PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO (PUC-SP)

Solange Contrera

AUTENTICIDADE EM LIVROS DIDÁTICOS PARA O ENSINO DE INGLÊS COMO LÍNGUA ESTRANGEIRA: UM ESTUDO DIACRÔNICO SOB A PERSPECTIVA DA LINGÜÍSTICA DE *CORPUS*

MESTRADO EM LINGÜÍSTICA APLICADA E ESTUDOS DA LINGUAGEM

SÃO PAULO

2010

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO (PUC-SP)

Solange Contrera

AUTENTICIDADE EM LIVROS DIDÁTICOS PARA O ENSINO DE INGLÊS COMO LÍNGUA ESTRANGEIRA: UM ESTUDO DIACRÔNICO SOB A PERSPECTIVA DA LINGÜÍSTICA DE *CORPUS*

Dissertação de mestrado produzida sob a orientação do Professor Doutor Antônio Paulo Berber Sardinha e apresentada à banca examinadora do departamento de Lingüística Aplicada e Estudos da Linguagem (LAEL), da Pontifícia Universidade Católica de São Paulo, como exigência parcial para a obtenção do título de mestre em Lingüística Aplicada e Estudos da Linguagem.

MESTRADO EM LINGÜÍSTICA APLICADA E ESTUDOS DA LINGUAGEM

SÃO PAULO

2010

AUTENTICIDADE EM LIVROS DIDÁTICOS PARA O ENSINO DE INGLÊS COMO LÍNGUA ESTRANGEIRA: UM ESTUDO DIACRÔNICO SOB A PERSPECTIVA DA LINGÜÍSTICA DE *CORPUS*

DISSERTAÇÃO DE MESTRADO

Solange Contrera

Banca examinadora:

_____

_____

_____

São Paulo, ... de ..... de 2010.

Pontifícia Universidade de São Paulo (PUC-SP)

São Paulo — 2010

Dedico este trabalho ao Wal, meu marido, meu amor, meu companheiro, por ter-me incentivado a fazer este mestrado, pela compreensão na minha ausência, e, acima de tudo, por estar sempre torcendo pelo meu sucesso.

Agradecimentos

Agradeço ao professor Doutor Antônio Paulo Berber Sardinha, meu orientador, que soube ensinar, exigir, bronquear, sorrir e rir quando foi preciso; por ter-me escolhido e me acolhido como orientanda; por ter acreditado e confiado no meu trabalho; pela paciência; pela seriedade; pelas revisões; pela descontração e, acima de tudo, pela amizade.

A todos os queridos professores do LAEL pelo exemplo e pela dedicação.

Às professoras doutoras Leila Darin e Solange Maria Sanches Gervai pelas valiosas contribuições durante o exame de qualificação.

A todos os amigos que estudaram comigo na PUC-SP; em especial, a todos os "tonetes" —Ana Júlia, Ciça, Cris, Eduardo, Elaine, Eliane, Evandro, Flávia, José, Lílian, Márcia Veirano, Marcinha, Paty (minha conterrânea), Renata, Rosana, Tadeu e Telma— pelos momentos de descontração e tensão, pelas risadas compartilhadas, pela ajuda de todas as horas e pelos seminários de pesquisa e orientação. Em especial, à Marcinha e à Telma pelo carinho, pelo apoio e pelo conhecimento compartilhado.

Aos amigos e alunos do CCAA, de Birigui, e da Hit Centre, de São Paulo, pelo incentivo, carinho e por ajudarem-me a atingir os meus objetivos acadêmicos sem deixar de lado a minha atuação profissional.

Aos meus pais, Pedro e Leonice, por serem meus eternos, grandes amigos e companheiros, a quem eu devo tudo o que tenho nesta vida.

Aos meus irmãos, Haroldo e Sandra, em especial a minha irmã, por sempre me dar o suporte necessário nas horas difíceis, por sempre acreditar no meu potencial e por ser a minha eterna melhor amiga.

Ao meu marido, Wal, pelo amor e pelo apoio de sempre. (Dê só uma olhadinha na dedicatória acima.)

Ao Eduardo Cassimiro pela atenção e revisão precisa.

Aos meus alunos, que são a minha paixão e a minha inspiração para buscar novas formas de ensinar.

Ao CNPq pela bolsa.

A todos as pessoas que colaboraram, direta ou indiretamente, na realização deste trabalho.

Resumo

A pesquisadora teve como objetivo principal verificar se os autores de livros didáticos atuais e voltados para o ensino de língua inglesa, de fato, empregam lexicogramática autêntica e se os autores de livros didáticos de décadas passadas, numa perspectiva diacrônica, também o fizeram. Para tanto, encontrou suporte teórico na Lingüística de *Corpus*, uma área cujos pesquisadores se preocupam com a coleta e com exploração criteriosas de *corpora* —em formatos legíveis por computador— com o objetivo de pesquisar uma língua. A pesquisadora fundamentou-se também nos pressupostos teóricos das abordagens de ensino de língua estrangeira conhecidas como (1) audiolingualismo, (2) comunicativismo e (3) informada por *corpus* para investigar o uso de língua autêntica no desenvolvimento dos livros didáticos. As questões de pesquisa norteadoras deste estudo concernem (1) à análise de pacotes lexicais convergentes e divergentes no *corpus* de estudo, (2) à descoberta de qual é a freqüência desses pacotes lexicais, (3) à análise dos pacotes convergentes e dos divergentes em cada texto dos livros didáticos e, por fim, (4) à constatação de quais são os livros compostos por um grau de autenticidade lingüística superior em relação aos demais investigados. Os *corpora* selecionados para a pesquisa são estes: (a) o *corpus* de estudo, formado por cinco livros didáticos para o ensino de inglês como língua estrangeira, para o nível intermediário, compostos por 25.485 palavras, e (b) os *corpora* de referência: o *British National Corpus* (*BNC*), composto por 100 milhões de palavras, e o Google *Corpus*, composto por 1 trilhão de palavras. A pesquisadora almejou, ao analisar o grau de autenticidade da língua constante desses livros didáticos —uns que foram, e outros que ainda são utilizados no ensino de inglês— trazer uma contribuição para a área da Lingüística de *Corpus*.

Palavras-chave: Lingüística de *Corpus*, Pacotes Lexicais, Livros didáticos. Abordagens. Língua inglesa, Autenticidade.

Abstract

The main aim of this researcher was to verify whether, in modern English teaching books, authentic bundles were used or not, and whether, in books of the last decades, in a diachronic perspective, if it was done or not. The main theoretical underpinning for the research is provided by Corpus Linguistic, which is the area that is concerned with the collection and analysis of criteriously selected corpora, which could be read by computers with the aim of linguistic research. The researcher was also based on the theoretical underpinning of the Teaching Approaches known as: (1) audiolingual; (2) communicative; and (3) corpus Informed to investigate the use of authentic language in the development of the English books. The research questions investigated in this study were: (1) to the analysis of convergent and divergent lexical bundles in this study corpus, (2) to discover which is the frequency of those lexical bundles; (3) to the analysis the convergent and divergent bundles in each text of the English books; and (4) and to verify which are the books composed by a degree of superior linguistic authenticity in relation to the others investigated. The corpora selected for the research are as follows: (a) the study of corpus, formed by five English books designed for a foreign language, for the intermediate level, composed of 25.485 words; (b) the reference corpora: British National Corpus (BNC), composed of 100 million words, and the Google Corpus, composed of 1 trillion words. The researcher analyzed, the degree of authenticity in the constancy of the language used in the English books analyzed -some that were, and others that are still in use in English teaching – bringing a contribution to the Corpus Linguistics field.

Keywords: Corpus Linguistic, Lexical Bundles, English Books, English language, Approaches, Authenticity.

Sumário

Introdução 122
Objetivos e questões de pesquisa 166
Organização da Dissertação 17
Capítulo 1 18
Fundamentação teórica 18
1.1. Lingüística de Corpus 18
1.1.1. Corpus 19
1.2 Definição 22
1.2.1 Padronização 23
1.2.2 Colocação 25
1.2.3 Coligação 26
1.2.4 Prosódia Semântica 28
1.2.5 Pacotes Lexicais 28
1.3 A Lingüística de *Corpus* no Ensino de Idiomas 31
1.3.1 Currículo lexical *(lexical syllabus)* 32
1.3.2 Abordagem lexical *(lexical approach*) 33
1.3.3 Aprendizado movido por dados *(data driving learning)* 34
1.4 Autenticidade 35
1.4.1 Definição de material autêntico 37
1.4.2 Uso de material autêntico no ensino de língua estrangeira 39
1.5 Métodos e abordagens relevantes a esta pesquisa 42
1.5.1 Abordagem 42
1.5.2 Método 42
1.6 Audiolingualismo 43
1.6.1 Definição 43
1.6.2 Breve Histórico 45
1.6.3 Material de ensino no audiolingualismo 47
1.7 Comunicativismo 51
1.7.1 Definição 51

1.7.2 Breve Histórico 53

1.7.3 Material de ensino no comunicativismo 56

1.8. Abordagem informada por *corpus* 60

1.8.1 Definição 60

1.8.2 Breve Histórico 61

1.8.3 Material de ensino na abordagem informada por *corpus* 62

Capítulo 2 66

Metodologia 66

2.1 Objetivos e questões de pesquisa 66

2.2. *Corpora* e critérios de elaboração e coleta 67

2.2.1 *Corpora* de estudo 67

2.2.2 *Corpora* de referência 70

2.2.2.1 Informações sobre o tamanho do Google Corpus 71

2.2.3 Critérios de elaboração e coleta 73

2.3 Procedimentos de Análise dos Dados 74

2. 3.1 Ferramentas de coleta 75

2.3.2 Análise dos dados 78

Capítulo 3 86

Análise dos Dados 86

3.1 Análise quantitativa dos *corpora* 86

3.1.1 Análise quantitativa inicial 87

3.1.2 Análise qualitativa recontagem 92

3.1.3 Análise qualitativa de variação de convergência texto por texto 103

3.2 Análise qualitativa dos corpora 107

3.2.1 Análise qualitativa dos pacotes lexicais divergentes 108

3.2.2 Análise dos pacotes lexicais convergentes 110

4 Considerações Finais 120

Referências Bibliográficas 124

LISTA DE TABELAS

Tabela 1: Alguns *corpora* de língua inglesa, a sua extensão, a sua composição e a sua disponibilidade **21**

Tabela 2: Porcentagem de convergência e divergência entre os pacotes lexicais dos livros didáticos e os do Google *Corpus* **87**

Tabela 3: Porcentagem dos pacotes lexicais constantes dos livros didáticos em relação aos do *British National Corpus* **90**
Tabela 4: Recontagem dos pacotes lexicais convergentes e divergentes do livro *English 900* comparado ao *BNC* **93**
Tabela 5: Recontagem dos pacotes lexicais convergentes e divergentes do livro *English 900* comparado ao Google *Corpus* **93**
Tabela 6: Recontagem dos pacotes lexicais convergentes e divergentes do livro *To the Top* comparado ao *BNC* **94**
Tabela 7: Recontagem dos pacotes lexicais convergentes e divergentes do livro *To the Top* comparado ao Google **95**
Tabela 8: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Inside Out* comparado ao *BNC* **95**
Tabela 9: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Inside Out* comparado ao Google *Corpus* **96**
Tabela 10: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Interchange* comparado ao *BNC* **97**
Tabela 11: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Interchange* comparado ao Google *Corpus* **98**
Tabela 12: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Touchstone* comparado ao *BNC* **99**
Tabela 13: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Touchstone* comparado ao Google *Corpus* **100**
T**abela 14:** Quantidade absoluta de pacotes convergentes em cada livro, por faixa de convergência **104**
Tabela 15: Quantidade relativa, em porcentagem, de pacotes convergentes em cada livro, por faixa de convergência **104**
Tabela 16: Quantidade relativa, em porcentagem, de pacotes convergentes em cada livro, em duas faixas principais **104**
Tabela 17: *Ranking* de convergência de acordo com a tabela **105**
Tabela 18: Os 20 primeiros *pacotes lexicais matches* do livro *English 900* Comparados ao Google *Corpus* **113**
Tabela 19: Os 20 primeiros pacotes lexicais *matches* do livro *To the Top: Way Ahead* comparados ao Google *Corpus* **113**
Tabela 20: Os 20 primeiros pacotes lexicais *matches* do livro *Interchange* comparados ao Google *Corpus* **115**
Tabela 21: Os 20 primeiros pacotes lexicais *matches* do livro *Inside Out* comparados ao Google *Corpus* **116**
Tabela 22: Os 20 primeiros pacotes lexicais *matches* do livro *Touchstone* comparados ao Google *Corpus* **118**
Tabela 23: Quadro do *ranking* dos livros antes da pesquisa **120**
Tabela 24: Quadro do *ranking* dos livros após as análises à luz do *BNC* **120**
Tabela 25: Quadro do *ranking* após as análises à luz do Google *Corpus* **121**
Tabela 26: Quadro do *ranking* após a análise quantitativa de variação de convergência texto por texto **121**

LISTA DE FIGURAS

Figura 1: Marcas de entonação (English 900, book two, p.14) **50**
Figura 2 : Exercícios estruturais (English 900, book two,p.14) **50**
Figura 3: Word sort activity – *Touchstone,* Level 1, Unit 8B **65**
Figura 4: Amostra do programa *WordSmith Tools* **79**

Figura 5: Pacotes lexicais do livro *English 900*, de 1961 **79**
Figura 6: Lista dos pacotes lexicais convergentes e divergentes **80**
Figura 7: Uma amostra do teste qui-quadrado **82**
Figura 8: Gráfico representativo dos pacotes lexicais dos livros didáticos comparados aos do Google *Corpus* **88**
Figura 9: Gráfico representativo dos pacotes lexicais constantes dos livros didáticos comparados aos do *BNC Corpus* **90**
Figura10: Grau de convergência entre os *corpora* **100**
Figura 11: Gráfico após a recontagem dos pacotes convergentes comparados ao *BNC* **101**
Figura 12: Gráfico após a recontagem dos pacotes lexicais convergentes comparados ao Google *Corpus* **102**
Figura 13: Gráfico de todos os *rankings* obtidos por meio da pesquisa **122**

LISTA DE QUADROS

Quadro 1: *BNC*: Descritivo do componente escrito **71**
Quadro 2: Número de pacotes lexicais extraído do Google *Corpus* **72**
Quadro 3: Número total de *tokens* do *corpus* de estudo **74**
Quadro 4: Classificação do *corpus* de acordo com o seu "tamanho" **74**
Quadro 5: Ferramentas, utilitários, instrumentos e funções disponíveis no W*ordSmith Tools* **77**

LISTA DE ANEXOS

Anexo 1: Script criado pelo orientador **130**
Anexo 2: Faixas com % de convergência **132**
Anexo 3: Texto número 18, do livro *Inside Out* **135**
Anexo 4: Texto número 13, do livro Inside Out **138**

Introdução

Os autores de muitos livros didáticos para o ensino de inglês como língua estrangeira, hoje em dia, enfatizam o "fato" de apresentarem ao aluno a língua "real", isto é, tal qual ela é usada, fora da sala de aula, pelos usuários dela nativos. "*Real English*" e outras expressões desse tipo aparecem nas contracapas e no material promocional de muitos livros, além de dicionários e gramáticas voltados ao público estudantil. Porém, nem sempre foi assim. Nos anos 1960s e 1970s, por exemplo, por causa do audiolingualismo, seria muito estranho (senão contraditório) se o autor do material didático se gabasse por oferecer ao consumidor daquele material didático a "língua real". Livros dessa linha, na verdade, punham em evidência os *drills*, que eram incluídos no material porque se visava à automatização de estruturas gramaticais. Ao longo do tempo, foi tomando corpo a idéia de que o livro didático para o ensino de inglês devesse incorporar as modalidades falada e escrita da língua —com as quais os estudantes haveriam de se deparar fora da sala de aula— a fim de prepará-los para o engajamento com falantes "verdadeiros" (nativos ou não) e com textos "autênticos" em situações reais de uso da língua inglesa. Contudo, essa noção de que o texto autêntico é melhor para o aluno, muitas vezes, acaba sendo colocada em xeque entre professores de inglês, que dizem ter aprendido inglês por meio de métodos "antigos", de *drills*, de repetições, de textos inventados, de frases soltas e descontextualizadas e de situações artificiais. Muitos até confessam que gostam desse tipo de material, mas não podem usá-lo porque "saiu de moda", e garantem que 'funciona'.

Em meio a essa problemática que é o processo de ensino e aprendizagem de língua estrangeira e às muitas questões que a envolvem, uma inquietação nos moveu desde o início do mestrado: saber se os materiais didáticos antigos e os atuais eram compostos, mesmo sem que os seus autores admitissem, de língua autêntica. A nossa hipótese —por assim dizer, já que não se trata de uma pesquisa quantitativa por meio da qual se possa confirmar ou rejeitar hipóteses— era a de que mesmo os livros "audiolinguais", cujos autores rejeitavam a autenticidade lingüística, na verdade, acabavam expondo os seus usuários à lexicogramática autêntica, nos seus textos inventados e, portanto, artificiais. Por lexicogramática, entendemos, à luz do arcabouço teórico da Lingüística de *Corpus*, padrões de uso recorrente da língua formados por itens lexicais próximos uns dos outros; tais estruturas são conhecidas por diversos nomes como *bundles*, *chunks*, colocações e pacotes lexicais (o termo

que usamos na pesquisa).

A partir desse nosso ímpeto inicial, passamos a nos preocupar com a descoberta de qual seria a quantidade de lexicogramática autêntica constante, dos textos e dos diálogos, de livros didáticos voltados para o ensino de inglês como língua estrangeira de diversas épocas e orientações pedagógicas.

Um texto autêntico pode ser entendido como "aquele que não foi criado com a finalidade de ensinar língua" (Berber Sardinha, 2007:275). Segundo o autor, "as seqüências de palavras de um texto autêntico ecoam nas vozes dos milhões de falantes de inglês do mundo"; já o texto artificial, por sua vez, "ecoa apenas parcialmente e não exibe a riqueza, o encanto e o desafio do texto autêntico".

Segundo Mishan [(2005) *apud* Lee (1995:324)]:

> Um texto, normalmente, é considerado autêntico se não tiver sido escrito para propósitos pedagógicos, mas para um propósito comunicativo da vida real onde [*sic*] o escritor tem uma mensagem para passar para o leitor. Como tal, um texto autêntico é aquele que possui [*sic*] uma qualidade comunicativa de forma intrínseca.[1]

Para viabilizar a pesquisa, fizemos um recorte referente aos livros didáticos a serem investigados. Selecionamos livros didáticos pertencentes a três abordagens marcantes na história do ensino de inglês como língua estrangeira, todos voltados para aprendizes de nível intermediário:

- ***Dois da abordagem audiolingual:***
  - ***English 900****,* publicado em 1961, pela editora Collier Publishing, o mais antigo dessa abordagem;
  - ***To the Top: Way Ahead***, publicado em 2006, pela Waldyr Lima Editora, o mais atual dessa abordagem e utilizado por uma escola de idiomas de grande porte, que está no mercado há cerca de 50 anos e tem, aproximadamente, 250 mil alunos;
- ***Dois da abordagem comunicativa:***
  - ***Interchange: English for International Communication***, publicado em 1991, pela editora Cambridge University Press; o mais antigo dessa abordagem;

[1] A text is usually regarded as authentic if it is not written for teaching purposes but for a real-life communicative purpose, where the writer has a certain message to pass on to the reader. As such, an authentic text is one that possesses an intrinsically communicative quality.

- ***Inside Out****,* publicado em 2004, pela editora Macmillan; o mais atual dessa abordagem;

- ***Um da abordagem informada por corpus:***

- ***Touchstone***, publicado em 2006, pela editora Cambridge University Press; o mais recente de toda a coleção estudada na pesquisa.

A seguir, faremos uma breve apresentação dessas abordagens, uma vez que a exposição completa será exposta no capítulo de fundamentação teórica.

A primeira abordagem é o audiolingualismo, que dominou os programas acadêmicos nas décadas de 1950 e 1960, nos Estados Unidos. No Brasil, ele começou a "instalar-se" nas escolas de idiomas no final da década de 1960.

Uma vez que o audiolingualismo é uma abordagem oral de ensino de línguas, é natural que o seu processo de ensino e aprendizagem envolva intensa atividade oral. O foco é, portanto, na fala; a gramática é apenas uma coadjuvante que dele participa por meio de breves explicações ou referências. A língua-alvo é usada inclusive para darem-se as instruções, e a inclusão da língua materna é firmemente desencorajada. O número ideal de alunos por sala de aula é de, no máximo, dez.

Os livros didáticos dessa abordagem contêm muitos exemplos artificiais, e os professores, influenciados pelo behaviorismo, tentavam impedir as hesitações e as pausas, contrariando as características naturais das interações conversacionais (Paiva, 2005). Havia pouco estímulo às atividades mentais, pois se valorizava o automatismo, e pouca ênfase era dada à capacidade do ser humano de pensar e produzir sentido; ou seja, os materiais, aparentemente, ostentavam pouca autenticidade. Tal limitação dessa abordagem foi um dos construtos que mais influenciou o nascimento da segunda abordagem contemplada pela nossa pesquisa: o comunicativismo.

A abordagem comunicativa surgiu como uma forte reação contra o audiolingualismo (Schutz, 2006). Nela, a função se sobrepõe à forma e ao significado, e as situações são inspiradoras da planificação didática e da confecção de materiais. A competência comunicativa passa ser o objetivo principal do processo de ensino e aprendizagem de língua estrangeira, em vez do acúmulo de conhecimento gramatical ou da estocagem de formas memorizadas.

Segundo Paiva (2005), o livro didático, na abordagem comunicativa, deixa de ser uma bíblia a ser seguida cegamente e se torna um dos recursos para a

aprendizagem, pois os alunos devem ser também expostos a insumos lingüísticos autênticos —como livros, jornais, revistas, filmes, programas de TV, etc., além dos diversos recursos da *internet*.

A autenticidade começava a desempenhar papel central na elaboração de livros para o ensino de idiomas, e os textos autênticos começavam, desse modo, a enriquecer o material didático.

Quando tratamos de autenticidade em livros didáticos, não podemos deixar de citar a terceira abordagem da nossa pesquisa, que é a informada por *corpus*, pois é nela que a noção de linguagem autêntica fica mais evidente, em virtude da centralidade que ocupa na disciplina fundadora dessa abordagem: a Lingüística de *Corpus*.

