1	REIGN	R EY N	1.856
162	23	TH-	1	THOMAS	T AA M AH S	6.345
163	24	-UTE	1	MINUTE	M IH N AH T	8.121

Tabela 3.70 – Relação de vocábulos com correspondência grafofonêmica atípica (continuação).

Nenhum vocábulo foi escolhido para as seguintes seqüências de grafemas: <aol>, <cial>, <igm>, <ism>, <ous>, <gn>, <kn> e <ew>.

3.5 Relação Final de Grafemas em Ordem Decrescente de Atipicidade

A seguir, na tabela 3.71, apresentamos a relação final de grafemas com correspondência grafofonêmica atípica em ordem decrescente de atipicidade.

A tabela 3.70 responde a pergunta de pesquisa sobre quais são os vocábulos que exibem uma relação atípica entre a ortografia e a pronúncia e que apresentam freqüência de uso relevante na língua inglesa. A tabela 3.71 responde a pergunta de pesquisa sobre quais são os grafemas ou seqüência de grafemas que exibem maior atipicidade do ponto de vista léxico-freqüencial.

	GRAFEMAS	R.FONÊMICAS. (1)	CMU	BNC (TOKENS) (2)	[(1) X (2)]/1000
1	**-EAR-**	5	365	220.997	11.050
2	**-OUR-**	4	280	221.782	8.871
3	**EX-**	3	369	218.789	6.564
4	**-OUGH**	5	82	127.135	6.357
5	**TH-**	1	643	529.651	5.297
6	**-AGE**	2	210	141.223	2.824
7	**-AID**	1	4	181.622	1.816
8	**-EY**	1	1.753	119.117	1.191
9	**-EW**	1	57	85.030	850
10	**-OUNT-**	1	9	50.675	507
11	**H-**	1	150	45.938	459
12	**-BT**	1	36	31.066	311
13	**OR-**	1	18	16.673	167
14	**-URY**	2	55	8.037	161
15	**P-**	1	119	13.409	134
16	**LEO-**	3	35	3.231	97
17	**-UTE**	1	3	8.121	81
18	**-OE**	2	10	1.982	40
19	**-UCE**	2	4	1.716	34
20	**-OMB**	2	7	1.518	30
21	**-ANGE**	1	1	2.511	25
22	**-REIGN-**	1	4	2.342	23
23	**-OUP**	1	1	1.792	18
24	**-AUGE-**	1	4	1.125	11
25	**-OUS**	2	3	344	7
26	**-AUGHT**	1	1	482	5
27	**-AOL**	0	0	0	0
28	**-CIAL**	0	0	0	0
29	**GN-**	0	0	0	0
30	**-IGM**	0	0	0	0
31	**-ISM**	0	0	0	0
32	**KN-**	0	0	0	0
	TOTAIS	**47**	**4.223**	**2.036.308**	**482,19**

Tabela 3.71 - Grafemas em ordem decrescente de atipicidade.

A seguir, passaremos às considerações finais.

Considerações Finais

Pour l'ortographe, mais contre la façon dont on l'enseigne ou plutôt dont on ne l'enseigne pas.

Maistre (1974:179)

O presente capítulo fecha nosso trabalho, retomando seus pontos principais, apontando limitações e fazendo sugestões de pesquisas futuras e possíveis aplicações pedagógicas dos resultados.

Conforme dito na Introdução, no Brasil, parece-nos que a maioria dos professores de inglês como língua estrangeira enfrenta problemas de pronúncia causados pela falta de formação na área e pela ortografia inglesa que pode conduzir a pronúncias errôneas.

A pesquisa aqui descrita buscou contribuir para a formação do professor brasileiro de inglês como língua estrangeira, estudando a correspondência grafofonêmica de alguns grafemas que podem causar dificuldades ao serem pronunciadas por falantes de português brasileiro. Buscamos também contribuir com informação que possa ser utilizada por elaboradores de material didático na criação de atividades que envolvam pronúncia.

Para tanto, nosso trabalho encontrou suporte teórico na Lingüística de Corpus, que é uma área que investiga a linguagem de modo empírico e objetivo, por meio de computadores, os quais analisam grandes amostras de linguagem armazenadas eletronicamente chamadas de corpora. Além da Lingüística de Corpus, fundamentamo-nos também nos princípios teóricos que dão suporte à correspondência grafofonêmica.