A abordagem informada por *corpus* esforça-se para trazer aos estudantes e aos professores mais confiança quanto ao uso genuíno do idioma-alvo, o que é essencial para uma boa comunicação nele. Assim, esse idioma nos permite falar com mais precisão e sofisticação, levando-os em suas metas de uma forma mais rápida e eficaz (McCarten & Sandiford, 2006).

Enfatizamos, nesta pesquisa, que a abordagem informada por *corpus* não tem caráter de ensino (como as abordagens audiolingual e comunicativa); por enquanto, ela só é utilizada na elaboração e no desenvolvimento de livros didáticos.

A utilização da Lingüística de *Corpus* na análise da autenticidade desses livros didáticos é de suma importância para a apreciação da língua constante dos textos e dos diálogos formadores do *corpus* de estudo.

A Lingüística de *Corpus* “ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados lingüísticos textuais coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística” (Berber Sardinha, 2004:3).

Por meio da Lingüística de *Corpus*, poderemos verificar até que ponto os textos e os diálogos existentes nos livros contêm padrões léxico-gramaticais, ou seja, pacotes lexicais (*bundles*) típicos da língua autêntica que indicam traços de autenticidade. Assim, a nossa metodologia é de cunho probabilístico (Berber Sardinha, 2004) e não categórico, pois não classificaremos cada texto como inteiramente autêntico ou não-autêntico, mas identificaremos cada texto e cada livro didático (com a ajuda de ferramentas computacionais), como mais ou menos convergente. Para essa aferição, empregaremos dois *corpora* de referência do

inglês: o *British National Corpus* (BNC) e o Google *Corpus*, que, atualmente, são, no mundo, os maiores *corpora*, em inglês, disponíveis para uma pesquisa como a nossa.

Conforme dissemos, tentamos descobrir se livros didáticos atuais e de décadas passadas empregam lexicogramática autêntica. Por lexicogramática autêntica entendemos aquela que é encontrada em textos autênticos, escritos e falados, reunidos em gigantescos bancos de dados chamados de *corpora*. A nossa suspeita, advinda da prática no ensino de inglês no país há muitos anos, é de que:

(1) mesmo livros didáticos antigos, de orientação estrutural, inclusive os de base audiolingual, ou funcional, ou os comunicativos e os informados por *corpus* também apresentem algum tipo de língua autêntica nos seus textos e diálogos;

(2) livros contemporâneos, inclusive os de base estrutural ou funcional, contenham mais material lingüístico autêntico do que os seus pares mais antigos;

(3) e livros didáticos elaborados com base em *corpora*, ainda que apregoem serem totalmente autênticos, não o são, haja vista que, durante a produção do livro didático, é preciso tomar decisões que, muitas vezes, implicam a simplificação e/ou a alteração da língua constante dos seus textos e diálogos.

Embora haja uma série de publicações que analisam livros didáticos para o ensino de línguas estrangeiras, há poucas pesquisas sobre análise da autenticidade da lexicogramática dos livros didáticos, de várias épocas, baseada na Lingüística de *Corpus*, conforme propomos neste estudo. Dessa forma, buscamos, por meio da pesquisa, fazer uma contribuição à literatura mediante as análises dos referidos livros didáticos, que foram e ainda são utilizados no ensino de inglês como língua estrangeira.

Objetivo e questões de pesquisa

Conforme exposto acima, nesta pesquisa, temos como objetivo geral verificar (1) se os autores de livros didáticos atuais —para o ensino de língua inglesa— empregaram, de fato, textos autênticos e, (2) até que ponto, os autores de livros didáticos de décadas passadas, numa perspectiva diacrônica, também o fizeram.

Diante desses objetivos, as questões de pesquisa norteadoras deste estudo são estas:

1) Quais são os pacotes lexicais existentes no *corpus* de estudo?

2) Qual é o grau de convergência e/ou divergência entre os pacotes lexicais constantes dos *corpora* de estudo e de referência ?

3) Qual é o grau de convergência e/ou divergência entre os pacotes lexicais constantes dos *corpora* de estudo e de referência tendo em vista cada texto dos livros didáticos?

4) Quais livros didáticos apresentam maior e menor grau de autenticidade, tendo em vista os índices de convergência e/ou divergência?

Organização da dissertação

A dissertação está organizada da seguinte maneira: o primeiro capítulo foi dedicado à fundamentação teórica da pesquisa. No capítulo 2, apresentamos, em detalhes, a metodologia do estudo, incluindo a descrição dos *corpora* de estudo e de referência, bem como a especificação dos procedimentos de coleta e de análise dos dados. No capítulo 3, fazemos a discussão e a análise dos dados e procedemos à resposta a cada uma das perguntas de pesquisa. O volume é encerrado pelas nossas considerações finais, que são seguidas das referências e dos anexos.

.

Capítulo 1

Fundamentação teórica

Este capítulo constitui o arcabouço teórico da pesquisa e está dividido em três seções principais: na primeira, tratamos de concepção de linguagem e de Lingüística de *Corpus*; na segunda, tratamos da questão da autenticidade e, na terceira, traçamos um breve histórico das abordagens constantes dos livros didáticos selecionados para esta pesquisa.

Assim, a primeira seção está voltada para a visão de linguagem e, nela, serão apresentados os princípios teóricos da Lingüística de *Corpus*, bem como uma explanação sobre o conceito de corpora. Apresentaremos, ainda, um breve histórico da Lingüística de *Corpus*, além dos critérios para a construção da pesquisa nessa perspectiva e dos tipos de *corpora* existentes.

Na segunda seção, é abordada a questão da autenticidade, dos materiais autênticos e do seu uso no ensino de língua estrangeira (doravante, LE), com base nas teorias de Berber Sardinha (2000), Coracini (1999), Guariento e Morley (2001), Mishan (2004 e 2005) e outros.

Na terceira e última seção, apresentamos uma análise das abordagens utilizadas pelos autores dos livros didáticos de ensino de inglês como LE selecionados para esta pesquisa, que podemos encontrar nos trabalhos de Richards & Rodgers (1986), Almeida Filho (1993), Kumaravadivelu (1994), e nos de outros. Desse modo, em tal seção, serão abordadas as abordagens de ensino de LE relevantes a este estudo, bem como as suas definições, um breve histórico e alguns exemplos oriundos de materiais utilizados para o ensino de idiomas.

1.1. Lingüística de *Corpus*

Antes da definição de Lingüística de *Corpus* (doravante, LC) e dos seus fundamentos teóricos e metodológicos, cabe, aqui, esclarecer o significado de *corpus*, que é o objeto de estudo da LC.

1.1.1. *Corpus*

Os estudiosos da área definem o termo *corpus* de diversas maneiras. Conforme Biber (1999:12), *corpus* é "uma extensa coletânea de textos naturais". Segundo Sampson & McCarthy (2004:6),

> um *corpus* se refere a um exemplo, de "tamanho" razoável, da língua (real) em uso —em inglês ou em outro idioma estudado—, que será compilado e utilizado como fonte de evidência para hipóteses gerais sobre a natureza da língua[1]. (Tradução minha.)

Corpus pode ser explicado também como "uma coletânea de textos naturais (*naturally occurring*), escolhidos para caracterizar um estado ou variedade da linguagem (...); é um corpo de linguagem natural (autêntica) que pode ser usado como base para pesquisa lingüística". [Sinclair (1991:171) *apud* Berber Sardinha (2004:16-17)].

Todavia, a definição que nos parece mais completa —e incorpora as características principais— de um *corpus* é esta:

> 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

[1] A corpus refers to a sizeable sample of real-life usage in English or another language under study, compiled and used as a source of evidence for generating hypotheses about the nature of the language.

> 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.
>
> [Sanchez (1995:8-9) a*pud* Berber Sardinha (2004:18)]

Por meio desta citação, algumas características inerentes a um *corpus* podem ser elucidadas:

> · quanto à origem: os textos devem ser autênticos;
>
> · quanto ao propósito: o corpus deve ter a finalidade de ser um objeto de estudo lingüístico;
>
> · quanto à composição: o conteúdo do corpus deve ser criteriosamente escolhido;
>
> · quanto à formatação: os dados do corpus devem ser representativos de uma língua ou variedade;
>
> · quanto à representatividade: o *corpus* deve ser vasto para ser representativo".
>
> (Berber Sardinha, 2004:16)

Na tabela a seguir, são apresentados os nomes de alguns corpora da língua inglesa, que, no caso deste estudo, é a língua do *corpus* analisado. Podemos notar que o *corpus* de referência empregado nesta pesquisa, o BNC, é composto de inglês britânico escrito e falado.

Nome do Corpus	Palavras	Composição	Lançamento	Disponível
BNC	100 milhões	Inglês Britânico escrito e falado	1995	Sim, on-line em: http://corpus.byu.edu/bnc/
Brown Corpus (Brown University Standard Corpus of Present-day American English)	1 milhão	Inglês americano escrito	1964	Não
Bank of English	450 milhões	Inglês Britânico	1987	Não
The Corpus of Contemporary American English (COCA)	Mais de 360 milhões	Inglês americano		Sim, on-line em: www.americancorpus.org

Tabela 1 – Alguns *corpora* de língua inglesa, a sua extensão, a sua composição e a sua disponibilidade.

Como os termos *corpus* de referência e *corpus* de estudo serão muito utilizados nesta pesquisa, cabe-nos defini-los. A esse respeito, Sinclair (1991:17-18) esclarece que *corpus* de referência

> (...) é um conjunto de materiais largamente homogêneos, mas reunida a partir de uma variedade de fontes de maneira que a individualidade de uma fonte seja obscurecida, a menos que o pesquisador isole um texto em particular. A diversidade de fontes é uma proteção essencial (...) para permitir uma amostragem eficiente[2]. (Tradução minha.)

Berber Sardinha (2004:21) também define *corpus* de referência como o "usado para fins de comparação com o *corpus* de estudo".

[2] It is a collection of material which is broadly homogeneous, but which is gathered from a variety of sources so that the individuality of a source is obscured, unless the researcher isolates a particular text. The diversity of sources in an essential safeguard (…) to allow efficient sampling.

Tanto a coleta quanto a análise de *corpora* são essenciais para a descrição de uma língua, ou de um determinado tipo de linguagem em uma abordagem empírica. É preciso, porém, que haja uma ciência que efetivamente teorize e caracterize a análise da linguagem em uso; tal ciência é a LC.

1.2 Definição de Lingüística de *Corpus*

A "LC pode ser entendida como o estudo da linguagem baseado em exemplos da língua usada na 'vida real'" (McEnery & Wilson, 1996:1-2). São muitos os autores que definem, de maneiras parecidas, a LC. Alguns exemplos de importantes vozes são os de Sinclair (1991), Stubbs (1993), McEnery & Wilson (1996), Biber *et al* (1998), Kennedy (1998), Hunston (2002) e Berber Sardinha (2004).

Berber Sardinha (2004:3) conceitua de forma concisa o que é LC, ao afirmar que

> (...) a Lingüística de *Corpus* ocupa-se da coleta e da exploração de Corpus, ou conjuntos de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou de uma variedade lingüística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador.

Podemos destacar quatro características típicas de uma análise baseada em *corpus*, segundo nos sugerem Biber *et al* (1998:5):

> i. É empírica, analisando os padrões atuais de uso em textos naturais.
>
> ii. Utiliza uma coleta de textos naturais, por meio de *softwares* específicos, e é realizada de maneira objetiva e ampla, conhecida por *corpus*, como a base para a análise.

iii. Faz uso extensivo do computador para a análise, utilizando-se das técnicas interativas e automáticas.

iv. Depende de ambas as técnicas analíticas: a quantitativa e a qualitativa.

Apesar de ter-se tornado mais conhecida no início dos anos 1990s, a LC vinha desenvolvendo-se fazia vários anos. Todavia, a influência da corrente gerativista e de algumas críticas em relação à coleta com *corpora* manuais fizeram com que a LC não fosse conhecida por muitos anos.

LC, no entanto, não é apenas uma metodologia nova —emergente— para o estudo da linguagem (como as correntes gerativistas afirmavam), mas tem uma nova visão da linguagem. Podemos resumir a posição teórico-metodológica da LC, segundo Leech (1992:106-107)[3], como aquela que tem foco:

— no desempenho lingüístico;

— na descrição lingüística em vez nos universais lingüísticos;

— no quantitativo, bem como nos modelos qualitativos da linguagem;

— empírico na pesquisa lingüística.

Este estudo, como já vimos, tem por foco a análise de pacotes léxico--gramaticais, que é uma das propostas pela LC. Assim sendo, nas subseções subseqüentes, daremos mais informações sobre esse assunto.

1.2.1 Padronização

No que tange ao conceito de padronização, Huston & Francis (1999:03) afirmam que “um padrão é uma fraseologia freqüentemente associada a (o

[3] Focus on linguistic performance, rather than competence;

Focus on linguistic description rather than linguistic universals ;

Focus on quantitative, as well as qualitative models of language ;

Focus on a more empiricist, rather than a rationalist view of scientific inquiry.

sentido de) uma palavra, particularmente em relação às preposições, grupos e orações que seguem aquela palavra" [4]. Especificamente, elas (1999:37) definem padrões como:

> (...) todas as palavras e estruturas que são regularmente associadas com a palavra e contribuem para seu sentido. Um padrão pode ser identificado se uma combinação de palavras ocorrer com relativa freqüência, se ela depender de uma escolha lexical específica, e se houver um sentido claro associado a ela[5].

Padronização é definida ainda como "regularidade expressa na recorrência sistemática de unidades co-ocorrentes de várias ordens (lexical, gramatical, sintática, etc.)" em Berber Sardinha (2004:31). Outros teóricos, como Hunston (2000), definem padronização como

> todas as palavras e estruturas, com as quais são regularmente associados, que contribuam para o seu significado. Um padrão pode ser identificado se uma combinação de palavras ocorre com relativa freqüência, se é dependente de uma palavra específica, e se há um significado claro associado.
>
> (Hunston, 2000, *apud* Berber Sardinha, 2004:39).

Como podemos observar nas definições supracitadas, sob a perspectiva da LC, a língua "é formada por porções lexicais (*chunks*) ou idiomas" (Berber Sardinha, 2004:33). Segundo essa visão, a descrição da linguagem é feita "pelo ponto de vista lexical, cuja perspectiva é a de descrição de quais agrupamentos lexicais são realmente empregados pelos falantes, ou seja, atestados pelo uso" (Berber Sardinha, 2004:33).

Desse modo, o conceito da padronização é um dos mais importantes da LC.

---

[4] (...) a pattern is a phraseology frequently associated with (a sense of) a word, particularly in terms of preposition, groups and clauses that follow that word.

[5] (…) all the words and structures which are regularly associated with the word and which contribute to its meaning. A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it.

A freqüência de co-ocorrência entre itens lexicais tem permitido aos lingüistas de *corpus* analisar os itens lexicais de acordo com fenômenos de padrões de associações conhecidos por colocações, coligações e prosódias semânticas. Faremos um breve esclarecimento de cada um a seguir.

1.2.2 Colocação

Firth foi o primeiro lingüista a tratar de colocações. Em 1957, ele já dizia que devemos "julgar uma palavra pela sua companhia" [6]**.** Esse princípio garante que as "palavras não ocorram ao acaso, em um texto"[7] (Sinclair, 1991:110). Assim, colocação é definida como a "ocorrência de duas ou mais palavras em uma curta 'distância' uma da outra em um texto" [8], de acordo com Sinclair (1991:170).

Colocação é definida a partir do conceito de sentido colocacional, por Leech (1974:20). Para ele, o "sentido colocacional é o que se constitui a partir das associações de uma palavra que tendem a ocorrer no 'ambiente natural' dela"[9].

Além dessas definições, há a de Hoey (1991:6-7): "colocação, há muito tempo, tem sido o nome dado ao relacionamento que um item lexical tem com itens que aparecem com muita probabilidade em seu contexto (textual)"[10].

Existem três definições principais de colocação na literatura:

> · Textual: colocação é a co-ocorrência de duas ou mais palavras distantes um "pequeno espaço" de texto uma das outras (219:170).
>
> · Psicológica: o sentido colocacional consiste nas associações que uma palavra faz por conta dos sentidos das outras palavras que tendem a ocorrer no ambiente (143:20).

[6] …you shall judge a word by the company it keeps.
[7] …words do not occur at random in a text.
[8] …an item collocates with another if it appears somewhere near it in a given text.
[9] …collocative meaning consists of the associations of a word which tend to occur in its environment.
[10] collocation has long been the name given to the relationship a lexical item has with items that appear with greater then random probability in its (textual) context.

> · Estatística: colocação tem sido o nome dado à relação que um item lexical tem com itens que aparecem com probabilidade significativa no seu contexto (textual).
>
> (Berber Sardinha, 2004:41)

É importante ressaltar que, para analisar uma colocação, é preciso diferenciar os conceitos de nódulo e colocado. Segundo Stubbs (2001:29), "nódulo (*node*) é a palavra ou o lema (forma canônica da palavra) que está sob investigação. O colocado (*collocate*) é uma palavra ou lema que, em um *corpus*, ocorre com o nódulo".

É possível dizer que todas essas definições têm em comum o fato de tratarem da colocação como uma associação entre itens lexicais que adquirem sentido a partir da sua junção. Um exemplo disso é o que Sinclair (1991:110) destaca. Consideremos a locução prepositiva inglesa *of course*: se analisarmos, separadamente, a preposição e o substantivo, não chegaremos a nenhuma conclusão a respeito do sentido de tal expressão. Contudo, se considerarmos o fato de essas duas palavras estarem colocadas, elas adquirem um sentido idiomático, que não poderia ser abstraído se as olhássemos como duas palavras separadas.

Esclarecido o conceito de colocação, cabe-nos, agora, discorrer sobre o que é coligação.

1.2.3 Coligação

Coligação pode ser definida como a companhia gramatical que a palavra mantém e as posições que ela prefere (Hunston, 2000). Coligação também é explicada como a "associação entre itens lexicais e gramaticais" por Berber Sardinha (2004: 40).

Hunston (2000:15) diz que utilizar o termo coligação é importante, pois ele chama a atenção para o fato de que "a evidência de muitas situações de linguagem autêntica pode ser utilizada para explicar o comportamento que é

tradicionalmente associado à gramática".[11] Hunston (2000:15) menciona também a maneira pela qual Hoey definiu colocação, em uma palestra, em 1998, como:

> a) a companhia gramatical que uma palavra "mantém" (ou "evita manter") ou no seu grupo, ou numa posição superior;
>
> b) as funções gramaticais que o grupo da palavra "prefere" (ou "evita");
>
> c) o lugar, em uma seqüência, que uma palavra "prefere (ou "evita")[12].
>
> [Hoey (1998) *apud* Huston (2000:15)] (Tradução minha.)

Coligação é a "combinação consagrada de elementos lingüísticos em que o colocado é uma palavra gramatical", para Hoey (1998:15). Existem vários tipos de coligações, como as classificadas por regência: verbos, substantivos, adjetivos e advérbios.

Uma das principais vantagens de estudar-se colocação e coligação consite no fato de esse estudo permitir que tenhamos expectativas acerca do que vai ser dito, ou melhor, de como vai ser dito. Ambas permitem que tenhamos à nossa disposição as estruturas semiconstruídas, das quais Sinclair fala. Temos, portanto, mais economia de tempo e mais facilidade de comunicação tanto na produção quanto na compreensão de um texto ou fala.

A seguir, explicaremos o que se entende por prosódia semântica.

---

[11] (…) the evidence of many instances of naturally occurring language can be used to explain behavior that is traditionally associated with grammar.

[12] (a) The grammatical company a word keeps (or avoids keeping) either within its own group or at a higher rank; (b) The grammatical functions that the word's group prefers (or avoids);

(c) The place in a sequence that a wor sd prefers (or avoids).

1.2.4 Prosódia semântica

Entende-se por prosódia semântica a associação entre itens lexicais e conotação (negativa, positiva ou neutra) ou instância avaliativa (Berber Sardinha, 2004:40). Segundo o autor, o nome “prosódia semântica” pode ser atribuído ao fato de certas palavras “prepararem o ouvinte, ou o leitor, para o conteúdo semântico que está por vir, da mesma maneira pela qual a prosódia, na fala, indica para o interlocutor que tipos de sons estão por vir a seguir”.

A prosódia semântica envolve a conotação das palavras, ou seja, é a carga semântica que cada indivíduo tem —em relação a uma palavra— de acordo com a experiência pessoal e cultural com a língua. Essa carga semântica pode influenciar positiva ou negativamente as palavras que acompanharão determinada palavra, ou o sentido do texto, ou a intenção do seu autor ao atribuir determinados sentidos às palavras ou expressões.

Neste nosso estudo, todavia, não serão destacadas questões de prosódia semântica.

1.2.5 Pacotes lexicais

A definição de pacotes lexicais *(*ou *lexical bundles)* como seqüências de palavras que ocorrem naturalmente no discurso é atribuída a Biber *et al* (1999). Ainda de acordo com Biber *et al* (1999:990-991), por um lado, as combinações de palavras em seqüências ininterruptas não são unidades estruturais completas ou bem-formadas, do mesmo modo que não são expressões lexicais fixas ou idiomáticas reconhecidas facilmente pelos falantes. Por outro lado, para que tais combinações sejam consideradas pacotes lexicais, devem recorrer, freqüentemente, numa vasta coleção de textos (em cinco ou mais textos distribuídos entre registros variados), e esse rigor na exigência de tão alta freqüência para tal classificação tem por finalidade evitar tendências idiossincráticas por parte de certo usuário da língua sejam consideradas pacotes

lexicais. Em inglês, o autor considera expressões como *the end of the*, *in addition to the*, *the point of view of*, e tantas outras, pacotes lexicais.

Essa afirmação é complementada por Stubbs (2007) ao apoiar-se no conceito de seqüência múltipla de palavras (ou *multi-word sequence*) para referir-se aos estudos baseados na extração do conjunto de palavras ininterruptas, recorrentes no texto, por meio de programas computacionais (*softwares*). O autor pressupõe que uma das definições possíveis nessa seqüência ininterrupta de palavras pode ser atribuída aos *n-grams*, os quais fornecem, com o auxílio de instrumentos computacionais, um determinado número de palavras, em conjunto, ao mesmo tempo, ordenando-as alfabeticamente ou por freqüência. Para o autor, não existe, em português, um termo padrão para *n-grams*, embora afirme que também que, em inglês, podem ser denominados *statistical phrases, recurrent word-combinations, lexical bundles, cluster, chains ou multi-word sequences*.

Por conseguinte*,* são, para Mazza (2009), pacotes lexicais caracterizados por combinações de três ou mais palavras identificadas, empiricamente, em um *corpus* de língua natural. Do ponto de vista da autora, a aquisição e o uso apropriado dessas combinações podem não ser um processo tão natural, em razão da importância de considerar-se o significado que essas expressões adquirem em determinados contextos. Ademais, as unidades multipalavras revelam padrões da língua em uso que podem ser identificados somente com a exploração de corpora específicos, pois esses padrões se tornam evidentes pela recorrência (Mazza, 2009).

Faz-se importante, todavia, saber que Biber & Cortes (2004) notaram, em seu artigo, duas coisas a respeito desses pacotes lexicais: “1) certas palavras ocorrem juntas (como um conjunto); 2) esse conjunto está aberto à adição de outras palavras, para completar o seu significado e a sua gramática”.[13] Assim**,** os autores, em seu artigo, enumeraram alguns tipos estruturais de pacotes lexicais:

1. pacotes lexicais que incorporam fragmentos de frase verbal. Por exemplo, o que você pensa (...)?;

[13] (…) notice two things about these lexical bundles: 1. certain words occur together as a set and 2. that set is open to addition of other words to complete the meaning and the grammar.

2. pacotes lexicais que incorporam fragmentos de orações subordinadas. Por exemplo, eu quero que você (...);

3. pacotes lexicais que incorporam sintagmas nominais e fragmentos de sintagmas preposicionais. Por exemplo, no fim do, ao término de.[14]

Biber *et al* (2009:283) completam essa afirmação ao discorrem sobre algumas características dos pacotes lexicais:

> Primeiro, pacotes lexicais são, por definição extremamente comuns (em contraste com a maioria das expressões idiomáticas e muitos 'padrões gramaticais', que tendem a ser raros). Em segundo lugar, muitos dos pacotes lexicais não são idiomáticos em significado e não são relativamente salientes. E, finalmente, pacotes lexicais usualmente não representam uma unidade estrutural completa[15]. (Tradução minha.)

Por meio das definições propostas nos parágrafos anteriores, buscamos entender melhor o conceito de pacotes lexicais, já que estes compõem a base deste estudo, em que serão analisados aqueles compostos por três palavras.

Esta seção foi dedicada à elucidação do que é LC e à abordagem de alguns dos conceitos que a permeiam. Na próxima seção, serão tratados aspectos relacionados ao uso da LC no ensino de idiomas.

[14] 1. Lexical bundles that incorporate verb phrase fragments, e.g., *what do you think*;

2. Lexical bundles that incorporate dependent clause fragments, e.g., *I want you to*;

3.Lexical bundles that incorporate noun phrase and prepositional phrase fragments, e.g., *the end of the, at the end of*.

[15] First, lexical bundles are by definition extremely common (in contrast to most idioms and any 'grammar patterns', which tend to be rare). Second, most lexical bundles are not idiomatic in meaning and not perceptually salient. And finally, lexical bundles usually do not represent a complete structural unit.

1.3 A Lingüística de *Corpus* no Ensino de Idiomas Estrangeiros

Muitas das pesquisas baseadas em *corpus*, desde o seu início, tiveram como objeto de pesquisa o ensino de línguas estrangeiras. Vários autores como Biber (1998) e Yoon & Hirvela (2004) acreditam que a aplicação das teorias e das ferramentas da LC no ensino é tremendamente benéfica, já que promove a exposição dos alunos a vários textos autênticos, o que facilita a expansão do seu entendimento das palavras em contexto e auxilia também na aprendizagem (ou apropriação) tanto da ordenação quanto da utilização das palavras nas frases.

Vale lembrar que um dos benefícios da utilização de *corpora* na sala de aula é a possibilidade de permitir que o aluno assuma o papel de pesquisador; ou seja, ele analisará (observará) a língua real e, a partir dessa análise, poderá verificar os padrões dessa língua (Leech, 1997).

Há um crescente interesse na utilização de *corpora* na sala de aula, de modo a aumentar a demanda de material didático para o ensino de línguas por meio de colocações (ou *bundles*, como normalmente são conhecidas no ensino de línguas baseado em *corpus*).

Para compreendermos a importância da LC no ensino de idiomas, é importante, primeiro, compreendermos as subdivisões da disciplina. Para tanto, Berber Sardinha (2004:254-255) resume a LC em quatro áreas majoritárias:

(1) descrição da linguagem nativa: essa descrição, normalmente, ocorre em meios acadêmicos e é incorporada aos materiais didáticos;

(2) descrição da linguagem do aprendiz: essa descrição é feita com base na linguagem produzida por aprendizes de língua estrangeira. Do mesmo modo que a anterior, não é utilizada diretamente em sala de aula e ainda está muito restrita ao meio acadêmico;

(3) transposição de metodologias de pesquisa acadêmica para a sala de aula: essa aplicação refere-se ao uso de linhas de concordâncias, em sala de aula, para tirar dúvidas sobre o uso de palavras;

(4) desenvolvimento de materiais de ensino, currículos e abordagens: com a criação de materiais, metodologias ou abordagens baseados na exploração de *corpora* ou nos conceitos da LC.

Este estudo se insere na quarta área, já que se refere à criação de metodologias ou abordagens de ensino inspiradas na exploração de *corpora* ou em conceitos da LC.

Berber Sardinha (2004) formula, pelo menos, três propostas de utilização no ensino: o currículo lexical, a abordagem lexical e o ensino movido por dados (*DDL*). Como essas abordagens ou metodologias não fazem parte do nosso objeto de estudo, nas próximas subseções, apresentaremos as três propostas e as suas principais características de forma breve.

1.3.1 Currículo lexical *(lexical syllabus)*

O currículo lexical foi desenvolvido por Dave Willis, com base nas pesquisas feitas pelos membros do projeto COBUILD, que tinham como objetivo a criação de um dicionário baseado na descrição lexical da linguagem (Willis, 1990). A partir dos dados compilados para a confecção do dicionário, surgiram os livros didáticos da série *Collins Cobuild English Course,* que podem ser considerados exemplos de aplicação dos conceitos e da metodologia da LC.

O curso é, basicamente, centrado no léxico que ocorre com mais freqüência no *corpus* e nos textos que exemplificavam esses itens (Quais???). Em cada nível do curso, havia um número de palavras que deveria ser ensinado, esperando-se que, até o final dele, os alunos tivessem sido expostos a 80% das palavras da língua inglesa.

A metodologia do curso é baseada em tarefas que privilegiam a fluência, e não a exatidão (*accuracy*). Desse modo, o aluno passa a ser o responsável pela aprendizagem, e o professor deixa de ser o conhecedor ou fornecedor do conhecimento.

As suas principais características segundo Berber Sardinha (2004:282) são:

- utilização de *corpora* nos materiais didáticos;
- ensino, primeiro, do que é mais freqüente;
- utilização de linguagem autêntica não-simplificada;
- abordagem de cunho indutivo, ou seja, por meio da qual o aluno descobre as regras pela absorção ou apropriação dos dados.

1.3.2 Abordagem lexical *(lexical approach)*

Criada por Michael Lewis, essa abordagem se baseia em conceitos da LC. Lewis (2000:126) considera que trabalhos baseados em *corpus*, muitas vezes, "contradizem crenças difundidas" [16].

Nessa abordagem, a gramática é vista como algo que tem papel secundário, uma vez que palavras gramaticais são consideradas como itens lexicais.

As porções (ou *bundles*) se configuram em colocações, e o seu ensino —nessa abordagem— é privilegiado. Os *chunks* são ensinados por meio de textos, e os alunos são estimulados a manterem cadernos lexicais em que anotam as porções de língua às quais foram expostos.

As principais características da abordagem, segundo Berber Sardinha (2004:288), são:

- o léxico desempenha papel central no conteúdo e na metodologia;
- há ênfase na colocabilidade do léxico (o que a distingue do currículo lexical);
- o léxico é descrito por meio de porções *(bundles)* léxico-gramaticais ensinadas por intermédio de textos (escritos e falados);
- há a utilização de corpora nos materiais didáticos;
- há a utilização de linguagem autêntica;

[16] (…) contradicting some belief which is widespread.

- há a revisão e a reciclagem regular da linguagem ensinada.

1.3.3 Aprendizado movido por dados *(data driven learning)*

O aprendizado movido por dados foi criado por Tim Johns, inicialmente, para o ensino de gramática de inglês, mas, atualmente, tem sido utilizado noutras áreas, além da gramática, e no ensino de outras línguas.

O preceito da proposta é o do aluno como pesquisador, ou descobridor. Tim Johns (1991:2) considera que "a pesquisa é algo muito sério para ser deixada para os pesquisadores" [17].

Trata-se de uma abordagem por meio da qual se utiliza o computador como elemento central na aprendizagem, como informante; no entanto, este não substitui o professor. Nela, "os alunos interagem com o *corpus* por meio de tarefas de pesquisa"[18] (Mishan, 2004:222), o que pode levá-los a descobrirem como a linguagem "funciona" em contexto real.

A abordagem é essencialmente indutiva; ou seja, é a partir da observação dos dados que os alunos chegam ao conhecimento. As suas principais características, de acordo com Berber Sardinha (2004), são:

- a utilização de linhas de concordâncias como uma das fontes de linguagem;
- o computador se torna um provedor de respostas na forma de exemplos extraídos de um *corpus*;
- foco nas colocações e nos padrões léxico-gramaticais;
- desenvolvimento da autonomia, da autoconfiança dos alunos e da habilidade de descoberta, visando a torná-los pesquisadores;
- abordagem de cunho indutivo: o aluno descobre as regras por meio das observações de dados.

A seguir, explicaremos o segundo arcabouço teórico da pesquisa, tratando do conceito de autenticidade.

[17] (…) research is too serious to be left to the researchers.
[18] (…) learners engaging with the corpus via research tasks.

1.4 Autenticidade

As definições dos termos ‘autêntico’ e ‘autenticidade’ e as suas aplicações no ensino de línguas têm sido motivo de muitas controvérsias ao longo das últimas três décadas. O motivo para tanta controvérsia se originou com a abordagem comunicativa [*communicative language teaching* (*CLT*)] nos anos 70. Os adeptos dessa abordagem rejeitaram aquelas estritamente estruturais para o ensino de LE e abriram-se para o uso de textos autênticos, ou seja, textos que tinham sido criados com um propósito comunicativo genuíno (Mishan, 2005).

Os debates sobre o papel da autenticidade, bem como aquele sobre o que significa ser autêntico, têm aumentado ao longo dos anos. Atualmente, abrange pesquisas de vários campos, como a análise do discurso, a aquisição da segunda língua, a autonomia de estudantes, a informação e as tecnologias de comunicação (ICT), a pesquisa de motivação e o desenvolvimento de materiais, entre outros (Gilmore, 2004).

Como podemos notar, muito se discute a respeito desse tema, seja como textos autênticos ou materiais didáticos autênticos. Esse assunto é o foco desta pesquisa, assim como de outros autores —como Berber Sardinha (2007), Gilmore (2004 e 2007), Mishan (2004 e 2005), Guariento e Morley (2001), entre outros.

A importância do tema, nesta pesquisa, deve ser atribuído ao fato de ela consistir na análise do grau de autenticidade dos textos contidos nos livros selecionados para este estudo (que são estes: *English 900*, *To the Top*, *Interchange*, *Inside Out* e *Touchstone*); ou seja, busca-se, por meio deste trabalho, verificar quão autêntica é a linguagem constante dos textos e dos diálogos desses livros. A premissa é que mesmo textos aparentemente artificiais (isto é, fabricados para serem incluídos nos livros didáticos) podem conter porções autênticas de língua em uso pelos falantes e escritores nativos. Assim, um texto não precisa ser ‘autêntico’ ou ‘não-autêntico’ por completo.

Busca-se, assim, uma melhor definição, ou seja, um melhor entendimento do que realmente é considerado autêntico —ou simplesmente inventado— para os autores citados acima.

Os textos autênticos podem ser utilizados, dentro de sala de aula, na tentativa de construir-ser uma ponte entre o contexto artificial da sala de aula e o contexto do mundo real, facilitando e estimulando, desse modo, a sua aprendizagem, segundo o que Guariento e Morley (2001) comentam no seu artigo. Para os autores, autenticidade não se refere apenas a textos genuínos, mas também à autenticidade das tarefas: a linguagem contida nos textos deverá ser usada com um propósito genuíno, para que os resultados realmente apareçam no ensino da língua-alvo.

Do mesmo modo, a autora Mishan (2004) acredita que autenticidade e autêntico estão relacionados com atributos positivos, puros, originais, etc. Por isso, a sua importância no ensino de línguas, nos materiais e nos textos apresentados aos aprendizes cresce. Mas, para a autora, o que é realmente autêntico pode ser contestável. Por exemplo, ela entende que o fato de mudar--se a forma ou a fonte dos *corpora* já implica a perda da sua autenticidade. Segundo ela, esse é um problema dos *corpora* eletrônicos, pois são compostos de textos —escritos e falados— reais, mas as suas formas originais são obscurecidas.

Mishan (2005) [*apud* Lee (1995:324)] traz, em seu livro, uma definição de texto autêntico:

> Um texto, normalmente, é considerado autêntico se não for escrito para propósitos pedagógicos, mas para um propósito comunicativo da vida real, onde [*sic*] o escritor tem uma mensagem para passar para o leitor. Como tal, um texto autêntico é aquele que possui uma qualidade comunicativa de forma intrínseca.[19]

[19] 'A text is usually regarded as authentic if it is not written for teaching purposes but for a real-life communicative purpose, where the writer has a certain message to pass on to the reader. As such, an authentic text is one that possesses an intrinsically communicative quality'.

Após analisar a literatura sobre esse tema —composta por posições e argumentos de vários teóricos—, , Mishan (2005:18), em seu livro, sintetiza todas as teorias, em cinco "critérios". A autora diz, resumidamente, que autenticidade é um conjunto de fatores:

> 1. Proveniência e autoria do texto.
>
> 2. Propósito comunicativo e sociocultural original do texto.
>
> 3. Contexto original do texto (por exemplo, fonte, contexto sociocultural).
>
> 4. Atividade de ensino gerada pelo texto.
>
> 5. As percepções dos aprendizes e as suas atitudes a respeito do texto e das atividades relacionadas a ele.[20] (Tradução minha.)

Com o mesmo pensamento de Lee (1995), Berber Sardinha (2007:275) entende como texto autêntico "aquele que não foi criado com a finalidade de ensinar língua". O autor também se manifesta em relação aos livros didáticos e ressalta que:

> (...) muitos livros didáticos geralmente não se utilizam de textos autênticos por razões variadas, mas, principalmente, porque sentem necessidade de controlar o vocabulário e a gramática do conteúdo do curso, com base em conceitos de que um texto se torna mais adequado na medida em que incorpora apenas certa quantidade ou tipo de vocabulário e/ou de estruturas gramaticais.

Segundo o autor, a não-utilização de textos autênticos na sala de aula pode ser decorrente da crença de alguns professores em que textos autênticos

[20] "Authenticity is a factor of the:

1. Provenance and authorship of the text; 2. Original communicative and socio-cultural purpose of the text; 3. Original context (e.g. its source, socio-cultural context) of the text; 4. Learning activity engendered by the text; 5. Learners' perceptions of and attitudes to, the text and the activity pertaining to it".

são mais difíceis de ser trabalhados com os alunos, muito embora ele acredite que "não [haja] texto fácil ou difícil do ponto de vista lingüístico". Todavia, Berber Sardinha (2007:282) afirma que não há desculpa para não se utilizar textos autênticos dentro da sala de aula, uma vez que o sucesso de um texto, como parte de uma atividade, depende muito do que será feito com ele, ou seja, de como ele será explorado.

Encontrar uma definição para 'autenticidade' não é uma tarefa muito fácil, mas tentamos mostrar nesta subseção, interpretações e entendimentos — dos principais autores— sobre esse tema. Assim, ressaltamos a importância do trabalho com textos autênticos nos livros didáticos e na sala de aula. Desse modo, explanaremos mais sobre o assunto nas próximas subseções.

1.4.1 Definição de material autêntico

Material autêntico é visto por Peacock (1997) como todo aquele que não foi produzido para o ensino de uma segunda língua. Por exemplo: jornais, revistas, músicas, etc. Ele realizou uma pesquisa tanto com materiais autênticos quanto com materiais artificiais e, a partir dela, ele concluiu que os materiais autênticos, embora fossem ricos em informações —por conterem a língua em uso—, não são considerados muito interessantes, no primeiro momento, aos olhos dos aprendizes; mas, após um período, as vantagens de usá-los no ensino de língua estrangeira são notáveis.

Gilmore (2007:103) ainda sugere, em seu artigo, que:

> Materiais autênticos, particularmente audiovisuais, oferecem uma fonte muito mais rica de contribuição para estudantes e têm o potencial para ser explorados de modos diferentes e em níveis diferentes, para desenvolver a competência comunicativa dos estudantes.[21]

[21] Authentic materials, particularly audio-visual ones, offer a much richer source of input for learners and have the potential to be exploited in different ways and on different levels to develop learners' communicative competence.

Alguns autores —como Little & Singliton (1991)— dizem, por um lado, que textos autênticos aproximam os aprendizes da cultura da língua-alvo, levando-os a aprender de forma mais interessada e motivada. A esse respeito, outros autores, como Williams (1983) e Morrison (1989), acreditam, por outro, que materiais autênticos reduzem a motivação dos alunos, em virtude da sua dificuldade de compreensão de tais materiais, mas todos concordam quanto aos benefícios trazidos pela autenticidade.

Como já dissemos, a LC, para Berber Sardinha (2004), pode ser dividida em quatro áreas, sendo duas delas a descrição da linguagem natural (língua autêntica) e o desenvolvimento de materiais de ensino. Percebemos, desse modo, que, cada vez mais, a LC tem ligação direta com livros didáticos de linguagem autêntica, mesmo porque uma das características inerentes a um *corpus* é ser autêntico. Todo e qualquer *corpus* utilizado nos livros didáticos tem a função de trazer a língua real (ou natural) para a sala de aula, de modo a promover o contato direto dos alunos com a realidade da língua.

Percebemos, então, que tem aumentado o interesse dos autores e das editoras no trabalho com materiais didáticos autênticos. Com isso, todos ganham: nós, professores, já que trabalharemos com um material rico em autenticidade, e os alunos, pois aprenderão com textos e atividades que realmente têm relação com a língua real.

1.4.2 Uso de material autêntico no ensino de língua estrangeira

O uso de materiais autênticos no ensino de língua estrangeira tem uma longa história, segundo Gilmore (2007). Henry Sweet (1899;177), por exemplo, ensinava e escrevia, no final do século XIX, e é considerado um dos primeiros lingüistas a fazer o uso regular de textos autênticos em seus livros. O autor tinha plena consciência das vantagens oferecidas pelos seus livros ao compará-los aos que continham materiais lingüísticos inventados:

> A principal vantagem de textos naturais, idiomáticos em relação aos ' métodos ou ' séries' artificiais é, por um lado, que eles acolhem todas as características do idioma(...) Os sistemas artificiais, por outro, tendem a apresentar repetição incessante de certas construções gramaticais, de certos elementos do vocabulário, de certas combinações de palavras, para a exclusão —quase total— de outros, que são igualmente, ou talvez até mesmo mais, essenciais. [22] (Tradução minha.)

Embora reconheçamos a importância da utilização de dados empíricos para uma descrição mais realista da linguagem, são ainda poucos —embora crescentes em número— os livros didáticos que baseiam o seu conteúdo ou as suas explicações em dados extraídos de *corpora*.

Muitos autores ainda se baseiam em intuições ao produzirem materiais de ensino. Biber (2001:101) afirma que

> O desenvolvimento de materiais para o ensino de línguas e avaliação requer julgamentos sobre o uso da linguagem, pois os autores decidem quais palavras e aspectos da linguagem devem incluir. Essas decisões são, normalmente, baseadas em impressões puramente pessoais e em casos isolados de como falantes e escritores usam a linguagem. Essas impressões, geralmente, não são conscientes e são consideradas verdades inquestionáveis.[23] (Tradução minha.)

Há muitas discussões entre os lingüistas de *corpus* sobre os prejuízos que a intuição de autores de materiais didáticos pode causar ao ensino de línguas. A esse respeito, Biber (2001:101) esclarece que

---

[22] The great advantage of natural, idiomatic texts over artificial 'methods' or 'series' is that they do justice to every feature of the language . . . The artificial systems, on the other hand, tend to cause incessant repetition of certain grammatical constructions, certain elements of the vocabulary, certain combinations of words to the almost total exclusion of others which are equally, or perhaps even more, essential".

[23]The development materials for language instructions and assessment require repeated judgments about language use, as authors make decisions about the words and linguistic features to include. These decisions have usually been based on the author´s gut-level impressions and anecdotal evidence of how speakers and writers use language. These impressions usually operate below the level of consciousness and are often regarded as accepted truths."

> Infelizmente, as intuições dos lingüistas estão, geralmente, erradas. Conseqüentemente, os materiais de ensino e avaliação não proporcionam um reflexo preciso da linguagem realmente utilizada por falantes e escritores em situações reais. [24] (Tradução minha.)

Os autores McEnery & Wilson (1996:104) também discutem a importância da utilização de materiais baseados em *corpus*:

> Materiais não baseados em dados empíricos podem ser enganadores, e os estudos de *corpus* deveriam ser utilizados para informar a produção de materiais para que as escolhas mais comuns de uso recebam mais atenção do que aquelas que são menos comuns.[25] (Tradução minha.)

Em vista do exposto, Mishan (2005:26) afirma que "a motivação é uma das justificativas-chave para o uso de textos autênticos no ensino de língua"[26]. O autor Peacock (1997) concorda com a autora afirma que esse tipo de 'motivação' —citada acima— é considerada "*intrínseca* (a motivação sendo do próprio interesse e curiosidade do aprendiz), que se difere da *extrínseca* (a motivação que vem de forças externas)"[27]. Quando o material didático oferece ao aprendiz um conhecimento autêntico, a motivação para utilizá-lo é apenas uma conseqüência.

Finalmente, diante das cada vez mais numerosas pesquisas baseadas em *corpus*, espera-se que os estudos emanem da academia para os livros didáticos e para outros matérias de ensino. A seguir, discutiremos os métodos e abordagens relevantes a esta pesquisa.

---

[24] Unfortunaly, linguists' intuitions about language use are often wrong. As a result, teaching and assessment materials often fail to provide an accurate reflection of the language actually used by speakers and writers in natural situations.