A investigação aqui relatada foi norteada pelas seguintes questões de pesquisa:

a) Com base nos grafemas presentes no trabalho de Lessa (1985), quais são os vocábulos que exibem uma relação atípica entre a ortografia e a pronúncia e que apresentam freqüência de uso relevante na língua inglesa?

b) Quais são os grafemas que exibem maior atipicidade do ponto de vista léxico-freqüencial?

A metodologia empregada na pesquisa consistiu na a) seleção de grafemas que causam dificuldades a falantes de português brasileiro ao pronunciar palavras em inglês, b) coleta no dicionário eletrônico de pronúncia CMU das palavras que contêm tais grafemas, c) coleta no corpus de inglês geral BNC das freqüências de uso de cada uma das palavras coletadas no CMU, d) análise e determinação dos grafemas mais atípicos e e) confecção de uma relação de palavras que apresentam correspondência grafofonêmica inconsistente, porém com freqüência de uso relevante.

Os resultados geraram como resposta à pergunta a) o quadro 3.70 e o quadro 3.71 como resposta à pergunta b), ambos expostos no capítulo 3.

Buscamos apresentar os resultados de maneira direta, ou seja, confeccionando uma relação de vocábulos que merecem maior atenção durante a formação do professor. Nossa intenção não foi criar uma lista de palavras para ser memorizada, mas sim mostrar quais são os vocábulos de correspondência grafofonêmica atípica e de uso freqüente na língua inglesa. Trata-se de uma lista gerativa, ou seja, os grafemas neles presentes participam de milhares de palavras que os professores encontrarão dentro e fora de sala de aula.

A relevância destes achados, discutida na Introdução, refere-se ao fato de buscarmos contribuir para que os professores sejam lingüisticamente competentes para ensinar seus alunos a se comunicarem sem causar distrações a seus ouvintes devido à pronúncia influenciada pela ortografia. Além disso, este trabalho visou mostrar aos elaboradores de material didático quais são os grafemas e vocábulos que requerem maior atenção em suas publicações.

Quisemos também chamar a atenção para a importância da inclusão da freqüência de uso nos estudos sobre pronúncia, especialmente no momento de decidir o que ensinar. Há vinte anos,

Lessa (1985:66) teve de selecionar os vocábulos que fariam parte de sua pesquisa com base em sua própria experiência. Por isso foram incluídos em seu trabalho alguns vocábulos que têm freqüência de uso muito baixa, tais como *sandage*, *furlough*, *slough*, *thyme* e *barley*, palavras estas que ficaram abaixo da margem de erro, motivo pelo qual cremos não ser necessário atribuir-lhes muita importância no processo de ensino e aprendizagem da pronúncia do inglês. Hoje, mais de vinte anos depois, dispomos de métodos empíricos, mais objetivos, falseáveis e replicáveis.

Nossa pesquisa também revelou que nem sempre a correspondência grafofonêmica mais freqüente no léxico da língua inglesa ou num corpus de inglês geral é também a mais freqüente para os não-nativos, como ficou claro em <-age>[62]. O que o corpus mostra deve ser analisado à luz de outras variáveis.

Os resultados de nossa pesquisa também revelaram que, em muitos casos, a inconsistência na relação grafofonêmica é apenas aparente, bastando conhecer algumas regras para dirimir as possíveis dúvidas sobre como pronunciar a palavra. São exemplos disso os grafemas <gn->, <kn-> e <-omb>, os quais não apresentaram inconsistência em nossa pesquisa. Trata-se mais de uma questão de falta de treinamento do que de falta de transparência na língua-alvo.

Morley (1991:495) também chama a atenção para a importância de conhecer algumas regras de correspondência grafofonêmica, mostrando que a ortografia é uma ferramenta-chave para o ensino da pronúncia. Para dominar a correspondência grafofonêmica do inglês, faz-se necessário treinar o olho, e não apenas o ouvido (Murphy, 1991:60; O'Connor, 1967:1; Kiran, Tuchtenhagen & Spelman, 2003:139). Falta ao professor de inglês brasileiro não nativo um conhecimento maior sobre a relação entre a escrita e a pronúncia do inglês, para que ele consiga deduzir, por meio da ortografia, a

[62] Ver seção 3.3.1

pronúncia das palavras com as quais ainda não está familiarizado, servindo assim como um bom modelo de pronúncia para seus alunos. Neste trabalho, proporcionamos resultados que ajudam nessa direção.

Ao iniciarmos esta pesquisa, esperávamos que a reforma ortográfica do inglês fosse quase um consenso entre os envolvidos no ensino e aprendizagem do inglês. Críamos que um sistema do tipo um grafema para representar um fonema fosse a melhor solução para eliminar a confusão na área grafofonêmica tanto para falantes não-nativos como quanto para nativos. Porém, após nossa investigação, passamos a concordar com Venezky (1970) que uma reforma ortográfica não pode ocultar as raízes morfológicas dos vocábulos na forma escrita.

O presente trabalho possui algumas limitações. Não entramos no mérito de *como* aprimorar o ensino da pronúncia a partir da forma escrita no processo de formação de professores. Porém, Celce-Murcia (1996:283) mostra algumas maneiras de ensinar e aprender a pronúncia do inglês. Algumas maneiras já têm uso há mais tempo, outras são mais recentes, a saber:

a) Treinamento fonético;
b) Gravação da produção oral em áudio ou vídeo;
c) Leitura em voz alta;
d) Recursos audiovisuais, como figuras explicativas, fotos, DVD, CD-ROM e outros;
e) Ouvir e imitar;
f) Exercícios com pares mínimos: *bit* x *beat*.

Idealmente, gostaríamos de analisar todas as seqüências de grafemas presentes em Lessa (1985), porém isso estaria fora do escopo de uma dissertação de mestrado. Desenvolvendo, porém, ferramentas mais poderosas, abrir-se-ão as portas para trabalhos ainda mais aprofundados e ainda mais abrangentes. Ferramentas que, por

exemplo, tivessem códigos que representassem o conjunto das vogais e das consoantes, ou ainda, que incluíssem a soma das freqüências de uso das formas lematizadas das palavras em estudo. Por exemplo, não somando apenas a freqüência de *bury*, mas sim as de *bury, buries*, *burying* e *buried*. Isso poderá ser incorporado em versões futuras do buscador do dicionário eletrônico CMU.

Ainda em relação à ferramenta de busca no CMU, ela poderia também incluir mais combinações de busca, como por exemplo duas ou três opções de localização dos grafemas ou fonemas ao mesmo tempo. Isso agilizaria a pesquisa de seqüências menores de grafemas ou até mesmo apenas um grafema ou apenas um fonema, porque se tornaria mais fácil precisar sua posição na palavra.

Outra sugestão para trabalhos futuros seria a de realizar a pesquisa usando a freqüência de uso fornecida por corpora de inglês geral de diferentes variantes de inglês (inglês americano, inglês britânico, inglês canadense, inglês australiano, inglês sul-africano etc). Isso daria um caráter mais internacional à pesquisa. O problema certamente está em ter acesso a esses corpora.

Pode-se também pesquisar como utilizar no processo de formação de professores a relação de vocábulos de correspondência ortografia-pronúncia atípica apresentada nesta dissertação.

Não temos conhecimento sobre estudos realizados com aprendizes brasileiros de inglês para medir a sensibilidade destes ao contexto grafêmico.

Esperamos que o trabalho aqui apresentado, envolvendo a Lingüística de Corpus e estudos sobre correspondência grafofonêmica, possa ser de auxílio para a formação de professores e elaboração de material didático. A pesquisa de mestrado que desenvolvemos nos mostrou que há ainda vários aspectos que precisamos abordar em relação ao ensino de pronúncia do inglês para brasileiros. Acima de

tudo, esta pesquisa que desenvolvemos nos ensinou o valor do ato de pesquisar e quanto ainda precisamos saber sobre esse aspecto tão importante da formação do professor de inglês como língua estrangeira que é à pronúncia e sua relação com a ortografia.

Referências Bibliográficas

Intelligible pronunciation is an essential component of communicative competence.

Morley (1991:488)

Agard, F. B. (1969). *The Sounds of English and Italian*: A Systematic Analysis of the Contrasts between the Sound Systems. Chicago: University of Chicago Press.