[25] Non-empirically based materials can be positively misleading and that corpus studies should be used to inform the production of materials, so that the more common choices of usage are given more attention than those which are less common.

[26] (…) the motivation factor is one of the key justifications for the use of authentic texts for language learning.

[27] (…) *intrinsic* (the motivation being the learner's own interest or curiosity), as distinct from e*xtrinsic* (the motivation coming from external forces).

1.5 Métodos e abordagens de ensino relevantes a esta pesquisa

Nesta subseção, cabe explicar a diferença entre 'abordagem' e 'método', não só porque esse assunto é matéria do nosso estudo, mas também porque se trata de um assunto que causa certa confusão com respeito ao ensino de idiomas. Para muitos educadores, os dois termos significam a mesma coisa; a diferença estaria, conseqüentemente, apenas no nome. Diversamente de tal concepção, acreditamos que haja diferença entre método e abordagem, e ela será explicada, abaixo, de forma resumida.

1.5.1 **Abordagem**:

Richard & Rogers (2001), explicam que *"abordagem"* se refere às teorias sobre a natureza da língua e da sua aprendizagem que servem de fonte para as práticas e princípios no ensino de idiomas. Michael Lewis (1993:2), por sua vez, resume *"abordagem"* de modo bem mais prático: "(...) abordagem é o porquê de ensinarmos o que ensinamos e do modo pelo qual ensinamos."[28]

1.5.2 **Método:**

A literatura é farta em definições para método. Jeremy Harmer (2001), entende que método é o que nos permite colocar a abordagem na prática. Métodos incluem vários procedimentos e técnicas como parte do seu corpo- -padrão. Com base nessa premissa, Michael Lewis (1993) resume o método na pergunta "**como vamos colocar a abordagem na prática?**" e acrescenta outras perguntas como

> (1) quais são meios que farão com que o professor ensine eficazmente aos seus alunos? (2) Qual é a melhor maneira para que os alunos aprendam? (3) Como fazer para que aquilo que eu considero importante seja aprendido e retido pelos aprendizes da língua?

[28] (...) approach is why we teach, what we teach and how we teach.

Embora divergências continuem a existir em relação ao conceito de abordagem e método, nesta pesquisa, verificaremos a abordagem implicitamente veiculada num dos elementos centrais, que pode contribuir para o sucesso do processo educativo: o livro didático. É de suma importância enfatizar que somente serão explicitadas, neste trabalho, as abordagens que correspondem aos livros didáticos analisados e pesquisados, sendo elas: audiolingualismo, comunicativismo e abordagem Informada por *corpus*. Essa análise torna-se essencial para o desenvolvimento deste trabalho, pois poderemos verificar se o grau de autenticidade contido nos livros didáticos selecionados tem influência ou não em seu grau de convergência, de acordo com a sua abordagem.

1.6 Audiolingualismo

1.6.1 Definição

O audiolinguismo dominou os programas acadêmicos, nas décadas de 1950 e 1960, nos Estados Unidos. No Brasil, ele começou a se instalar nas escolas no final da década de 1960.

De acordo com Pedreiro (2002), a abordagem audiolingual é a fusão da lingüística estruturalista e da psicologia behaviorista que nos leva à teoria de que aprender línguas é um processo de condicionamento. Preconizavam-se modelos de ensino que, na prática, induziram ao condicionamento e à formação de hábitos —imitação— que se combinam a práticas padronizadas da abordagem audiolingual.

Segundo Sánchez Pérez (1997), a língua, no audiolingualismo, é:

> (...) um conjunto de estruturas hierarquicamente organizadas que tem como finalidade a transmissão de significado. A estruturação da língua se dá em três níveis: fonológico, morfológico e sintático (o nível semântico é pouco relevante nessa concepção lingüística).

No audiolingualismo, a língua é, sobretudo, um fenômeno oral, e é aprendida obedecendo-se a uma ordem hierárquica das habilidades

lingüísticas: ouvir, falar, ler e escrever. Deve-se aprender a falar sem se preocupar com como a língua é estruturada. A língua é ensinada, mas não a sua formação estrutural. Em alguns momentos, o professor utiliza a análise contrastiva, pois se considera que as línguas são diferentes umas das outras, e essa diferença pode causar problemas. A memorização de diálogos e modelos estruturais (*drills*) padronizados são os meios para se conseguirem respostas condicionadas até que, nos cursos mais avançados, chegue o momento de os alunos se expressarem espontaneamente. Os pilares mais robustos da teoria da abordagem audiolingual são as técnicas elaboradas de prática oral e auditiva, assim como a separação pedagógica de línguas em habilidades.

Uma vez que o audiolingualismo é uma abordagem oral de ensino de línguas, parece natural que o processo envolva intensa atividade oral. Dessa forma, o foco é na fala; a gramática é apenas uma coadjuvante por meio de breves explicações ou referências. A língua-alvo é usada para dar as instruções, e a inclusão da língua materna são firmemente desencorajadas. O número ideal de alunos na sala de aula é de, no máximo, dez.

Pedreiro (2002) acredita que, no audiolingualismo, o aprendiz de uma língua estrangeira recebe um estímulo —aquilo que se deseja ensinar na língua-alvo (por exemplo, uma frase falada ou escrita)—; esse estímulo deve produzir uma resposta (por exemplo, a repetição da frase), que, por sua vez, vai gerar o reforço, que é a aprovação do professor e dos colegas à resposta certa na língua-alvo.

No audiolingualismo, o papel do professor é central e ativo. É uma abordagem em que ele desempenha um papel dominante. É ele quem direciona os ensinamentos da língua-alvo, monitora e corrige o desempenho dos aprendizes. Ele tem de saber motivar os alunos por meio da variação dos exercícios e tarefas e saber escolher situações relevantes para a prática das estruturas que pretende treinar com os alunos. Tudo gira em torno do professor e do aluno.

Essa abordagem trouxe vários resultados positivos no que se refere à aprendizagem de uma LE. No entanto, houve quem não se adaptasse a ela. Uns por exaustão em meio a tanta prática e repetição que, muitas vezes, tornava-se sem sentido; outros por sentirem-se com medo de errar, pois o

professor cobrava uma produção correta; outros, ainda, até por terem um estilo de aprendizagem que priorizava o conhecimento da gramática antes da produção oral.

1.6.2 Breve histórico

A Segunda Guerra Mundial revelou muitos horrores, mas também mostrou ao mundo uma necessidade: a de se aprenderem línguas estrangeiras. Os funcionários do governo e das forças armadas norte-americanas, principalmente os das secretarias de inteligência, precisavam, com urgência, aprender a se comunicar nas línguas dos países envolvidos no conflito. Além disso, muitos estrangeiros estavam imigrando para os Estados Unidos e tinham de aprender a se comunicar em inglês. Diante de uma necessidade lingüística tão premente, o governo americano criou o Programa de Treinamento Especializado do Exército (*Army Specialized Training Program*) com o objetivo de elaborar um método de ensino de línguas estrangeiras. Participaram do projeto 55 (cinqüenta e cinco) universidades norte-americanas, e Leonard Bloomfield foi um dos lingüistas envolvidos no programa.

Bloomfield era um lingüista que via a língua como um conjunto de estruturas gramaticais. Pesquisou e descreveu línguas indígenas da América do Norte, o que o levou a elaborar técnicas aprimoradas de descrição lingüística. Era adepto do empirismo, e, por isso, adotou a teoria behaviorista da linguagem como um dos pilares do método de ensino de línguas estrangeiras que desenvolveu, o qual fora encomendado pelo governo norte-americano. De posse da sua teoria estruturalista de língua e da teoria behaviorista da aprendizagem, Bloomfield elaborou aquele que viria a ser o método mais difundido de ensino de línguas estrangeiras: o audiolingualismo, também chamado de abordagem audiolingual.

O audiolingualismo foi muito bem-sucedido durante a guerra, o que levou muitas pessoas a acreditarem que aquele era um bom método de ensino de línguas. Entretanto, após a guerra, percebeu-se que o audiolingualismo não era tão eficiente assim. No entanto, existiam fatores positivos que pareciam

provar a eficiência dessa abordagem para a época, como podemos observar a seguir:

- Um fator era o caráter de imersão que as aulas tinham no período da guerra. O tempo para se aprender a língua era pouco, e, por isso, era necessário que os estudantes estivessem expostos à língua o máximo de tempo possível, justificando-se a imersão.
- O segundo fator era o alto grau de motivação dos estudantes, que precisavam aprender a língua pela razão mais motivadora que existe: a sobrevivência.

Sem esses dois fatores atuando na aprendizagem, o ensino por meio do audiolingualismo não apresentava os mesmos resultados.

A abordagem, que tinha por objetivo criar hábitos automáticos de fala para comunicação, era orientada pelos seguintes princípios:

1. aprender a língua e não sobre a língua;
2. as estruturas deviam ser apresentadas em uma seqüência gramatical;
3. as estruturas deviam ser praticadas até serem automatizadas;
4. as regras gramaticais deviam ser aprendidas intuitivamente, por meio de analogia.
5. os hábitos lingüísticos são formados pela saturação da prática.

Apesar da novidade —de utilizar as quatro habilidades lingüísticas e alertar para o uso real da língua—, essa abordagem, como as anteriores, também se baseou em exercícios de padrões estruturais. O aluno continuava, portanto, a repetir estruturas ‘prontas’.

A desconfiança em relação ao audiolingualismo aumentou com a publicação do texto “Verbal Behavior”, de Noam Chomsky, no final dos anos 1950s. O texto é uma crítica veemente ao behaviorismo, demonstrando a incapacidade dessa teoria de explicar a aquisição da linguagem. O argumento

de Chomsky se centra na criatividade lingüística, a capacidade que o falante-ouvinte tem de produzir e entender sentenças originais.

Para ele, se a teoria behaviorista, centrada na formação de hábitos corretos por meio de um mecanismo de estímulo-resposta-reforço (ou punição) estivesse correta, uma criança de três anos de idade não seria capaz de produzir e entender sentenças que nunca ouviu antes. Contudo, a verdade é que as crianças conseguem produzir e entender sentenças originais mesmo aos três anos de idade (Chomsky, 1978).

Após a decepção com os modelos lingüista-gerativo-transformacionais e a descrença nos pressupostos do audiolingualismo —para sua aplicação no ensino de LE—, inaugurou-se, na área, uma fase transitória, na qual apareceram, no final da década de 60 e ao longo da década de 70, vários métodos ligados aos avanços da Psicologia Humanista.

1.6.3 Material de ensino no audiolingualismo

As principais características do material de ensino no audiolingualismo são apontadas por Pedreiro (2002: 38):

> O novo material [lingüístico] é apresentado em forma de diálogo; há a dependência de imitação (mimetismo), memorização de jogos de frases; as estruturas são seqüenciadas, por meio de análise contrastiva, e ensinadas uma de cada vez; padrões estruturais são ensinados utilizando-se de *drills* de repetição; há pouca ou quase nenhuma explicação sobre a gramática: esta é ensinada por analogia indutiva, e não por explanação dedutiva; o vocabulário é limitado e aprendido em contexto; é vasto [*sic*] o uso de fitas cassetes, laboratórios de línguas e recursos audiovisuais; é dada muita importância à pronúncia; a língua materna não deve ser usada; as respostas corretas são prontamente reforçadas porque se prestam a procurar levar os alunos a produzirem falas sem erros; há uma tendência de manipular a língua e desconsiderar o significado, ou seja, o que se quer realmente transmitir, comunicar.

Em seu artigo, Ticks (2005) utiliza a classificação de Kumaravadivelu (2004), em que este divide as abordagens e métodos em três grandes grupos: *centrados na linguagem, centrados na aprendizagem e centrados no aprendiz*. O Audiolingualismo, segundo essa classificação, faz parte do grupo *"centrado na linguagem"*, uma vez que reúne as teorias cujo foco é a forma da língua.

Façamos, aqui, uma segunda ressalva no sentido de esclarecer que utilizamos a classificação de Kumaravadivelu (2004) apenas como ponto de partida para explicar e exemplificar os materiais didáticos nas abordagens selecionadas para esta pesquisa.

As abordagens *centradas na linguagem* promovem oportunidades para que os aprendizes pratiquem estruturas lingüísticas pré-selecionadas, pré--seqüenciadas, por meio de exercícios cujo foco está na forma, assumindo que a preocupação com ela resultará, em última instância, em aprendizagem. O professor introduz estruturas gramaticais e itens de vocabulário, um de cada vez, e ajuda o aluno a exercitá-los até a internalização.

As atividades *centradas na linguagem* assumem a forma de exercícios de pronúncia e/ou de repetição de itens lexicais, também conhecidos como *drills*. Após ouvir e ler, o aluno repete, conforme exemplo 1 (Ticks, 2005).

Listen and repeat

Bank coffee shop bakery pharmacy movie theater

open-air market

Train station restaurant library bookstore

parking lot post office

Exemplo 1 – Exercício centrado na linguagem

No audiolingualismo, os materiais didáticos continham muitos exemplos artificiais, e os professores, influenciados pelo behaviorismo, tentavam impedir as hesitações e as pausas, contrariando as características naturais das interações conversacionais. Havia uma ênfase exagerada nas diferenças entre as duas línguas, de modo que se ignoravam as semelhanças e oferecia-se pouco estímulo às atividades mentais, pois se almejava o automatismo. A abordagem costumava agradar às crianças e desagradar aos adultos por causa das atividades —cansativas— de repetição.

Os críticos do audiolinguismo diziam que, de tanto repetir as mesmas frases, os aprendizes criariam hábitos automáticos e se transformariam em papagaios. Em um dos livros da época (English 900, 1964) havia o seguinte diálogo:

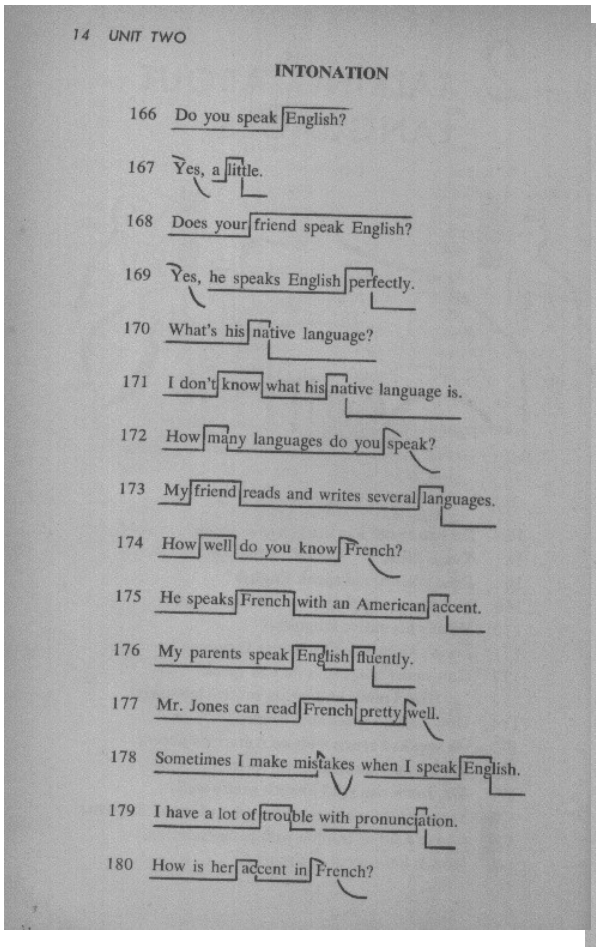
A: — Where do you come from?

B: — I come from a little town not far from here.

Com base na exposição a essa abordagem, diante da pergunta "*Where do you come from?"*, os aprendizes responderiam, automaticamente, "*I come from a little town not far from here"*. Assim, o audiolingualismo ignorava a capacidade do ser humano de pensar e produzir sentido, mesmo que ele nunca tenha antes ouvido ou produzido determinado enunciado.

Nesta pesquisa, utilizaremos dois livros didáticos, da abordagem audiolingual, que serão avaliados por meio dos seus textos e diálogos. São eles: *English 900,* publicado em 1964, pela editora Collier Publishing, e *To the Top: Way Ahead*, publicado em 2006, pela Waldyr Lima Editora.

Segundo Paiva (2005), o livro *English 900*, foi um material muito usado no Brasil, na década de 70. Abaixo, exibimos duas páginas do livro:



14 UNIT TWO

INTONATION

166 Do you speak English?

167 Yes, a little.

168 Does your friend speak English?

169 Yes, he speaks English perfectly.

170 What's his native language?

171 I don't know what his native language is.

172 How many languages do you speak?

173 My friend reads and writes several languages.

174 How well do you know French?

175 He speaks French with an American accent.

176 My parents speak English fluently.

177 Mr. Jones can read French pretty well.

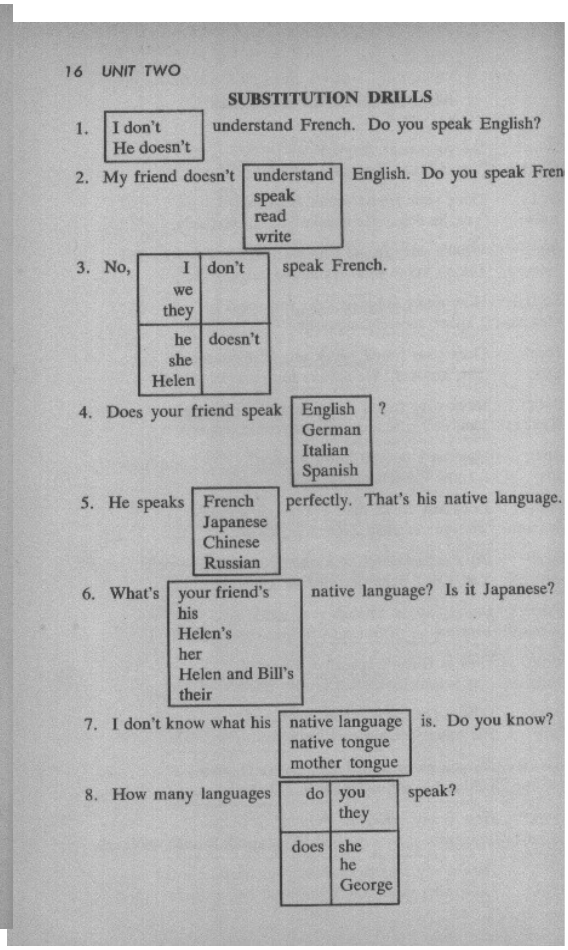
178 Sometimes I make mistakes when I speak English.

179 I have a lot of trouble with pronunciation.

180 How is her accent in French?

FIGURA 1 – Marcas de entonação

(English 900, book two, p.14)



16 UNIT TWO

SUBSTITUTION DRILLS

1. I don't / He doesn't understand French. Do you speak English?
2. My friend doesn't [understand / speak / read / write] English. Do you speak Fren
3. No, [I / we / they] [don't] / [he / she / Helen] [doesn't] speak French.
4. Does your friend speak [English / German / Italian / Spanish] ?
5. He speaks [French / Japanese / Chinese / Russian] perfectly. That's his native language.
6. What's [your friend's / his / Helen's / her / Helen and Bill's / their] native language? Is it Japanese?
7. I don't know what his [native language / native tongue / mother tongue] is. Do you know?
8. How many languages [do / you / they] / [does / she / he / George] speak?

FIGURA 1 – Exercícios estruturais

(English 900, book two, p.14)

Na figura 1, percebe-se a preocupação com a entonação na apresentação das estruturas da lição. Essas mesmas estruturas eram, depois, trabalhadas em forma de breves diálogos e de exercícios estruturais.

Na figura 2, temos uma amostra dos exercícios estruturais, de substituição, que eram trabalhados da seguinte forma: o professor repetia as estruturas básicas várias vezes e, depois, ia promovendo a substituição de um determinado vocábulo ou sintagma. No exercício 4, por exemplo, após repetir, várias vezes, a frase "*Does your friend speak English?*", o professor enunciaria a palavra *German* e os alunos fariam a substituição repetindo a frase "*Does your friend speak German?*", e assim sucessivamente.

Nos outros tipos de exercício, as orações eram trabalhadas por meio da sua recombinação ou transformação em formas negativa, interrogativa, afirmativa, passiva, etc. Eram também comuns os exercícios preparados pelo

autor do livro didático com o objetivo de propor ao estudante que este completasse lacunas. Alguns materiais traziam os pares mínimos (*minimal pairs*) para que, por meio deles, os estudantes pudessem treinar os sons, nos seus contrastes mínimos. Por exemplo: *sheep/ship; chin/tin; beach/bitch; bet/bat*; etc.

Acredito que nenhum profissional que tenha vivenciado esse período da história do ensino de línguas negaria que o método teve o mérito de proporcionar aos aprendizes alguma fluência, pois as habilidades orais eram desenvolvidas desde o primeiro dia do curso. As fitas gravadas forneciam bons modelos de pronúncia e entonação; as dificuldades eram graduadas; era, assim, possível prever quais erros poderiam ser cometidos, e, diante dessa previsão, oferecer a prática necessária para tentar impedir o seu cometimento (Paiva, 2005).

O outro livro que compõe o *corpus* de pesquisa deste estudo é o *To the Top,* que é considerado audiolingual, pois também se compõe dos famosos *drills* e de outras características da abordagem. Esse material didático foi escolhido porque é utilizado por uma escola de idiomas de grande porte, que está no mercado há cerca de 50 anos e tem, aproximadamente, 250 mil alunos; essa escola é adepta e defensora do audiolingualismo e norteia o seu método por ele até hoje.

1.7 Comunicativismo

1.7.1 Definição

A *abordagem comunicativa* se caracteriza pelo foco no sentido, no significado e na interação propositada entre os sujeitos que estão aprendendo uma nova língua (Almeida Filho, 1993).

Para o autor (1993), o ensino comunicativo de LE é aquele em que se organizam as experiências de aprender em termos de atividades/tarefas de real interesse e/ou necessidade para o aluno, para que este se capacite a usar a língua-alvo na prática de ações autênticas na interação com outros falantes--usuários dessa língua. Além disso, os adeptos desse tipo de ensino não

tomam as formas da língua descritas nas gramáticas como modelos suficientes para a organização das experiências de aprender uma LE, embora não descartem a possibilidade de criar, na sala de aula, momentos de explicitação de regras e de prática de rotinização dos subsistemas gramaticais, como o dos pronomes, o das terminações de verbos, etc.

Nunan (1993) [*apud* Brown (1994)] lista cinco características da *abordagem comunicativa*:

● ênfase no aprender a comunicar-se por meio da interação com a língua-alvo;

● a introdução de textos autênticos na situação de aprendizagem;

● a provisão de oportunidades para os alunos, não somente na linguagem, mas também no processo de sua aprendizagem;

● intensificação das experiências do aluno como elementos importantes na contribuição para aprendizagem em sala de aula; e

● tentativa de ligar aprendizagem da linguagem (em sala de aula) com a ativação da linguagem (fora dela).

A competência comunicativa é entendida como um sistema que inclui sistemas subjacentes de conhecimento e habilidades requeridas para que haja a comunicação (por exemplo, conhecimentos de vocabulário e habilidade de usar as convenções sociolingüísticas da língua), de acordo com Canale e Swain (1980). Os autores também definem a competência comunicativa como composta de quatro competências interligadas, cujo desenvolvimento paralelo tornará o aprendiz de LE proficiente na língua-alvo. As competências a serem desenvolvidas são (1) a gramatical, (2) a sociolingüística, (3) a discursiva e (4) a estratégica.

Canale e Swain (1980) também contribuíram para uma melhor compreensão do conceito de competência comunicativa, por meio da propositura de um modelo constituído de quatro componentes, que são: (1) a **competência gramatical** (código da língua: vocabulário, morfologia, pronúncia, ortografia, etc.), (2) a **competência sociolingüística** (fatores sociais *versus* fatores lingüísticos), (3) a **competência discursiva** (unidade

temática: coesão e coerência) e (4) a **competência estratégica** (técnicas / estratégias comunicativas para a eficácia do discurso). Segundo os autores, cada um desses componentes deve receber igual ênfase, desfazendo-se, assim, o mal-entendido de que, em uma abordagem comunicativa, o componente gramatical deva ser ignorado.

À luz do comunicativismo, definir o que vem a ser ensinar uma língua falada não é tarefa fácil, porque os princípios comunicativos podem ser aplicados ao ensino de qualquer habilidade, em qualquer nível, e, por causa da vasta variedade de atividades de sala de aula e tipos de exercícios discutidos na literatura sobre o ensino comunicativo de línguas, descrever os procedimentos típicos —para o uso em uma aula baseada nos seus princípios— não é possível (Richards & Rodgers, 1986).

Por fim, Brown & Levinson (1987) afirmam que talvez seja impossível sintetizar todas as definições —a respeito do que venha a ser o ensino comunicativo de línguas— que têm sido oferecidas.

1.7.2 Breve histórico

A abordagem comunicativa foi inspirada por uma determinada leitura da teoria lingüística do norte-americano Noam Chomsky e das teorias —de psicologia cognitiva— de Piaget e Vygotsky. Foi motivada pela crescente demanda por métodos mais eficazes de ensino de línguas; surgiu como uma forte reação contra o audiolingualismo.

Na abordagem comunicativa, a unidade básica da língua, que requer atenção, é o ato comunicativo, em vez da frase. A função se sobrepõe à forma, de modo que o significado e as situações são os elementos inspiradores da planificação didática e da confecção de materiais. A competência comunicativa é o objetivo dessa abordagem —em vez do acúmulo de conhecimento gramatical ou da estocagem de formas memorizadas.

Segundo Richards & Rodgers (1986:64), as origens do ensino comunicativo de línguas, ou seja, as mudanças na tradição do ensino

britânico, surgiram no final da década de 1960. Assim que a teoria lingüística subjacente ao audiolingualismo foi rejeitada nos Estados Unidos, na metade dos anos 1960s, lingüistas aplicados começaram a questionar as hipóteses teóricas subjacentes ao ensino situacional de língua, que, até então, representava o principal método britânico para o ensino de inglês como LE.

Lingüistas aplicados britânicos começaram, então, a enfatizar o potencial funcional e comunicativo da língua, já que sentiam a necessidade de focar-se no ensino de línguas e na proficiência comunicativa, em vez de no mero domínio de estruturas. Defensores dessa visão de língua valeram-se dos trabalhos de lingüistas funcionais, como John Firth, M.A.K. Halliday; de trabalhos de sociolingüistas norte-americanos, como Dell Hymes, John Gumperz e William Labov, assim como do trabalho de filósofos da linguagem, como Austin e Searle.

As mudanças que vinham ocorrendo no cenário educacional da Europa também impulsionavam a criação de diferentes abordagens ao ensino de LE, e, com a interdependência crescente dos países do Mercado Comum Europeu, tornou-se visível a necessidade de criação de métodos alternativos para o ensino de línguas. Isso se tornou possível graças à intervenção do Conselho da Europa, que passou a patrocinar conferências internacionais sobre o ensino de línguas, e desempenhou papel ativo na formação da Associação Internacional de Lingüística Aplicada.

Começava a haver um interesse por estudos das necessidades dos aprendizes de línguas européias, e, em 1972, David Wilkins publicou um documento por meio do qual propôs uma definição funcional ou comunicativa da língua, a qual contribuiu para uma análise dos sentidos comunicativos de que um aprendiz de LE precisa para compreendê-la e nela expressar-se.

Mais tarde (1976), Wilkins revisou e expandiu esse documento de 1972, publicando o livro intitulado *Notional Syllabuses*, que é considerado o marco do movimento comunicativo nocional-funcional. Nessa obra, ele discute três abordagens de ensino —a gramatical, a situacional e a nocional—, apontando as limitações das duas primeiras e as vantagens proporcionadas pela última. Às duas categorias de sentido apresentadas em

seu documento de 1972, ele acrescenta mais uma —a **categoria de significado modal**— e sugere a inclusão de itens de significado ou uso, ao lado dos itens gramaticais, nos planejamentos de curso.

Entretanto, o planejamento nocional de Wilkins também recebe várias críticas de autores como Widdowson (1979), que declara que Wilkins apresenta listas de funções isoladas, da mesma forma por que um planejamento estrutural apresenta as unidades gramaticais. Para Widdowson, o que esse planejamento não faz é representar a linguagem como discurso, não podendo, assim, contribuir para o desenvolvimento da competência comunicativa, que ele define não como uma compilação de itens na memória, mas como um conjunto de estratégias ou procedimentos criativos, que se presta à realização dos valores de elementos lingüísticos em contextos de uso. Sendo assim, Widdowson acredita que o foco de atenção do planejamento nocional deva estar em itens, e não em estratégias, em componentes do discurso, nem no processo da sua criação. Nesse sentido, tal concepção não se difere, essencialmente, do planejamento estrutural, que também lida com itens e componentes. Para Widdowson, trata-se da mesma mercadoria, porém disfarçada como um rótulo novo, e ele cria essa imagem usando a conhecida citação "*old wine in new bottles – or perhaps old wine in old bottles with new labels."*

Contudo, apesar de apontar para algumas limitações do planejamento nocional, Widdowson acredita que a mudança de enfoque da oração para a noção representa um avanço, embora ele não esconda que ainda resta muito a ser feito.

Após a publicação de Wilkins, muitos outros trabalhos surgiram: Widdowson (1978, 1979, 1984, 1989, 1990), Brumfit & Johnson (1979), Breen Candlin (1980), Canale & Swain (1980), Canale (1983), Castaños (1989), Almeida Filho (1985, 1986, 1990, 1993), todos eles, dentre muitos outros, com o intuito de firmar bem as bases teóricas desse movimento. Essa busca —incessante— pelo estabelecimento dos princípios orientadores dessa metodologia resultou em mudanças ao modelo que Wilkins propôs inicialmente.

> Não foram poucas as contribuições que esses e outros trabalhos trouxeram ao ensino de línguas, e, ainda que elas tenham sido mais teóricas do que práticas, visaram a melhor estabelecer as bases norteadoras do comunicativismo e, assim, a orientar o professor de LE a tomar novos rumos na sua prática pedagógica.

Essa abordagem, entretanto, não escapou de questionamentos referentes aos próprios princípios nos quais ela se embasa. Os termos competência (de Chomsky) e, mais tarde, competência comunicativa (de Dell Hymes), têm sido alvo de muitos artigos e discussões como os de Thompson (1996) e Castaños (1989).

Apesar de existir, ainda hoje, bastante controvérsia no que tange ao conceito de competência comunicativa, Kumaravadivelu (1994) trouxe um ponto novo para a área de ensino e aprendizagem de LE: após o seu estabelecimento, começou-se a falar em contextos situacionais, registros de fala, funções comunicativas e outros. Sem dúvida, a abordagem comunicativa representa uma evolução inteligente em direção a um processo de ensino e aprendizagem de LE mais humano e centrado nos interesses do aprendiz. É a abordagem comunicativa a que inspira, hoje, os métodos mais eficazes.

1.7.3 Material de ensino no comunicativismo

Segundo a classificação de Kumaravadivelu (1994) explicada acima, a abordagem comunicativa pertence à última categoria de abordagens, ou seja, à *centrada na função.* Nela, o material didático propicia ao aluno a oportunidade de praticar estruturas pré-selecionadas, pré-seqüenciadas, focando-se não apenas na forma, mas também em propriedades funcionais e nocionais da linguagem. As preocupações com a forma e a função resultarão, em última instância, em aprendizagem. O professor introduz itens formais e funcionais, um de cada vez, e ajuda o aluno a praticá-los até que os internalize.

Os livros oferecem diálogos pré-selecionados de forma estruturada para que o aluno reproduza falas pré-determinadas pelo material, nas quais funções da linguagem são focalizadas.

Ticks (2005) oferece, no seu artigo, exemplos de exercícios e de atividades de material didático baseado no comunicativismo, conforme demonstraremos no exemplo 1.

Exemplo 1 – Exercício centrado na função

– Livro Azul (Unidade 1, p.4)

Practice introductions with a partner. Learn your classmates' names.

A: Hi. My name´s_____

B: Hi,_____. Nice to meet you.

A: Nice to meet you, too.

As atividades dos materiais didáticos *centrados na função* exploram, portanto, recortes de eventos comunicativos, que são fatiados possivelmente para que os alunos possam imitá-los (Ventola, 1987, p.399) e "aprender mais facilmente".

Essas atividades segundo TICKS (2005) se configuram, por exemplo, na forma de pequenos diálogos convencionalizados pelos livros e apresentam, via de regra, um enunciado indicando o que o aluno deve fazer e um modelo de diálogo, conforme demonstraremos no exemplo 2.

Exemplo 2 – Exercício centrado na função

How to get someone´s attention/

ask about price/agree to buy

Pair practice

Practice the conversation and vocabulary with a partner.

Use your own words.

Segundo o autor Sánchez Pérez (1997) as tipologias das atividades, no comunicativismo, são:

1. atividades interativas, uma vez que a comunicação é fundamentalmente interação e pressupõe o intercâmbio de informação entre dois ou mais interlocutores;
2. atividades fundamentadas na transmissão de conteúdo que seja de interesse para aqueles que intervêm na comunicação;
3. uso de vários tipos de auxílio para o esboço de atividades: gibis, filmes, tarefas, etc;
4. as atividades devem estar baseadas na realização de uma tarefa concreta. Requerem, portanto, uma funcionalidade;
5. as atividades devem ser variadas, de acordo com a riqueza que caracteriza as situações comunicativas.

O autor também ressalta elementos pedagógicos no desenho das atividades nessa abordagem:

1. nas atividades, é tão importante o procedimento (o como), quanto o conteúdo (o quê);
2. as atividades devem conduzir, sempre, para o estabelecimento de uma situação comunicativa pertinente;
3. na seleção dos materiais implicados pelas atividades, deve-se ter em conta uma progressão de elementos menos (dos menos até os mais complexos);
4. o aprendizado é o resultado de processos dedutivos e indutivos (prática e

compreensão).

O autor afirma também que a motivação do discente constitui um elemento central e chave: só assim propiciar-lhe-á a aprendizagem individual, que é a única aprendizagem verdadeira. E acrescenta a importância do contexto nessa abordagem. Deve-se considerar tudo aquilo que caracteriza o processo comunicativo global: a existência de outros componentes, além do gramatical.

Nesta pesquisa, utilizaremos dois livros didáticos da abordagem comunicativa, que serão avaliados por meio dos seus textos e diálogos. São eles: *Interchange: English for international communication*, publicado em 1991, pela editora Cambridge University Press e *Inside Out,* publicado em 2004, pela editora Macmillan.

Segundo resenha encontrada neste *site* (http://www.americanas.com.br/home/begin.do?home=AcomProd&departmentId=1472&itemId=786650#features), *Interchange* é um dos cursos de inglês mais bem-sucedidos do mundo para alunos jovens e adultos (do nível básico até o intermediário). Oferece uma gramática abrangente, fácil de ser compreendida e prática, mais oportunidades para desenvolver as habilidades de audição e fluência, assim como um novíssimo conteúdo em cada unidade.

Os livros *Interchange* trazem também tópicos fundamentais, como foco em fluência e precisão, índice dividido por unidades e seus respectivos temas, estruturas, funções, vocabulário e pronúncia.

De acordo com o *site* da Macmillan (http:/www.macmillanenglish.com/insideout/HTML/Original-pages/Original_insideout_Home.htm), o livro *Inside Out* é "um curso de inglês testado em sala de aula, projetado para desenvolver habilidades comunicativas da vida real. Escrito por professores, ele é um curso agradável e animado para adultos e jovens adultos."

1.8 Abordagem informada por *corpus*

1.8.1 Definição

A abordagem informada por *corpus* é descrita por McCarthy (1998) como "(...) aquela que enfatiza a estrutura pedagógica, das informações, da língua organizada por *corpus* em ilustrações significativas e contextualizadas de atividades comunicativas reais".[29]

Perez-Llantada (2009) comenta que

> (...) essa abordagem pode ter ajudado estudantes a colocarem em prática os seus conhecimentos de uso gramatical, contribuindo, assim, para a superação do problema do conhecimento inerte, que é associado às abordagens tradicionais de ensino de gramática.[30]

A abordagem informada por *corpus* pode conferir a estudantes e professores mais confiança quanto ao uso do idioma, que é essencial para uma boa comunicação. Dessa forma, esse idioma lhes permite falar com mais precisão, levando-os a alcançar as suas metas mais rápida e eficazmente (McCarthy, McCarten & Sandiford, 2006).

Essa abordagem há de trazer muitos benefícios e informações tanto para nós, professores de LE, quanto aos aprendizes, pois, por meio do *corpus*, poderemos não apenas aprender, mas também vivenciar a língua real sem sairmos da sala de aula.

Temos o dever de enfatizar, nesta pesquisa, que a abordagem informada por *corpus* não tem caráter de ensino, ou seja, ela não é ensino como a abordagem audiolingual e a abordagem comunicativa; por enquanto, só é

---

[29](…) that emphasizes the pedagogical framing of corpus-rendered language data into purposeful and contextualized illustrations of actual communicative activity."(Lingua e linguagem neste estudo serão usadas de forma permutável).

[30] (...) this approach may have helped the students put into practice their knowledge of grammar usage, thus contributing to overcoming the inert knowledge problem associated with the traditional approach to teaching grammar.

utilizada na elaboração de material didático.

1.8.2 Breve histórico

Recentemente, a tecnologia desenvolvida para as pesquisas em LC tem demonstrado muito potencial para a aplicação no ensino LE, tendo resultado em instruções por meio de integração de vocabulário, gramática, e padrões de discurso de determinados tipos de escrita (Gledhill, 2000; Hyland, 2002; Jabbour, 1997, 2001;Tribble, 1999, 2002).

Um número expressivo de estudos baseados em *corpus* tem implicado o desenvolvimento (1) de programas de estudo informado por *corpus*, (2) de materiais pedagógicos, e (3) de atividades para sala de aula (Conrad, 1999; Flowerdew, 1998; Thurstun & Candlin, 1998). Os resultados desses estudos têm enfatizado que a abordagem com base em *corpus* não só pode aumentar a consciência dos estudantes quanto à padronização léxico-gramatical dos textos, mas também pode nutrir a aprendizagem indutiva.

Considerando que a pesquisa baseada em *corpus* teve muito impacto no rápido desenvolvimento de materiais didáticos, pesquisadores começaram a olhar para o discurso escrito acadêmico, em combinação com a análise de gênero, para informar os materiais de ensino de inglês com propósito acadêmico (J. Flowerdew, 2002) e "ajudar estudantes a desenvolver competência como escritores dentro de domínios acadêmicos específicos" [31] (Tribble, 2002, p.131).

Atentos à escassez dos estudos na área, Yoon e Hirvela (2004) examinaram o comportamento e as atitudes de estudantes —de *corpus*— de inglês como língua estrangeira. Lançando mão de análises quantitativas e qualitativas, eles constataram que o uso de *corpus* no ensino da LE ajudou os estudantes a aprenderem padrões, de prática geral, de palavras que resultaram no aumento da sua confiança no ato da escrita na língua-alvo.

[31] help students to develop competence as writers within specific academic domains.

Gaskell & Cobb (2004) concordam em pesquisas preliminares, que *concordâncias* também podem ajudar estudantes —de nível intermediário— na aprendizagem da gramática. Eles promoveram a avaliação de escrita guiada por dados em relação aos erros típicos dos estudantes, usando um *software* de concordâncias *on-line*. Os estudantes foram direcionados a *links* e deveriam, com base no *software*, corrigir todo e qualquer erro com que se deparassem. Após a pesquisa, os autores concluíram que, embora os resultados não indicassem uma diminuição dramática no cometimento dos erros —por parte dos de estudantes—, muitos destes afirmaram que as concordâncias foram-lhes úteis e poderiam ser uma rica fonte de informação.

Assim, a abordagem informada por *corpus* vem-se fortalecendo em pesquisas e na elaboração de materiais didáticos, comprovando a todos que a autenticidade das informações contidas na prática de ensino e nos materiais didáticos para a aprendizagem de LE é de suma importância para se construir um conhecimento da língua real.

1.8.3 Material de ensino na abordagem informada por *corpus*

Um *corpus* pode ser um recurso muito rico para escritores de livros didáticos e de outros materiais de ensino de LE, porque nos pode propiciar uma visão detalhada de como os nativos daquela língua falam e escrevem em situações cotidianas. Pode-nos fornecer informação sobre vocabulário, gramática, formalidade e informalidade, diferenças entre o idioma falado e o escrito, como se executam funções básicas; como se fazem pedidos, cumprimentos e desculpas; como as pessoas iniciam e encerram conversações; como mudamos o assunto; como interrompemos um ao outro; como fazemos perguntas, e muito mais.

Segundo McCarthy (2004), “um *corpus* também pode prover estatísticas muito úteis para ajudar os escritores de livros voltados ao ensino de LE a apresentarem itens gramaticais da melhor forma”. [32]

[32] A corpus can also provide very useful statistics to help textbook writers present grammar items in the best way.

Professores e estudantes devem esperar que, na maioria das vezes, materiais didáticos informados por *corpus* pareçam-se com materiais tradicionalmente preparados. Certamente, professores não precisarão de nenhum conhecimento adicional para usá-los. Porém, materiais informados por *corpus*, de acordo com McCarthy (2004), são genuinamente especiais pelos seguintes motivos:

1- estão baseados em uso atual;

2- os exemplos contidos neles, embora, às vezes, possam ser editados ou adaptados, são um reflexo do uso real; não são inventados;

3- o estudo (ou seja, tanto os itens a serem ensinados como também a seqüência em que eles serão apresentados) é informado por meio de freqüência. Por exemplo, podemos priorizar gramáticas e vocabulários que são mais freqüentes e mais úteis;

4- os contextos —nos quais palavras e estruturas de gramática são usadas— são autênticos, baseados nos contextos que compõem o *corpus*;

5- a apresentação e as atividades podem focar nas diferenças importantes entre a língua falada e escrita;

6- os materiais podem incluir língua que foi ignorada ou não notada no passado, mas faz parte da comunicação real;

7- *corpora* especializados podem ser analisados para satisfazer as necessidades de um grupo particular de estudantes;

8- estudantes não têm de viver no ambiente da língua-alvo para experimentarem a língua autêntica.

Sendo assim, embora possam não parecer muito diferentes dos textos tradicionais, os materiais informados por *corpus* são revolucionários no modo pelo qual trazem para a sala de aula o uso real da língua.

As vantagens de usar-se *corpus* no desenvolvimento de materiais de ensino são indubitavelmente muitas, mas não podemos perder de vista a necessidade de equilibrar o uso de dados de *corpus* no material didático.

Para esta pesquisa, utilizaremos o livro *Touchstone*, publicado em 2006, pela editora Cambridge University Press. Para alcançar o equilíbrio comentado acima, os autores de *Touchstone* gastaram horas interpretando e mediando a pesquisa de seu *corpus*, tendo em mente três principais metas (McCarthy, McCarten & Sandiford, 2006):

1- identificar língua autêntica e motivadora;

2- envolver os seus achados em um programa cuidadosamente feito;

3- criar livros voltados ao ensino de LE que contenham estruturas familiares e fáceis de usar.

Abaixo, apresentamos um exemplo de exercício —retirado do livro *Touchstone*— encontrado em McCarthy (2004:8). Para o autor, este tipo de atividade dá aos aprendizes a chance de associar o nome das cores a informações das pessoas, e isso os levará a uma aprendizagem efetiva:

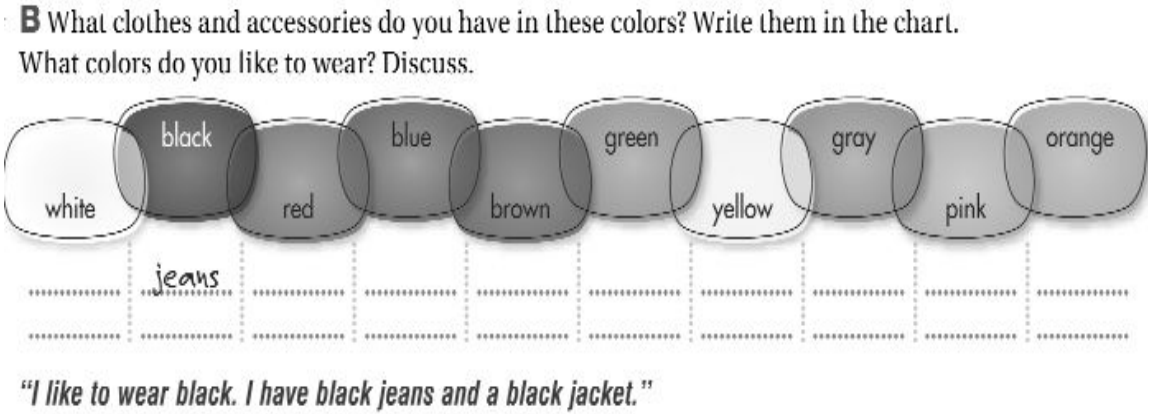


Figura 3: Word sort activity – *Touchstone,* Level 1, Unit 8B, Cambridge University Press.