Almeida Filho, J. C. P. & Schmitz, J. (1998). *Glossário de Lingüística Aplicada*. Campinas, SP: Pontes.

Atechi, S. N. (2004). The intelligibility of native and non-native English speech: a comparative analysis of Cameroon English and American and British English. Dissertação para obtenção do grau de doutor em filosofia. Alemanha: Universidade Técnica de Chemnitz.

Bahns, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21 (1), 101-114.

Baker, M. (1995). Corpora in translation studies: an overview and some suggestions for future research. *Target,* 7, 223-243. John Benjamins.

Bamgbose, A. (1998). Torn between the norms: innovation in world Englishes. *World Englishes*, 17 (1), 1-14.

Berber Sardinha, A. P. (2000). Lingüística de Corpus: histórico e problemática. *D.E.L.T.A.,* 16 (2), 323-367.

_____. (2004). *Lingüística de Corpus*. São Paulo: Manole.

_____. (2004b). Um Quadro Teórico e Prático para Produção de Atividades Didáticas com Corpora Eletrônicos para Ensino de Inglês como Língua Estrangeira. *14º InPLA. LAEL, PUC-SP*.

_____. (2005). Ver a Língua Portuguesa no Computador. In Berber Sardinha (org.), *A Língua Portuguesa no Computador*. São Paulo: Mercado das Letras.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics:* Investigating Language Structure and Use. Cambridge: Cambridge University Press.

Brezolin, A., Allegro, A. L. V., & Campos, R. M. (2001). *Pequeno Dicionário de Expressões Idiomáticas e Coloquialismos.* São Paulo.

British National Corpus (BNC). Disponível na Internet no endereço: http://www.natcorp.ox.ac.uk. Acessado em 15 de março de 2004.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, 22 (4), 593-606.

Brown, G. (1995). *Speakers, Listeners and Communication*. Cambridge: Cambridge University Press.

Capovilla, F. C., Capovilla, A. G. S. & Macedo, E. C. (2001). Rota perilexical na leitura em voz alta: tempo de reação, duração e segmentação na pronúncia. *Psicologia: Reflexão e Crítica*, 14 (2), 409-427. Disponível na Internet no endereço: http://www.scielo.br/scielo.php?script=sciarttext&pid=S0102 -79722001000 200015&lng=en&nrm=iso. Acessado em 9 de setembro de 2005.

Celce-Murcia, M., Brinton, D. & Goodwin, J. M. (1996). *Teaching Pronunciation*: A Reference for Teachers of English to Speakers of Other Languages. Cambridge: Cambridge University Press.

Celce-Murcia, M. & Goodwin, J. M. (1991). Teaching pronunciation. In Celce-Murcia. M. (ed). *Teaching English as a second or foreign language* (pp. 136-153). New York: Newbury House.

Chomsky, N. (1957). *Syntactic Strutures*. The Hague. Mouton.

Connelly, V. (2002). Graphophonemic awareness in adults after instruction in phonic generalisations. *Learning and Instruction,* 12, 627-649.

Coulmas, F. (2000). *The Writing Systems of the World*. Oxford Blackwell.

Crystal, D. (1997). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

Deschamps, A., Fournier, J., Duchet, J. & O'Neil, M. (2004). *English Phonology and Graphophonemics*. Paris: Ophrys.

D'Eugenio, A. (1982). *Major Problems of English Phonology*. Foggia, Itália: Atlântica.

Dickerson, W. (1975). The wh-question of pronunciation: an answer from spelling and generative phonology. *TESOL Quarterly*, 9 (3), 299-309.

Dickerson, W. B. (1985). The invisible Y: a case for spelling in pronunciation learning. *TESOL Quarterly*, 19 (2), 303-317.

Ferreiro, E. & Teberosky, A. (1988). *Los Sistemas de Escritura en el Desarrollo del Niño*. México: Siglo XXI Editores.

Firth, J. R. (1957). *Papers in Linguistics* – 1934-1951. Oxford: Oxford University Press.

Firth, J. R. (1957b). A synopsis of linguistic theory – 1930-1955. In *Studies in Linguistic Analysis*, pp. 1-32. Oxford: Philological Society.