Desse modo, esperamos que o número de livros didáticos informados por *corpus* aumente muito, com base nas descobertas e análises feitas pelos pesquisadores, sejam eles os alunos, professores ou os próprios autores dos livros didáticos.

Em resumo, o *corpus* informa os escritores de livro didático, que, depois, usam tais informações para desenvolverem atividades familiares com o idioma autêntico, que reflete contextos naturais do *corpus*. Isso ajuda os estudantes a

usarem o idioma mais naturalmente nas próprias conversações dentro e fora da sala de aula.

.

Capítulo 2

Metodologia

Este capítulo é constituído de três partes: a primeira, uma descrição dos objetivos e das questões que nortearam esta pesquisa; a segunda, uma descrição dos *corpora* utilizados, bem como a especificação dos procedimentos de coleta; e a terceira, a demonstração das ferramentas empregadas e a explicação dos procedimentos de análise dos dados.

2.1 Objetivos e questões de pesquisa

A pesquisadora teve como objetivo geral verificar (1) se os autores de livros didáticos atuais —para o ensino de língua inglesa— empregaram, de fato, textos autênticos e, (2) até que ponto, os autores de livros didáticos de décadas passadas, numa perspectiva diacrônica, também o fizeram.

O objetivo específico desta pesquisa foi o de, a partir da análise de padrões léxico-gramaticais (Sinclair, 1991; Hunston & Francis,1999; Jacobi, 2001), ou melhor, de pacotes lexicais (*bundles*) constantes dos *corpora* de estudo e de referência, descobrir (1) quais e quantos pacotes lexicais existem em cada livro didático do *corpus* de estudo e (2) qual é o grau de convergência e/ou divergência dos livros didáticos em relação aos *corpora* de referência —*British National Corpus* (*BNC*) e *Google Corpus*— e, a partir dessas descobertas, (3) verificar o grau de autenticidade —do uso lingüístico— dos livros pesquisados.

A fim de atingir esses objetivos, foram elaboradas as seguintes questões de pesquisa:

1) Quais são os pacotes lexicais existentes no *corpus* de estudo?

2) Qual é o grau de convergência e/ou divergência entre os pacotes lexicais constantes dos *corpora* de estudo e de referência ?

3) Qual é o grau de convergência e/ou divergência entre os pacotes lexicais constantes dos *corpora* de estudo e de referência tendo em vista cada texto dos livros didáticos?

4) Quais livros didáticos apresentam maior e menor grau de autenticidade, tendo em vista os índices de convergência e divergência?

2.2 *Corpora* e critérios de elaboração e coleta

A pesquisa baseou-se nos princípios da Lingüística de *Corpus* (LC) ao fazer uso de dois *corpora*, que são: (1) os *corpora* dos livros didáticos de ensino de inglês como língua estrangeira (*corpus* de estudo) e (2) o *British National Corpus* e o *Google Corpus* (já que ambos constituem o *corpus* de referência).

Além de fornecer o fundamento teórico deste estudo, os princípios da LC também serão utilizados para o tratamento dos dados, a sua análise e a sua interpretação. A LC, segundo Hunston (2002), possibilita a adoção de uma abordagem empírica de observação de dados, em formato eletrônico, que permitirá a análise deles tanto quantitativamente quanto qualitativamente, além do levantamento de pacotes lexicais da língua estudada, bem como a determinação dos pacotes lexicais recorrentes ou exclusivos de todos os *corpora*.

2.2.1 *Corpora* de estudo

Como já dissemos no capítulo anterior, de fundamentação teórica, o arcabouço teórico da LC tem contribuído, de maneira decisiva, para que os conteúdos dos materiais didáticos para o ensino de LE deixem de ser escolhidos segundo a simples intuição dos seus

autores.

Quanto à sua tipologia, segundo critérios listados por Berber Sardinha (2004:20-22), cabe esclarecer que se trata de um *corpus* de estudo:

- escrito para ser falado;
- contemporâneo (representa o tempo corrente);
- de amostragem (composto por porções de textos ou variedades textuais, planejado para ser uma amostra finita da língua estudada);
- de conteúdo especializado (textos de tipo específico; neste caso, de linguagem oral e informal);
- de língua nativa;
- cuja finalidade é servir como fonte de estudo e fonte de língua para a elaboração do material complementar.

.

Segundo Berber Sardinha (2004:26), há quatro pré-requisitos básicos para a formação de um *corpus* computadorizado:

> 1) o *corpus* deve ser composto de textos autênticos, em linguagem natural;
>
> 2) a autenticidade dos textos subentende que tenham sido escritos por falantes nativos;
>
> 3) o conteúdo do *corpus* deve ser escolhido criteriosamente;
>
> 4) e o *corpus* deve ser representativo de uma variedade lingüística ou mesmo de um idioma.

O *corpus* de estudo utilizado nesta pesquisa é composto por cinco livros didáticos, de épocas distintas, elaborados com base em três abordagens diferentes (a audiolingual, a comunicativa e a informada por *corpus*), todos voltados para aprendizes —de inglês como LE— de nível intermediário. A seguir, discorreremos sobre os livros didáticos escolhidos para neste estudo:

- **Pesquisamos dois livros didáticos cuja elaboração baseou-se nos pressupostos teóricos da abordagem audiolingual**:

  - ***English 900****,* publicado em 1961(portanto, mais antigo), pela editora Collier Publishing;

- ***To the Top: Way Ahead***, publicado em 2006 (logo, mais atual), pela Waldyr Lima Editora. É utilizado por uma escola de idiomas, de grande porte, que está no mercado há cerca de 50 anos e tem, aproximadamente, 250 mil alunos;

- **Pesquisamos dois livros didáticos cuja elaboração baseou-se nos pressupostos teóricos da abordagem comunicativa:**

  - ***Interchange: English for International Communication***, publicado em 1991 (mais antigo), pela editora Cambridge University Press;
  - ***Inside Out****,* publicado em 2004 (mais atual), pela editora Macmillan;

- **Pesquisamos um livro didático cuja elaboração baseou-se nos pressupostos teóricos da abordagem informada por *corpus*:**

  - ***Touchstone***, publicado em 2006, pela editora Cambridge University Press.

Para a compilação de um *corpus*, todavia, precisamos atentar para alguns critérios:

a) origem: os dados devem ser autênticos;

b) propósito: a seleção do *corpus* deve ter por finalidade utilizá-lo como objeto de estudo lingüístico;

c) composição: o conteúdo deve ser criteriosamente escolhido pelo pesquisador;

d) formatação: os dados devem ser legíveis para os programas específicos de computador;

e) representatividade: o *corpus* deve ser representativo de uma língua ou variedade lingüística;

f) extensão: o *corpus* deve ser vasto para ser representativo.

[Pearson (2002:42) *apud* Berber Sardinha (2004:16)]

Após a descrição dos *corpora* de estudo, apresentaremos os *corpora* de referência da pesquisa.

2.2.2 *Corpora* de referência

O objetivo principal da utilização de um *corpus* de referência foi o de validar os pacotes lexicais encontrados nos *corpora* de estudo, apesar de ter sido demonstrada a sua semelhança com a língua natural, conforme a descrição feita na seção anterior.

Como explicitado acima, para os fins desta pesquisa, optou-se pela utilização do *British National Corpus (BNC)* como corpus de referência da língua inglesa em uso. O *corpus* britânico supramencionado, lançado em 1995, foi o primeiro a atingir o marco de 100 milhões de palavras, das quais 90% são de inglês escrito, e 10%, de inglês oral. Por meio do quadro a seguir, descrevemos a disposição do *BNC* quanto ao seu modo escrito:

Tipo escrito	**Número de textos**	**Número de palavras**	**% de *corpus* escrito**
Imaginativos	625	19.666.309	22%
Ciências naturais	144	3.752.659	4%
Ciências aplicadas	364	7.369.290	8%
Assuntos internacionais	453	16.507.399	18%
Comércio	284	7.118.321	8%
Artes	259	7.253.846	8%
Lazer	374	9.990.080	11%
Crenças e pensamentos	146	3.053.672	3%
Não classificado	50	1.740.527	2%

TOTAL	3.209	89.740.544	99%[33]

Quadro 1: *BNC*: Descritivo do componente escrito.[34]

Pela observação do quadro 1, percebemos que o *BNC*, contém uma vasta variedade de gêneros e fontes (livros, revistas, jornais, manuscritos). Por isso, o *corpus* é visto como de referência da língua inglesa.

Foi também utilizado como *corpus* de referência o *Google Corpus*: coletado pelo Google Inc., contém palavras e pacotes lexicais de língua inglesa e as suas observadas freqüências. O "cumprimento" dos pacotes lexicais varia —de unigramas (únicas palavras) a qui-gramas (cinco palavras). Os responsáveis pela coleta desse *corpus* tiveram por objetivo a utilização desses dados em pesquisas lingüísticas, por exemplo, para tradução computadorizada ou para reconhecimento de fala, como também para outros usos.

O *corpus* contém, aproximadamente, 1 trilhão de palavras coletadas de textos, de páginas da *internet*, que são publicamente acessíveis.

Os dados foram padronizados respeitando-se a algumas exceções notáveis:

- palavras hifenizadas estão normalmente separadas, e números hifenizados, normalmente, formam um *token*;
- cada seqüência de números separados por barras (por exemplo, as datas) forma um *token*;
- cada seqüência que se parece com URLs ou endereços de *e-mail* forma um *token*.

2.2.2.1 Informações sobre o tamanho do Google Corpus

[33] Meyer (2002) informa, em seu livro, que, por causa do arredondamento das frações, as freqüências acumuladas na tabela não atingem 100%.
[34] Fonte: Meyer (2002:31). (Minha tradução do quadro).

O arquivo que comporta o Google *Corpus* é de, aproximadamente, 24 GB, comprimido (gzip'ed), e formado por arquivos de texto. A seguir, exporemos o quadro 2, por meio do qual apresentamos o número de unidades do *corpus*.

Número de tokens: 1,024,908,267,229
Número de orações: 95,119,665,584
Número de unigramas: 13,588,391
Número de bigramas: 314,843,401
Número de trigramas: 977,069,902
Número de quadrigramas: 1,313,818,354
Número de qui-gramas: 1,176,470,663

Quadro 2: Número de pacotes lexicais extraído do Google *Corpus*.

A escolha desse *corpus* está relacionada ao fato de que o inglês dos textos coletados dos livros didáticos é predominantemente Inglês Americano, então houve a necessidade da utilização de um corpus considerado grande que contém tanto Inglês Americano quanto Britânico para a realização dessa pesquisa.

Vale ressaltar que o Google *Corpus* não é escrito apenas por falantes nativos: nele, encontramos textos, de fonte desconhecida, que consistem em traduções, textos comuns, *e-mails*, etc. e podem ter sido escritos por nativos da língua inglesa ou não. Mas o ponto de corte elimina, de certo modo, vários casos de imprecisão, porque as freqüências mais altas serão as das formas mais estáveis, e estabilidade está relacionada a autenticidade. Berber Sardinha (comunicação pessoal, 2009) diz que estabilidade refere-se às associações léxico-gramaticais mais estabelecidas na língua. Para o autor, formas estáveis tendem a ter freqüências mais altas em um *corpus* tão extenso quanto o Google *Corpus*, enquanto formas emergentes tendem a ter freqüências mais baixas.

Dessa forma, como mencionamos no início desta seção, os *corpora* de referência foram usados para validar os pacotes lexicais encontrados no *corpus* de estudo; procedimento esse que será descrito

na próxima seção.

2.2.3 Critérios de elaboração e coleta

Foram coletados, de cada livro didático, textos como diálogos, reportagens, notícias e entrevistas.

Eles foram digitados no processador de textos *Microsoft Word* e salvos, um a um, em formato de texto *(.txt)*, num total de cinco arquivos. Cada um deles representava um livro didático. Depois de o *corpus* ter sido convertido para o formato .txt, pôde ser submetido ao processamento por meio do programa *WordSmith Tools* (Scott, 1996).

Os livros didáticos continham, em relação uns aos outros, números diferentes de palavras; logo, cada livro teve um número de *tokens* diferente dos demais. O *corpus* de estudo teve um total de 25.485 *tokens*. Segundo Hunston (2000:17), *tokens* são "seqüências de letras separadas por espaços ou pontuação". No quadro abaixo, poderemos demonstrar o número de palavras (*tokens*) de cada livro didático, ou seja, de cada arquivo de texto:

Livros coletados	Número de *tokens*
English 900 **(audiolingualismo)**	**3.338**
Way Ahead: To the Top **(audiolingualismo)**	**4.050**
Interchange **(comunicativismo)**	**6.486**
Inside Out **(comunicativismo)**	**6.489**
Touchstone **(abordagem informada por *corpus*)**	**5.122**

Total de *tokens* do *corpus* de estudo	**25.485**

Quadro3: Número total de *tokens* do *corpus* de estudo.

O *corpus* de estudo, segundo a classificação abaixo (quadro 4), é "pequeno". Para ter representatividade, o *corpus* tem de ser o maior possível. Como o tamanho do nosso *corpus* de estudo é limitado pelo número de livros didáticos pesquisados (apenas cinco), entendemos que o nosso *corpus* de estudo é representativo, apesar de "pequeno".

Tamanho em palavras	**Classificação**
Menos de 80 mil	Pequeno
80 a 250 mil	Pequeno-médio
250 mil a 1 milhão	Médio
1 milhão a 10 milhões	Médio-grande
10 milhões ou mais	Grande

Quadro 4: Classificação do *corpus* de acordo com o seu "tamanho", Berber Sardinha (2004:26).

Tendo exposto a forma pela qual os dados utilizados na pesquisa foram selecionados, passaremos à descrição dos procedimentos de análise deles.

2.3 Procedimentos de análise dos dados

A análise foi feita em consonância com os objetivos da pesquisa. De acordo com eles, primeiro, tivemos de coletar os *corpora*; em segundo lugar, foi feito o levantamento dos pacotes lexicais constantes dos textos existentes em cada um dos livros didáticos por meio do programa *WordSmith Tools* e, finalmente, esses pacotes lexicais foram comparados aos pacotes lexicais do *BNC* e do Google *Corpus*.

Um dos critérios definidos para a análise do *corpus* de estudo foi a extensão dos pacotes lexicais. A extensão de um pacote lexical (*bundle*) tem que ver com o número de termos que o comporão. Scott & Tribble (2006) analisaram os pacotes lexicais na escrita acadêmica em inglês. Fizeram uma análise das 20 palavras de mais ocorrência em

dois *corpora* e notaram que essas ocorrências eram muito similares em ambos. Os autores fizeram a mesma análise para pacotes lexicais de duas, três e quatro palavras (bigramas, trigramas e quadrigramas, respectivamente). Concluíram, então, que o mais indicado seria analisar trigramas (numa análise dos colocados à direita), para estudar o contraste entre os diferentes estilos de escrita ou o produto dos diferentes grupos de escritores, e quadrigramas (que são fortes fatores discriminantes entre diferentes registros).

Com base nesses achados, optamos por trabalhar apenas com trigramas, seguindo a argumentação de Biber (2000) : "(...) os feixes de três palavras podem ser considerados como uma espécie de associação mais extensa de colocados e, dessa forma, são extremamente comuns". (Minha tradução.)[35].

A seguir, conheceremos as ferramentas utilizadas nesta pesquisa.

2. 3.1 Ferramentas de coleta

Dentre os diversos *softwares* disponíveis para auxiliar o lingüista de *corpus*, um deles se destaca: o *WordSmith Tools*.

Utilizamos o programa, primeiro, por ser um pacote de ferramentas, utilitários e acessórios muito úteis; em segundo lugar, por ter sido idealizado especificamente para trabalhar-se com *corpus*, segundo a concepção terminológica da LC; e, também, por ser um programa bastante difundido, o que facilitaria a reprodução de um estudo por outros pesquisadores, que contariam com os mesmos recursos, sem precisarem lançar mão de muitos programas diferentes para terem o conjunto de ferramentas nele contidas à sua disposição.

O programa, desenvolvido por Mike Scott, da Universidade de Liverpool(Reino Unido), contém, entre outras, três ferramentas: *WordList*, *KeyWords* e *Concord*. Todavia, nesta pesquisa, utilizamos

[35] Three-word bundles can be considered as a kind of extender collocational association, and are, thus, extremely common.

apenas a ferramenta *WordList,* ou lista de multipalavras.

Segundo Berber Sardinha (1999), o *WordSmith Tools* é um programa de análise lexical por computador que faz análise lingüística por meio de um *corpus*. O programa, na sua versão 3.0, foi empregado nesta análise lingüística por ser um combinado de ferramentas eficientes e apropriadas para este tipo de estudo. Segundo (Berber Sardinha, 2004:86)

> O programa coloca à disposição do analista uma série de recursos que, bem usados, são extremamente úteis e poderosos na análise de vários aspectos da linguagem, como a composição lexical, a temática de textos selecionados e a organização retórica e composicional de gêneros discursivos.

No quadro 5, reproduzido e adaptado por Berber Sardinha (1999:89), há uma lista das as ferramentas, dos utilitários e dos instrumentos disponíveis no *WordSmith Tools*, bem como suas funções.

Ferramentas, utilitários, instrumentos e funções disponíveis no *WordSmith Tools*

Componentes: O *WordSmith Tools* é composto por (a) ferramentas, (b) utilitários, (c) instrumentos, e (d) funções.

Há três **ferramentas** e quatro **utilitários**, nomeadamente:

(a) **Ferramentas**:

WordList; KeyWords; Concord

(b) **Utilitários**:

Renamer; *Text Converter; Splitter; Viewer*

(c) **Instrumentos** de análise disponíveis (com os nomes em inglês entre parênteses):

WordList:

1. Lista de palavras individuais (‘*wordlist*’).

2. Lista de multipalavras (‘*wordlist, clusters activated*’).

3. Lista de palavras individuais de consistência (‘*detailed consistency*’).

4. Lista de multipalavras de consistência (‘*detailed consistency, clusters activated*’).

5. Lista de dimensões e densidade lexical (‘*statistics*’).

Concord:

1. Concordância (‘*concordance*’).

2. Lista de colocados (‘*collocates*’).

3. Lista de agrupamentos lexicais (‘*clusters*’).

4. Lista de padrões de colocados (‘*patterns*’).

5. Gráfico de distribuição da palavra de busca (‘*plot*’).

KeyWords:

1. Lista de palavras-chave (‘*keywords*’).

2. Banco de dados de listas de palavras-chave (‘*database*’).

3. Lista de palavras-chave chave (‘*key keywords*’).

4. Lista de palavras-chave associadas (‘*associates*’).

5. Lista de agrupamentos textuais (‘*clumps*’).

6. Gráfico de distribuição de palavras-chave (‘*keyword plot*’).

7. Listagem de elos entre palavras-chave (‘*keyword plot links*’).

As principais **funções** (d) distribuídas nas três ferramentas são:

1. Lematização: agrupamento de duas ou mais formas diferentes em mesmo item.
2. Classificação: ordenação de listas e concordâncias.
3. Delimitação: escolhas de quais partes do *corpus* serão lidas pelo programa.

Quadro 5: Ferramentas, utilitários, instrumentos e funções disponíveis no *WordSmith Tools*: reproduzido de Berber-Sardinha (1999:89), com adaptações e grifos nossos.

Utilizamos o programa *WordSmith Tools* para a obtenção dos pacotes *lexicais* (trigramas). Posteriormente, como precisávamos obter a lista dos trigramas exclusivos do *corpus* de estudo, e o *WordSmith* não nos disponibilizava esse recurso, desenvolvemos rotinas escritas, na linguagem de programação Shell, que foram rodadas no programa

Cygwin (que pode ser usado em ambiente Windows), como se estivéssemos utilizando o sistema operacional Unix. Dessa forma, a linguagem Shell possibilitou a criação de um *script* , que pode ser conferido no anexo 1.

Abaixo, descrevemos os passos adotados para a comparação dos pacotes lexicais encontrados no *corpus* de estudo.

2.3.3 Análise dos dados

Nesta subseção, serão detalhados os procedimentos utilizados na análise dos pacotes lexicais dos livros didáticos. É importante ressaltarmos que todos os procedimentos adotados para a análise foram aplicados, igualitariamente, para todos os livros didáticos pesquisados.

As palavras ‘convergentes’ e ‘divergentes’, nesta pesquisa, são apenas maneiras de operacionalizar a investigação de autenticidade do texto no âmbito da léxico-gramática.

Destacamos também que cada pesquisa desenvolvida a partir de uma abordagem baseada em *corpus* é única. Sendo assim, os procedimentos adotados em cada uma delas são igualmente únicos: norteados pelos dados e pelas necessidades impostas pela pesquisa no decorrer da análise. Desse modo, dividiremos a presente subseção em passos, para facilitarmos a visualização do que foi feito. Observemos os passos dos procedimentos adotados na análise.

Primeiro passo: fizemos, primeiro, uma lista dos pacotes lexicais de cada livro didático. Essa lista foi feita por meio do programa *WordSmith Tools* (versão 3.0), mediante o uso da ferramenta *WordList.*

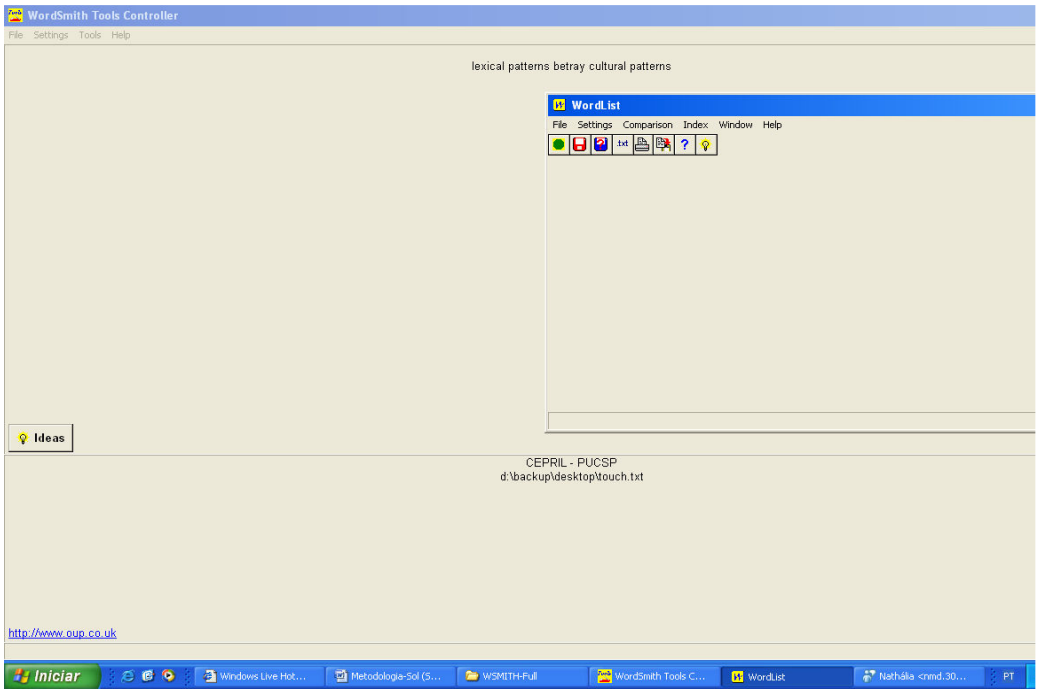


Figura 4: Amostra do programa *WordSmith Tools*.