Fox, G. (1998). Using corpus data in the classroom. In Tomlinson, B. (org.), *Materials Development in Language Teaching*, pp. 25-43. Cambridge: Cambridge University Press.

Gama-Rossi, A. J. A. & Almeida, S. S. (2004). Reavaliação de resultados experimentais sobre a fonotaxe do português brasileiro: transições entre fones e grau de aceitabilidade em logatomas. *Intercâmbio*, vol. XIII. São Paulo.

Granger, S. (2002). A bird's-eye view of learner corpus research. In Granger, S., Hung, J. & Petch-Tyson, S. (orgs.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, (pp. 3-33). Amsterdam: John Benjamins.

Guimarães, S. R. K. (2002). Dificuldades no desenvolvimento da lectoescrita: o papel das habilidades metalingüísticas. *Psicologia*: *Teoria e Pesquisa*, 18 (3), 247-259.

Hanna, P. R., Hanna, J. S., Hodges, R. E. & Rudorf, E. H. (1966). *Phoneme–Grapheme Correspondences as Cues to Spelling Improvement*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Hewings, M. & Goldstein, S. (1998). *Pronunciation Plus*: Practice through Interaction – North American English. Cambridge: Cambridge University Press.

Hoey, M. (1997). From concordance to text: new uses for computer corpora. In Lewandowska-Tomaszczyk, B. & Melia, P. J. (orgs.), *PALC '97*, *Practical Applications in Language Corpora*. Lodz: Lodz University Press.

Houaiss, A., Villar, M. S. & Franco, F. M. M. (2004). *Dicionário Houaiss da Língua Portuguesa.* Rio de Janeiro: Objetiva.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. & Francis, G. (1999). *Pattern Grammar*: A Corpus-Driven Approach to the Lexical Grammar of English. Amsterdam: John Benjamins Publishing Company.

James, C. (1998). *Errors in Language Learning and Use*: Exploring Error Analysis. London: Longman.

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford University Press.

_____. (2003). *World Englishes:* A Resource Book for Students. London: Routledge.

_____. (2004). ELF at the gate: the position of English as a lingua franca. In Pulverness, A. (org.), *Liverpool Conference Selections*. IATEFL Publications.

Johns, T. (1994). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (org.), *Perspectives on Pedagogical Grammar* (pp. 293-312). New York: Cambridge University Press.

Kato, M., Moreira, N. & Tarallo, F. (1998). *Estudos em Alfabetização*: Retrospectivas nas Áreas de Psico e da Sociolingüística. Campinas: Pontes Editores.

Katsiavriades, K. & Qureshi, T. (2005). *The KryssTal*. The Origin and History of the English Language. Disponível na Internet no endereço: http://www.krysstal.com/english.html. Acessado em 10 de março de 2005.

Kessler, B. & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, 37, 592-617.

_____. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, 44, 592-617.

Kiran, S., Tuchtenhagen, J. & Spelman, C. (2003). Effect of training phoneme to grapheme conversion in improving written and oral deficits. *Brain and Language*, 87 (1), 139-141.

Kjellmer, G. (1992). Grammatical or Nativelike? In Leitner, G. (org.), *New Directions in English Language Corpora*: Methodology, Results, Software Developments (pp. 329-344). Berlin: Mouton de Gruyter.

Kreidler, C. W. (1999). *The Pronunciation of English*: A Coursebook in Phonology. Oxford: Blackwell Publishers.

Kriedler, C. (1972). Teaching English spelling and pronunciation. *TESOL Quarterly*, 5 (1), 3-12.

Krishnamurthy, R. (1997). Keeping good company: collocation, corpus and dictionaries. In *Lexic, Corpus I Diccionaris*: Cicle de Conferencies 95-96, IULA – Institut Iniversitari de Lingüístca Aplicada, Universitat Pompeu Fabra, Barcelona, Spain, pp. 31-56.

Leech, G. (1992). Corpora and theories of linguistic performance. In Svartik, J. (org.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991.* Berlin, New York: De Gruyter.

Lessa, A. B. C. T. (1985). A ortografia como um fator de interferência da pronúncia do inglês como língua estrangeira. Dissertação de mestrado. São Paulo: Programa de Lingüìstica Aplicada e Estudos da Linguagem, PUC-SP.

Lewis, M. (1996). *The Lexical Approach*: The State of ELT and a Way Forward. Hove: LTP.

Lieff, C. D. & Nunes, Z. A. (1993). English pronunciation and the Brazilian learner: how to cope with language transfer. *Speak Out!*, 12, 22-27.

Linell, P. (1983). *The Written Language Bias in Linguistics*. Linköping: University of Linköping Studies in Communication.

Llol, M. P. (1999). Análisis de errores grafemáticos en textos libres de estudiantes de enseñanzas medias. Tese de doutorado. Universidad de Barcelona.

Longman Dictionary of Contemporary English (2003). New edition. London: Longman.

Lopes, E. (1987). *Fundamentos da Lingüística Contemporânea*. São Paulo: Cultrix.

Louw, B. (1993). Irony in the text or insincerity in the writer: the diagnostic potential of semantic prosodies. In Baker, M., Francis, G. & Tognini-Bonelli, E. (orgs.), *Text and Technology: Essays in Honor of John Sinclair*. Amsterdã/Atlanta: John Benjamins.

Luria, A. R. (2001). *Pensamento e Linguagem*: As Últimas Conferências de Luria. São Paulo: Artmed Editora.

Maistre, M. de. (1974). *Pour ou contre L'orthographe*?. Paris: Editions Universitaires.

Martinet, A. (1971). *Elementos de Lingüística Geral*. Lisboa: Sá da Costa.

Massini-Cagliari, G. & Cagliari, L. (2004). Fonética. In Mussalim, F. & Bentes, A. C. (orgs.), *Introdução à Lingüística: Domínios e Fronteiras*, vol. 1. São Paulo: Cortez.

McCarthy, M. (2001). *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.

McEnery, T. & Wilson, A. (1997). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Medgyes, P. (1994). *The Non-Native Teacher*. London: MacMillan.

Monaghan, J. (1979). *The Neo-Firthian Tradition and its Contribution to General Linguistics*. Tübingen: Max Niemeyer Verlag.

Morais, J. (1994). *A Arte de Ler*. São Paulo: editora da UNESP.

Mori, A. C. (2004). Fonologia. In Mussalim, F. & Bentes, A. C. (orgs.), *Introdução à Lingüística: Domínios e Fronteiras*, vol. 1. São Paulo: Cortez.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25 (3), 481-520.

Murphy, J. M. (1991). Oral communication in TESOL: integrating speaking, listening, and pronunciation. *TESOL Quarterly*, 25 (1), 51-75.

Nelson, M. (2005). *Semantic Associations in Business English:* A Corpus-Based Analysis. Finland: University of Turku, no prelo.

O'Connor, J. D. (1967). *Better English Pronunciation:* Language and Speech. Cambridge: Cambridge University Press.

Olso, D. (1994). *The World on Paper:* The Conceptual and Cognitive Implications of Writing and Reading. Cambridge: Cambridge University Press.

Parish, C. (1977). A practical philosophy of pronunciation. *TESOL Quarterly*, 11 (3), 311-317.

Pennington, M. C. & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly*, 20 (2), 207-225.

Popper, K. (1968). *The Logic of Scientific Discovery*. New York: Harper.

Prator, C. & Robinett, B. (1985). *Manual of American English Pronunciation*. San Francisco: Holt, Reinhart and Winston.

Quirk, R., Greenbaum S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Rundell, M. (2002). Macmillan English Dictionary for Advanced Learners of American English. MacMillan.

Sampson, G. (1996). *Sistemas de Escrita:* Tipologia, História e Psicologia. São Paulo: Ática.

_____. (2001). *Empirical Linguistics.* New York/Londres: Continuum.

Sanchez, A. & Cantos, P. (1996). Cumbre - Curso de Español. Madri: SGEL.

Santaella, L. (1983). *O que é Semiótica*. São Paulo: Brasiliense.

Saussure, F. (2001). *Curso de Lingüística Geral*. 23a. edição. São Paulo: editora Cultrix.

Schirmer, C. R., Fontoura, D. R. & Nunes, M. L. (2004). Distúrbios da aquisição da linguagem e da aprendizagem. *Jornal de Pediatria,* 80 (2), 95-103. Rio de Janeiro.