Após a utilização do programa, os pacotes lexicais foram salvos, um a um, em formato de texto *(.txt)*, num total de cinco arquivos. Cada desses arquivos representava um livro didático. Abaixo, exibimos um exemplo desse arquivo:

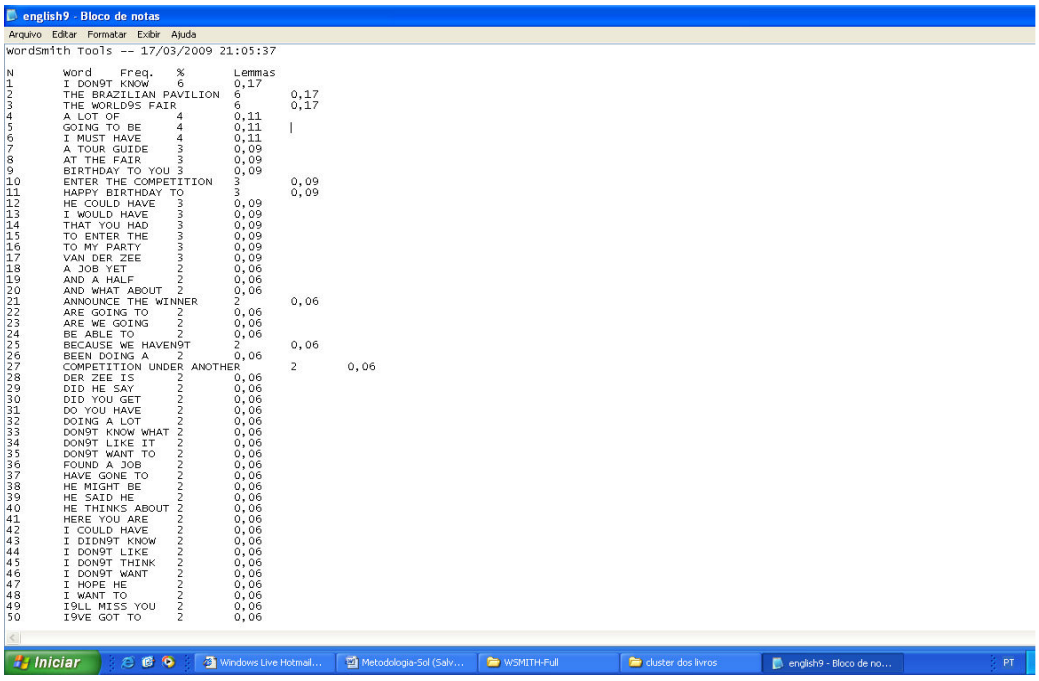


Figura 5: Pacotes lexicais do livro *English 900*, de 1961.

Segundo passo: fizemos uma comparação dos pacotes lexicais de cada livro didático com os dos *corpora* de referência. Essa

comparação foi feita por meio de um *script* desenvolvido pelo orientador deste estudo, o qual pode ser conferido no anexo 1. A figura abaixo ilustra a convergência e/ou a divergência após essa comparação:

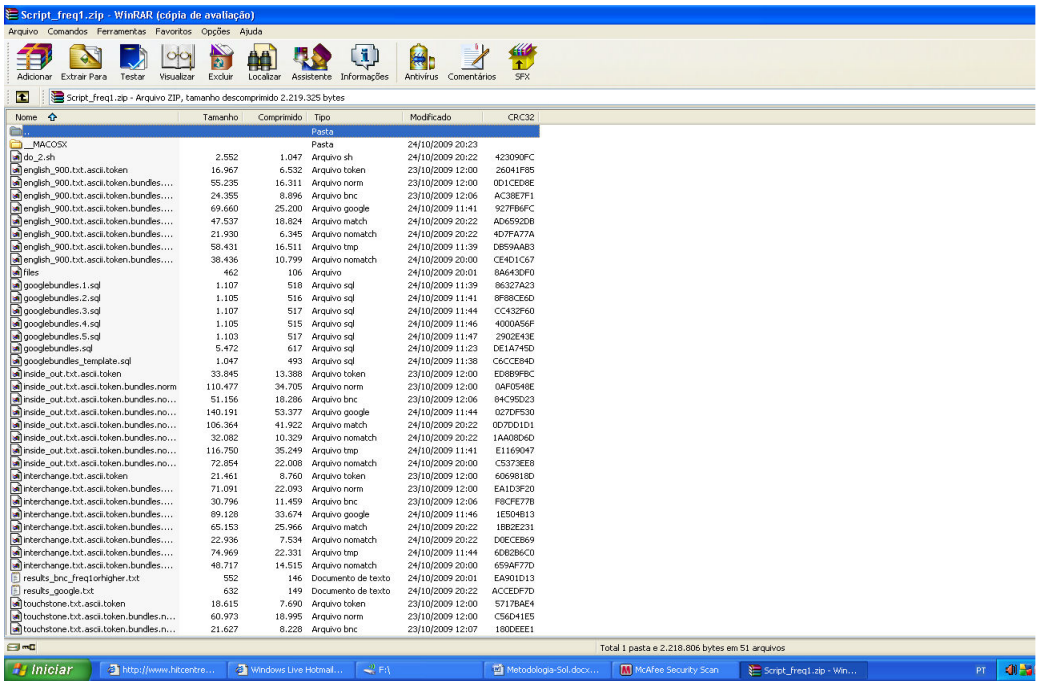


Figura 6: Lista dos pacotes lexicais convergentes e divergentes de cada *corpus* de referência.

Terceiro passo: após essa comparação, determinaram-se quais eram as semelhanças e diferenças dos pacotes lexicais em ambos os *corpora*. Procuramos, desse modo, analisar os livros didáticos de acordo com o seu grau de convergência e divergência, classificando-os como livros cujo nível de autenticidade era maior ou menor.

Nesse passo, dividimos a análise em duas partes: a análise quantitativa e a análise qualitativa.

- **Análise quantitativa:** analisamos os livros e textos por meio de porcentagens, tabelas, gráficos, testes e *scripts*. Essa análise foi dividida em três etapas:

a) *Análise quantitativa inicial:* nessa etapa, buscou-se demostrar o grau de convergência e divergência entre os *corpora* Google e *BNC* (*corpora* de referência) e os livros didáticos (*corpora* de

estudo), em número de pacotes lexicais e em porcentagem.

Para isso, tivemos de contar com recursos de computador, como tabelas feitas no programa Excel e, em uma etapa posterior, em gráficos (para melhor ilustrar os resultados obtidos).

Analisamos também a proporção de pacotes lexicais convergentes. O cálculo para obtermos esse tipo de padrão foi feito da seguinte maneira:

(total de pacotes lexicais / total de pacotes lexicais divergentes)

O resultado dessa divisão nos permitiu entender quantos pacotes lexicais de cada livro didático eram divergentes.

Além do resultado obtido, submetemos as freqüências dos pacotes lexicais convergentes e divergentes ao teste estatístico qui-quadrado, a fim de sabermos se havia diferença expressiva entre as freqüências. O cálculo foi feito por meio do *site* http://faculty.vassar.edu/lowry/newcs.html, como se pode visualizar abaixo:

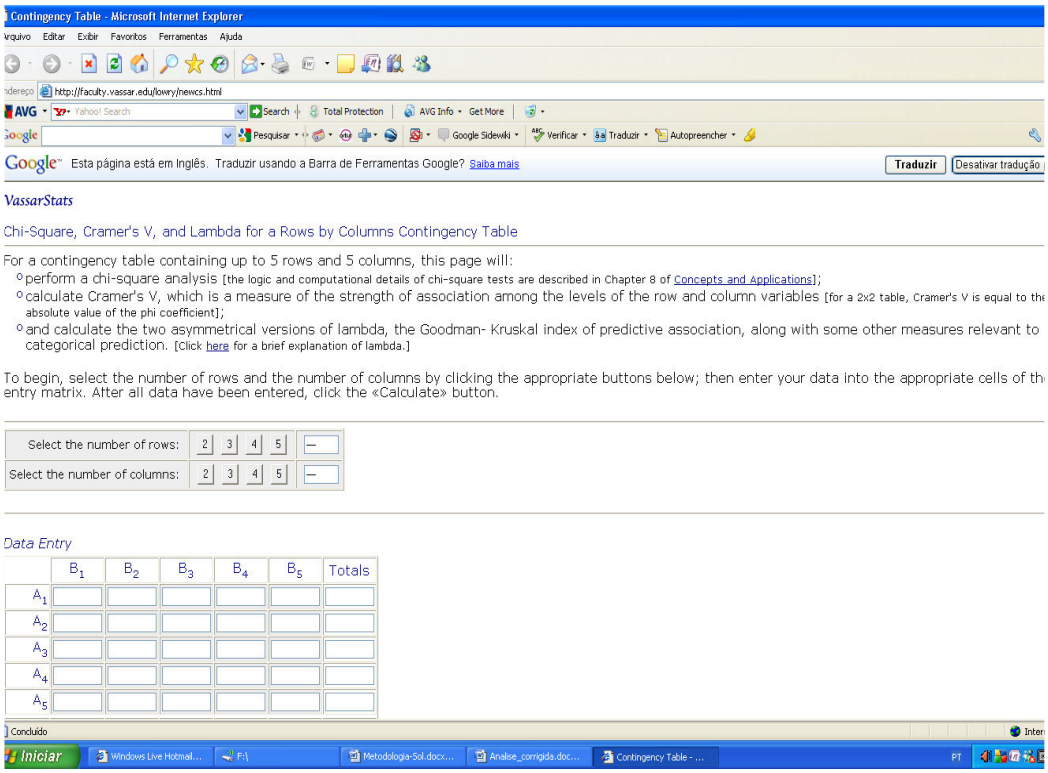


Figura 7: Uma amostra do teste qui-quadrado.

Por fim, fizemos um gráfico a partir somente do grau de convergência dos *corpora* de estudo e de referência (*BNC* e Google). Conseguimos, então, visualizar, de forma bem clara, a diferença do grau de convergência entre os *corpora*.

b) Análise quantitativa recontagem: julgamos necessária essa análise porque, ao verificarmos os pacotes lexicais divergentes, encontramos algumas características específicas dos *corpora* de estudo, como nomes próprios, numerais e pontuação, que acabaram influenciando o resultado do grau de autenticidade.

Fizemos, então, uma recontagem. Para que ela fosse adequada e bem--sucedida, determinamos o tamanho da amostra em 100. Assim, para cada livro escolhido, selecionamos 100 instâncias de pacotes lexicais divergentes e procedemos ao exame manual de cada um deles, retirando as características específicas. Depois, recontamos o número de pacotes lexicais divergentes e adicionamos ao número de pacotes lexicais convergentes.

Fizemos, ainda, uma tabela ao final de cada livro didático, com os números recontados e a nova porcentagem correspondente.

Em seguida, demonstramos os pacotes convergentes em gráficos (um com os dados do *BNC* e outro com os dados do Google *Corpus*).

Por fim, os novos resultados foram submetidos ao teste estatístico qui-quadrado (já comentado).

Terminada a explicação dos procedimentos adotados, passaremos para a terceira etapa da nossa análise.

c) Análise quantitativa de variação de convergência texto a texto: também julgamos necessária essa análise, já que, anteriormente, havíamos feito a análise dos pacotes lexicais convergentes e divergentes em todos os livros didáticos, mas não em todos os textos.

Essa análise também foi feita por meio de um *script* desenvolvido pelo professor-orientador. Esse *script* demostrou a porcentagem de convergência existente em cada texto dos livros didáticos. O resultado desse *script* pode ser conferido no anexo 2, do qual consta a porcentagem das faixas de convergências de cada texto do *corpus* de estudo. A partir disso, as faixas foram classificadas em 'alta', 'média', 'baixa' e 'muito baixa', de acordo com o seu grau de convergência. Após essa etapa, somamos a porcentagem das faixas 'altas' e 'média' e 'baixa' e 'muito baixa'. Com tais resultados, conseguimos obter um *ranking* dos livros mais e menos convergentes. Após a classificação, resolvemos apontar tanto o texto mais convergente quanto o menos convergente no nosso *corpus* de estudo.

- **Análise qualitativa:** interpretamos os dados, os resultados e os pacotes lexicais convergentes e divergentes. Essa etapa foi dividida em duas partes.

a) *Análise qualitativa dos pacotes lexicais divergentes*: nessa análise, observaram-se as características específicas dos pacotes lexicais divergentes. Notamos 3 (três) características que aumentavam o número de pacotes lexicais divergentes: essas características específicas não são, na verdade, marcadores de falta de autenticidade; são pacotes lexicais autênticos, que, no entanto, contêm nomes próprios, numerais e pontuações. Por isso, resolvemos exemplificar cada característica específica com 5 (cinco) pacotes lexicais e as

sentenças encontradas nos cinco livros didáticos; ou seja, cada exemplo foi retirado de um livro didático diferente. Desse modo, verificamos que aqueles pacotes lexicais que eram considerados divergentes, na verdade, eram convergentes, mas continham uma daquelas características especificas. Discorreremos, agora, sobre a segunda etapa desta pesquisa.

b) *Análise qualitativa dos pacotes lexicais convergentes:* já que havíamos analisado as características dos pacotes divergentes, resolvemos analisar melhor os pacotes lexicais convergentes. Para isso, selecionamos os 20 (vinte) primeiros pacotes lexicais convergentes de todos os livros didáticos e compará-los ao *corpus* de referência Google, já que ele apresenta um grau de convergência maior do que o do *BNC*.

Elaboramos, para essa análise, uma tabela com os 20 (vinte) pacotes lexicais dos livros didáticos, separados por suas abordagens. As interpretações baseadas nelas e sua classificação tiveram como base a *Longman Grammar of Spoken and Written English*, de Biber *et al* (1999). Primeiro, fizemos a análise dos livros elaborados com base nos pressupostos teóricos da abordagem audiolingual (*English 900* e *Way Ahead: To the Top*); em segundo lugar, a dos livros elaborados com base nos pressupostos teóricos da abordagem comunicativa (*Interchange* e *Inside Out*) e, por fim, a do livro elaborado com base nos pressupostos teóricos da abordagem informada por *corpus* (*Touchstone*). Após a conclusão dessas tabelas, classificamos os 20 primeiros *bundles* de cada livro didático de acordo com Biber (1999).

Finalmente, após as análises descritas acima, para concluir a metodologia, fizemos um '*ranking esperado'*, em ordem crescente, com os livros didáticos que considerávamos mais autênticos, assim como outros três *'rankings obtidos'* (um com o *BNC*, um com o *Google* e o último com a análise quantitativa de variação de convergência texto a texto), também em ordem crescente e com os livros didáticos cujo grau de autenticidade distribuído era maior. Fizemos um gráfico com todos os *rankings* obtidos por meio da pesquisa.

Após a apresentação da metodologia empregada no estudo, apresentaremos, no próximo capítulo, a interpretação dos dados e os resultados.

Capítulo 3

Análise dos Dados

Neste capítulo, apresentamos a análise dos dados advindos dos *corpora* selecionados para a pesquisa. A partir da sua coleta, os dados foram analisados e interpretados de forma quantitativa e qualitativa, a fim de respondermos às perguntas de pesquisa (explicitadas no capítulo anterior). Para responder a todas elas, dividimos a nossa análise em duas principais modalidades: quantitativa e qualitativa.

Os procedimentos de coleta e de análise foram apresentados no capítulo de metodologia.

3.1 Análise quantitativa dos *corpora*

É importante ressaltar que, por meio desta análise, buscamos representar, de forma quantitativa, os dados e resultados obtidos mediante o uso do programa computacional *WordSmith Tools* e de rotinas escritas na linguagem de programação Shell, todos esses passos também descritos na metodologia.

Para melhor interpretarmos os dados obtidos, esta análise quantitativa foi dividida em três subpartes:

- Análise quantitativa inicial;
- Análise quantitativa recontagem;
- Análise quantitativa de variação de convergência texto a texto.

Em cada uma dessas análises, tentamos dar uma diferente interpretação aos resultados obtidos por meio dos pacotes lexicais convergentes e divergentes a partir dos *corpora* pesquisados.

3.1.1 Análise quantitativa inicial

Para esta análise, tivemos de contar com recursos de computador, como tabelas feitas no programa Excel e, em uma etapa posterior, gráficos para melhor ilustrar os resultados obtidos.

A seguinte tabela contém os resultados do grau de convergência entre os *corpora* Google (*corpus* de referência) e os livros didáticos (*corpus* de estudo). Lembramos que o termo "convergência" refere-se, nesta pesquisa, à quantidade de pacotes lexicais em comum entre o *corpus* de estudo e o de referência. Segundo a metodologia, quanto mais elevado for o grau de convergência, mais probabilidade haverá de o grau de autenticidade da língua contida nos livros didáticos seja elevado. Essa autenticidade é em relação à língua em uso real —representada no *corpus* de referência— no que tange à presença de pacotes lexicais.

Assim, observaremos, em porcentagem, o grau de autenticidade existente nos textos e nos diálogos contidos nos livros didáticos, ou seja, lembrando-nos de que quanto mais elevado for o grau de convergência, mais elevado será o grau de autenticidade contida nos livros didáticos; e quanto mais elevado for o grau de divergência, mais baixo será o grau de autenticidade da língua contida nos livros didáticos.

Livros didáticos	***Total de pacotes lexicais***	***Pacotes lexicais convergentes***	***%***	***Pacotes lexicais divergentes***	***%***	***Proporção de pacotes lexicais convergentes***
English 900 **(audiolingualismo)**	3185	**2132**	66.9 %	**1053**	33.1%	**3.0**
To the Top **(audiolingualismo)**	3920	**2723**	69.4 %	**1197**	31.6%	**3.2**
Inside Out **(comunicativismo)**	6168	**4712**	76.3 %	**1456**	23.7%	**4.2**
Interchange **(comunicativismo)**	3820	**2817**	73.7 %	**1003**	26.3%	**3.8**
Touchstone **(informada por *corpus*)**	3331	**2370**	71.1 %	**961**	28.9%	**3.4**

Tabela 2: Porcentagem de convergência e divergência entre os pacotes lexicais dos livros didáticos e os do Google *Corpus*.

Na tabela acima, procuramos analisar também a proporção de pacotes lexicais convergentes, calculada assim: total de pacotes lexicais dividido pelo total de pacotes lexicais divergentes. Por exemplo, o livro *English 900*, da abordagem audiolingual: 3185/1053 = 3.02 = 3. Com esses dados, entendemos que, em cada grupo de 3 pacotes lexicais do livro *English 900*, 1 é divergente.

As freqüências das ocorrências nos pacotes lexicais convergentes e divergentes foram submetidas ao teste estatístico qui-quadrado, e o valor de qui-quadrado foi 119.05, que corresponde a p<.0001. Isso indica que há diferença expressiva nas quantidades de pacotes lexicais convergentes e divergentes entre os livros didáticos.

Abaixo, representaremos esses dados, por meio de gráficos, para melhor ilustrar os resultados obtidos acima.

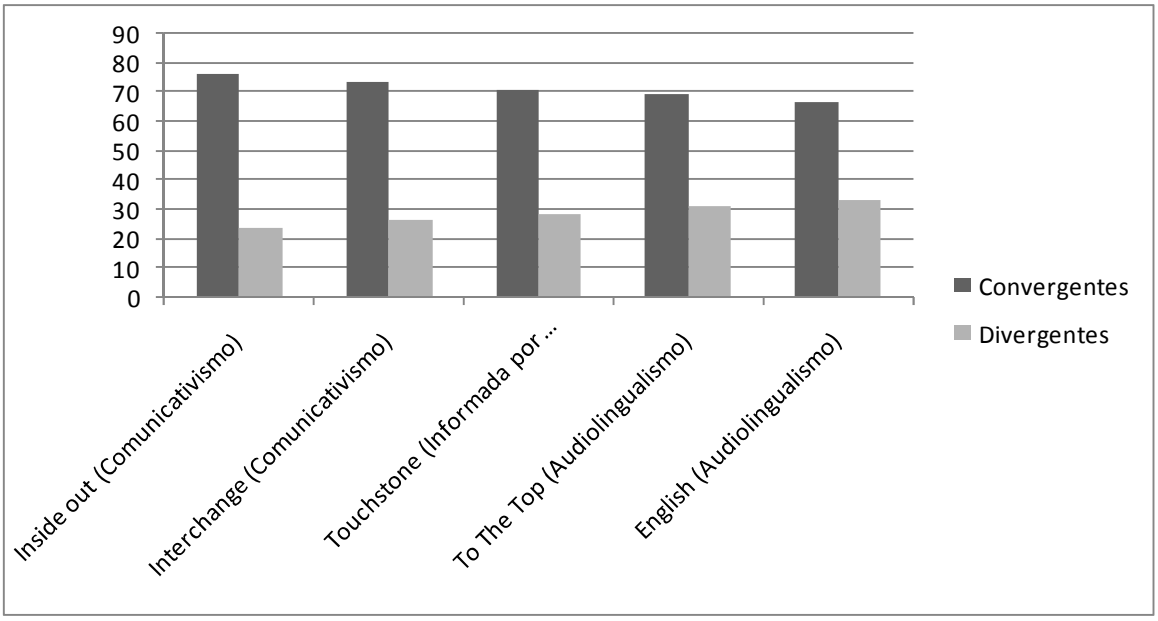


Figura 8: Gráfico representativo dos pacotes lexicais dos livros didáticos comparados aos do Google *Corpus*.