Schmitz, J. R. (2003). *Pronunciation Teaching and Learning*: Standard Varieties and International Varieties. 10th Braz Tesol Pronunciation Conference, 11 de outubro. São Paulo.

_____. (2004). Taking linguistics seriously: on the varied dimensions of applied linguistics. *Lingua*, 114 (2), 95-100.

Schoolcraft, H. R. (1851). *Historical and Statistical Information*: Respecting the History, Condition, and Prospects of the Indian Tribes of the United States. Part 1, Philadelphie.

Schütz, R. (2005). História da Língua Inglesa. Disponivel na Internet no endereço: http://www.sk.com.br/sk-enhis.html. Acessado em 30 de março de 2005.

Scliar-Cabral, L. (2003). *Princípios do Sistema Alfabético do Português do Brasil*. São Paulo: Editora Contexto.

Shepherd, D. (1987). Portuguese speakers. In Swan, M. & Smith, B. (orgs.), *Learner English:* A Teacher's Guide to Interference and Other Problems. Cambridge: Cambridge University Press.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1995). From theory to practice. In Leech, G., Myers, G. & Thomas, J., *Spoken English on Computer*: Transcription, Mark-Up and Application. London: Longman.

_____. (1996). The dictionary of the future. In Foley, J. (org.), *J. M. Sinclair on Lexis and Lexicography*. Singapore: UniPress.

Smith, L. E. & Nelson, C. L. (1985). International intelligibility of English: directions and resources. *World Englishes*, 4, 333-342.

Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmidt & M. McCarthy (orgs.), *Vocabulary Description, Acquisition and Pedagogy*, pp. 237-257. Cambridge: Cambridge University Press.

Steinberg, M. (1985). *Pronúncia do Inglês Norte-Americano*. São Paulo: Ática.

Stubbs, M. (1993). British traditions in text analysis: From Firth to Sinclair. In M. Baker, G. Francis & E. Tognini-Bonelli (orgs.), *Text And Technology:* In Honour of John Sinclair. Amsterdam: John Benjamins.

Stubbs, M. (1995). Collocations and semantic profiles: on the cause of the trouble with quantitative studies. *Functions of Language*, 2 (1), 23-55.

Succi, O. (2003). A utilização da Lingüística de Corpus e da Gramática de padrões na análise de alguns adjetivos presentes em um livro didático de inglês para negócios. Dissertação de mestrado. São Paulo: Programa de Lingüística Aplicada e Estudos da Linguagem, PUC-SP.

Tagnin, S. E. O. (2001). Corpus Técnico da FFLCH-USP. Organizado pelos alunos do Curso de Especialização em Tradução da USP. Citrat: Centro Interdepartamental de Tradução e Terminologia. Disponivel em CD.

_____. (2005). *O Jeito que A Gente Diz*. São Paulo: Disal.

Treiman, R., Kessler, B. & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language*, 47, 448-468.

Treiman, R., Mullennix, J., Bijeljac-Babic, R. & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology*: *General*, 124, 107-136.

Venezky, R. L. (1970). *The Structure of English Orthography*. The Hague: Mouton.

Vygotsky, L. S. (2000). *A Formação Social da Mente:* O Desenvolvimento dos Processos Psicológicos Superiores. São Paulo: Martins Fontes.

Wanke, E. T. (1987). *A Ortografia que nos Atormenta:* Reflexões e Dados sobre o Problema Ortográfico e Sugestões para a Desburocratização da Escrita. Rio de Janeiro: Codpoe.

Welna, J. (1978). *A Diachronic Grammar of English.* Part 1: Phonology. Warszawa: Pa'nstwowe Wydawnictwo Naukowe.

Wijk, A. (1966). *Rules of Pronunciation of the English Language*: An Account of the Relationship between English Spelling and Pronunciation. London: Oxford University Press.

Wikipédia, *A Enciclopédia Livre*. Disponível na Internet no endereço http://www.wikipedia.org/. Acessado em 18 de agosto de 2005.

Wimmer, H. & Goswami, U. (1994). The influence of orthographic consistency on reading development: word recognition in English and German children. *Cognition*, 51, 91-103.

Woolard, G. (2005). Noticing and learning collocation. *English Teaching Professional*, 40, 46-48.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, 32 (4), 213-231.