A depreensão mais imediata que a observância dos dados nos permite é a de que o grau de convergência é alto, em todos os livros. Tínhamos a expectativa de que as quantidades seriam, primeiro, muito diferentes entre os livros cujo preparo, em tese, tivesse sido pautado por mais comprometimento com a autenticidade lingüística (*Inside Out*, *Interchange* e *Touchstone*) e aqueles cujo preparo, também em tese, tivesse sido pautado por menos comprometimento com o emprego da língua real (*English 900* e *To the Top*). As diferenças, porém, foram mínimas entre eles.

Em segundo lugar, mas ainda relacionadas à primeira expectativa, esperávamos que as quantidades de pacotes convergentes fossem baixas para os livros do segundo grupo (*English 900* e *To the Top*). A quebra de ambas as expectativas ilustrou, mais uma vez, como a Lingüística de *Corpus* traz-nos dados que contradizem a nossa intuição. Além disso, mostra que as diferenças entre os livros, do ponto de vista quantitativo, são sutis e precisam ser exploradas qualitativamente, ou seja, talvez a principal diferença entre os livros não esteja tanto na quantidade de pacotes convergentes, mas na possível predominância de um ou outro tipo de pacote.

Diante desses dados, notamos também que, quanto aos livros baseados na abordagem comunicativa (*Interchange* e *Inside Out*) e ao baseado na abordagem informada por *corpus* (*Touchstone*), há um aumento sutil no grau de autenticidade —em relação aos livros baseados na abordagem audiolingual (*English 900* e *To the Top*).

Notamos, assim, que os livros que contêm textos menos comprometidos com a autenticidade, ou seja, cuja língua deles constante é julgada mais inventada, acabam apresentando menos convergência com o *corpus* de referência (representativo da língua autêntica, ou em uso pelos falantes nativos), enquanto os livros que são mais comprometidos com a autenticidade demonstram mais convergência.

A partir deste momento, analisaremos, por meio de tabela, a porcentagem dos pacotes lexicais constantes dos livros didáticos, em comparação aos do BNC, que é um dos *corpora* de referência na pesquisa.

Livros didáticos	***Total de pacotes lexicais***	***Pacotes lexicais convergentes***	***%***	***Pacotes lexicais divergentes***	***%***	***Proporção de pacotes lexicais convergentes***
English 900 **(audiolingualismo)**	3188	**1299**	40.7%	**1889**	59.3%	1.6
To the Top **(audiolingualismo)**	3922	**1381**	35.2%	**2541**	64.8%	1.5
Inside Out **(comunicativismo)**	6186	**2700**	43.6%	**3486**	56.4%	1.7
Interchange	3832	**1601**	41.7%	**2231**	59.3%	

(comunicativismo)						1.7
***Touchstone* (informada por *corpus*)**	3345	**1158**	34.6%	**2187**	65.4%	1.5

Tabela 3: Porcentagem dos pacotes lexicais constantes dos livros didáticos em relação aos do *British National Corpus*.

Com base na tabela acima, pudemos ainda analisar a proporção de pacotes lexicais convergentes, calculada da seguinte maneira: total de pacotes lexicais dividido pelo total de pacotes lexicais divergentes. Por exemplo, o livro *English 900*, da abordagem audiolingual: 3188/1889 = 1,6, ou seja, com esses dados, entendemos que 1 pacote lexical, em cada 1,6 do livro *English 900*, é divergente.

As freqüências das ocorrências nos pacotes lexicais convergentes e divergentes foram submetidas ao teste estatístico qui-quadrado, e o valor de qui--quadrado foi 117.64, que corresponde a p<.0001. Isso indica que há diferença expressiva nas quantidades de pacotes lexicais convergentes e divergentes entre os livros didáticos.

Abaixo, representaremos esses dados por meio de gráficos para melhor ilustrar os resultados obtidos acima.

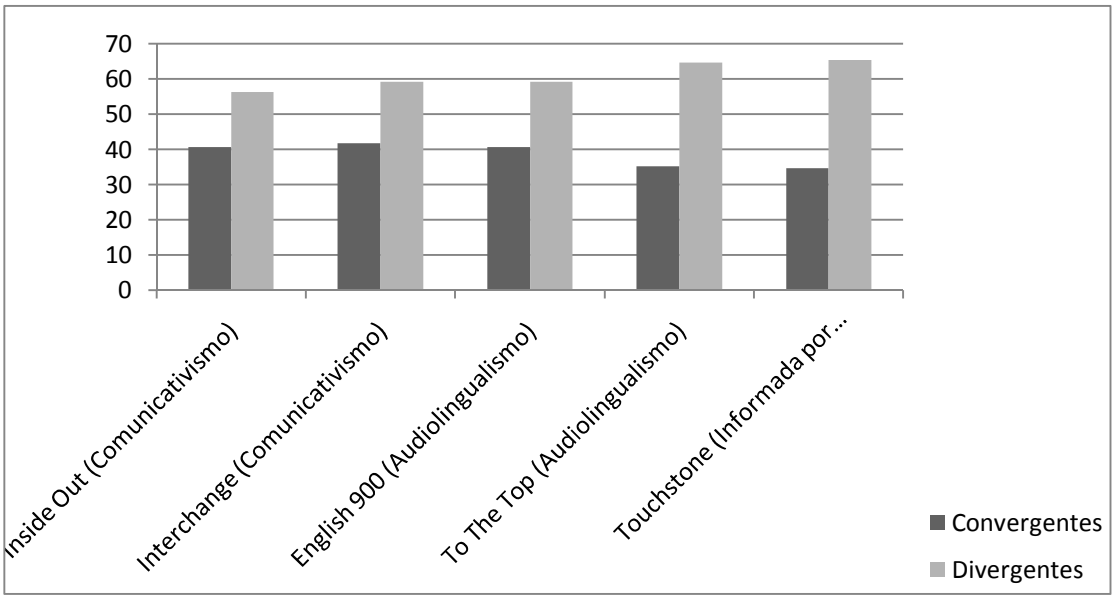


Figura 9: Gráfico representativo dos pacotes lexicais constantes dos livros didáticos comparados aos do *BNC Corpus*.

Diante da tabela e do gráfico acima, notamos que o grau de convergência dos livros didáticos, em comparação ao *BNC*, é baixo; ou seja, o número de pacotes lexicais divergentes é superior ao número dos convergentes. Do mesmo

modo, notamos que o grau de convergência entre os livros didáticos é bem "igualado". Diante desse fato, percebemos que os livros da abordagem comunicativa, *Inside Out* e *Interchange*, e o da abordagem audiolingual, *English 900*, ao serem comparados ao *BNC*, têm grau de autenticidade superior ao dos demais livros.

Com base nos dados acima, podemos observar que o resultado do grau de autenticidade dos textos e dos diálogos contidos nos livros didáticos pode sofrer alteração, de acordo com o *corpus* de referência comparado, como podemos notar no gráfico abaixo, em que apresentaremos o grau de convergência de todos os livros didáticos.

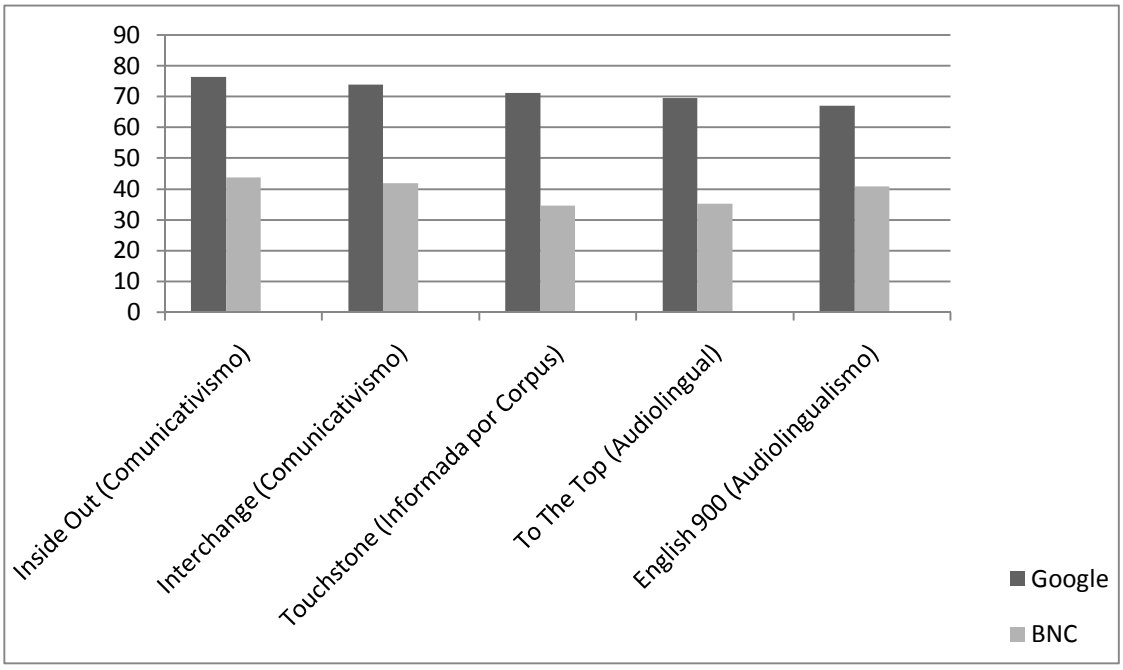


Figura10: Gráfico do grau de convergência entre o *corpus* de estudo e o *corpus* de referência.

Como podemos notar, mesmo que o grau de convergência do Google *Corpus* seja muito superior ao grau de convergência do *BNC*, há concordância entre os dois *corpora* (*BNC* e Google), e ainda que a sua proporção seja diferente, essa discrepância será justificada logo abaixo, na subseção intitulada *análise qualitativa dos corpora*, na qual explicaremos as características em que diferem o *BNC* e o Google *Corpus*.

3.1.2 Análise qualitativa recontagem

Para conferir os resultados apresentados no parágrafo anterior, resolvemos recontar os pacotes lexicais divergentes manualmente. Como o número de pacotes lexicais divergentes é muito alto, decidimos fazer essa recontagem por meio de amostragem.

Dessa forma, determinamos o tamanho da amostra em 100; ou seja, para cada livro, selecionamos 100 instâncias de pacotes lexicais divergentes e procedemos ao exame manual de cada um deles. Assim, cada pacote lexical foi julgado, e analisamos se ele era um caso de características específicas, ou melhor, se era uma instância de qualquer um dos seguintes pacotes lexicais divergentes (como nome próprio, numeral e pontuação). Essas características específicas serão mais bem explicadas na análise qualitativa. Começamos a recontagem com os livros da abordagem audiolingual.

a) Livro *English 900* comparado ao *BNC*

Dos 100 pacotes lexicais divergentes selecionados, 30 pacotes lexicais têm as características específicas, e 70 não as têm. Logo, 70% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 1889 x 70%= 1322, percebemos que havia 1322 pacotes lexicais divergentes verdadeiros. A diferença apresentada foi de 567 pacotes lexicais, que, por sua vez, foi acrescida entre os pacotes lexicais convergentes. Desse modo, o número de pacotes lexicais convergentes, que, antes, era de 1299 (originais) + 567 (corrigidos) = 1866, resultou em 1866 pacotes lexicais convergentes.

Livro	**Pacotes lexicais convergentes ANTES**	**Pacotes lexicais convergentes DEPOIS**	**Pacotes lexicais divergentes ANTES**	**Pacotes lexicais divergentes DEPOIS**
English 900 (audiolingualismo)	**1299**	**1866**	**1889**	**1322**
Porcentagem	**40.7%**	**58.5%**	**59.3%**	**42.5%**

Tabela 4: Recontagem dos pacotes lexicais convergentes e divergentes do livro *English 900* comparado ao *BNC*.

Como podemos observar, com essa recontagem, o número de pacotes lexicais convergentes aumentou de 40,7% para 58%, aumentando, assim, o grau de convergência, ou seja, o grau de autenticidade do livro *English 900* quando comparado ao *BNC*.

b) Livro *English 900* comparado ao Google *Corpus*

Dos 100 pacotes lexicais divergentes selecionados, 43 pacotes lexicais contêm as características específicas, e 57 não as contêm; ou seja, 57% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração o fato de que a contagem inicial total de pacotes lexicais divergentes era de 1053 pacotes lexicais x 57%= 600, tivemos, como resultado, 600 pacotes lexicais divergentes verdadeiros. A diferença (453 pacotes lexicais) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, que, antes, era de 2132 (originais) + 453 (corrigidos) = 2585, resultou em 2585 pacotes lexicais convergentes.

Livro	Pacotes lexicais convergentes ANTES	Pacotes lexicais convergentes DEPOIS	Pacotes lexicais divergentes ANTES	Pacotes lexicais divergentes DEPOIS
English 900 (audiolingualismo)	**2132**	**2585**	**1053**	**600**
Porcentagem	**66.9%**	**81.1%**	**33.1%**	**19.9%**

Tabela 5: Recontagem dos pacotes lexicais convergentes e divergentes do livro *English 900* comparado ao Google *Corpus*.

Como podemos notar, com essa recontagem, o número de pacotes lexicais convergentes aumentou de 66.9% para 81.1%, aumentando também o grau de convergência, ou seja, o grau de autenticidade do livro *English 900* quando comparado ao Google *Corpus*.

c) Livro *To the Top* comparado ao *BNC*

Dos 100 pacotes lexicais divergentes selecionados, 17 pacotes lexicais têm as características específicas, e 83 não as têm; ou seja, 83% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 2541 pacotes lexicais x 83%, entendemos que o resultado seja de 2109 pacotes lexicais divergentes verdadeiros. A diferença (432 pacotes lexicais) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, que, antes, era de 1381 (originais) + 453 (corrigidos) = 1813, agora, soma 1813 pacotes lexicais convergentes.

Livro	**Pacotes lexicais convergentes ANTES**	**Pacotes lexicais convergentes DEPOIS**	**Pacotes lexicais divergentes ANTES**	**Pacotes lexicais divergentes DEPOIS**
To the Top (audiolingualismo)	**1381**	**1813**	**2541**	**2109**
Porcentagem	**35.2%**	**46.2%**	**64.8%**	**53.8%**

Tabela 6: Recontagem dos pacotes lexicais convergentes e divergentes do livro *To the Top* comparado ao *BNC*.

Notemos que, por meio dessa recontagem, o número de pacotes lexicais convergentes aumentou de 35.2% para 46.2%, aumentando, assim, o grau de convergência, ou seja, o grau de autenticidade do livro *To the Top*, se o compararmos ao *BNC*.

d) Livro *To the Top* comparado ao Google *Corpus*

Dos 100 pacotes lexicais divergentes selecionados, 21 pacotes lexicais contêm as características específicas, e 79 não as contêm; ou seja, 79% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 1197 pacotes lexicais x 79%, obtivemos este resultado: 945 pacotes lexicais divergentes são verdadeiros. A diferença (252 pacotes lexicais) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, que, antes, era de 2723 (originais) + 252 (corrigidos), resultou em 2975 pacotes lexicais convergentes.

Livro	**Pacotes lexicais convergentes ANTES**	**Pacotes lexicais convergentes DEPOIS**	**Pacotes lexicais divergentes ANTES**	**Pacotes lexicais divergentes DEPOIS**
To the Top (audiolingualismo)	**2723**	**2975**	**1197**	**945**
Porcentagem	69.4%	75.8%	30.6%	24.2%

Tabela 7: Recontagem dos pacotes lexicais convergentes e divergentes do livro *To the Top* comparado ao Google.

Como podemos observar, mediante essa recontagem, o número de pacotes lexicais convergentes aumentou de 69.4% para 75.8%, aumentando, desse modo, o grau de convergência, ou de autenticidade do livro *To the Top* quando comparado ao Google *Corpus*. Agora, analisaremos os livros da abordagem comunicativa.

e) Livro *Inside Out* comparado ao *BNC*

Dos 100 pacotes lexicais divergentes selecionados, 25 pacotes lexicais têm as características específicas, e 75 não as têm; ou seja, 75% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 3486 pacotes lexicais x 75%, obtivemos o resultado de que 2614 pacotes lexicais divergentes são verdadeiros. A diferença (872 pacotes lexicais) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes antes era de 2700 (originais) + 872 (corrigidos), de modo a resultar em 3572 pacotes lexicais convergentes.

Livro	Pacotes lexicais convergentes e antes	Pacotes lexicais convergentes e depois	Pacotes lexicais divergentes e antes	Pacotes lexicais divergentes e depois
Inside Out (comunicativismo)	**2700**	**3572**	**3486**	**2614**
Porcentagem	43.6%	57.7%	66.4%	42.3%

Tabela 8: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Inside Out* comparado ao *BNC*.

Como podemos constatar, a partir dessa recontagem, o número de pacotes lexicais convergentes aumentou de 43.6% para 57.7%, aumentando o grau de convergência, ou de autenticidade, do livro *Inside Out* quando comparado ao *BNC*.

f) Livro *Inside Out* comparado ao Google *Corpus*

Dos 100 pacotes lexicais divergentes selecionados, sabemos que 38 pacotes lexicais contêm as características específicas, e 62 não as contêm; ou seja, 62% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 1456 pacotes lexicais x 62%, tivemos um resultado de 902 pacotes lexicais divergentes que são verdadeiros. A diferença (554 pacotes lexicais) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, que, antes, era de 4712 (originais) + 554 (corrigidos), modificou-se para 5266 pacotes lexicais convergentes.

Livro	Pacotes lexicais convergentes ANTES	Pacotes lexicais convergentes DEPOIS	Pacotes lexicais divergentes ANTES	Pacotes lexicais divergentes DEPOIS
Inside Out (comunicativismo)	**4712**	**5266**	**1456**	**902**
Porcentagem	76.3%	85.3%	23.7%	14.7%

Tabela 9: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Inside Out* comparado ao Google *Corpus*.

Como podemos observar, mediante essa recontagem, o número de pacotes lexicais convergentes aumentou de 76.3% para 85.3%, aumentando, assim, o grau de convergência, ou de autenticidade, do livro *Inside Out* quando comparado ao Google *Corpus*.

g) Livro *Interchange* comparado ao *BNC*

Dos 100 pacotes lexicais divergentes selecionados, 26 pacotes lexicais têm as características específicas, e 74 não as têm; ou seja, 74% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 2231 pacotes lexicais x 74%, obtivemos o resultado de 1650 pacotes lexicais divergentes verdadeiros. A diferença (581 *pacotes lexicais*) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, que, antes, era de 1601 (originais) + 581 (corrigidos), tornou-se em 2182 pacotes lexicais convergentes.

Livro	**Pacotes lexicais convergentes ANTES**	**Pacotes lexicais convergentes DEPOIS**	**Pacotes lexicais divergentes ANTES**	**Pacotes lexicais divergentes DEPOIS**
Interchange (comunicativismo)	**1601**	**2182**	**2231**	**1650**
Porcentagem	41.7%	56.9%	59.3%	43.1%

Tabela 10: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Interchange* comparado ao *BNC*.

Como podemos perceber, a partir dessa recontagem, o número de pacotes lexicais convergentes aumentou de 76.3% para 85.3%, aumentando, assim, o grau de convergência, ou de autenticidade, do livro *Interchange* quando comparado ao *BNC*.

h) Livro *Interchange* comparado ao Google *Corpus*

Dos 100 pacotes lexicais divergentes selecionados, 22 pacotes lexicais contêm as características específicas, e 78 não as contêm; ou seja, 78% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 1003 pacotes lexicais x 78%= 782, 782 pacotes lexicais divergentes são verdadeiros. A diferença (221 *pacotes lexicais*) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, antes, era de 2817 (originais) + 221 (corrigidos) = 3038 pacotes lexicais convergentes.

Livro	**Pacotes lexicais convergentes ANTES**	**Pacotes lexicais convergentes DEPOIS**	**Pacotes lexicais divergentes ANTES**	**Pacotes lexicais divergentes DEPOIS**
Interchange (comunicativismo)	**2817**	**3038**	**1003**	**782**
Porcentagem	73.7%	79.5%	26.3%	20.5%

Tabela 11: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Interchange* comparado ao Google *Corpus*.

Como podemos notar, mediante tal recontagem, o número de pacotes lexicais convergentes aumentou de 73.7% para 79.5%, aumentando, assim, o grau de convergência, ou de autenticidade, do livro *Interchange* quando comparado ao Google *Corpus*. Agora, analisaremos o livro da abordagem informada por *corpus*.

i) Livro *Touchstone* comparado ao *BNC*

Dos 100 pacotes lexicais divergentes selecionados, 19 pacotes lexicais têm as características específicas, e 81 não as têm; logo, 81% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 2187 pacotes lexicais x 81%, 1771 pacotes lexicais

divergentes são verdadeiros. A diferença (416 pacotes lexicais) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, antes, era de 1158 (originais) + 416 (corrigidos) = 1574 pacotes lexicais convergentes.

Livro	**Pacotes lexicais convergentes ANTES**	**Pacotes lexicais convergentes DEPOIS**	**Pacotes lexicais divergentes ANTES**	**Pacotes lexicais divergentes DEPOIS**
Touchstone (informada por *corpus*)	**1158**	**1574**	**2187**	**1771**
Porcentagem	34.6%	47.0%	65.4%	53.0%

Tabela 12: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Touchstone* comparado ao *BNC*.

Como podemos observar, com essa recontagem, o número de pacotes lexicais convergentes aumentou de 34.6% para 47%, aumentando, assim, o grau de convergência, ou de autenticidade, do livro *Touchstone* quando comparado ao *BNC*.

j) Livro *Touchstone* comparado ao Google *Corpus*

Dos 100 pacotes lexicais divergentes selecionados, 44 pacotes lexicais têm as características específicas, e 56 não as têm; ou seja, 56% dos pacotes lexicais divergentes são verdadeiros.

Levando em consideração que a contagem inicial total de pacotes lexicais divergentes era de 961 pacotes lexicais x 56%, 538 pacotes lexicais divergentes são verdadeiros. A diferença (423 pacotes lexicais) foi acrescida entre os pacotes lexicais convergentes. Assim, o número de pacotes lexicais convergentes, antes, era de 2370 (originais) + 423 (corrigidos) = 2793 pacotes lexicais convergentes.

	Pacotes	**Pacotes**	**Pacotes**	**Pacotes**

Livro	lexicais convergentes ANTES	lexicais convergentes DEPOIS	lexicais divergentes ANTES	lexicais divergentes DEPOIS
Touchstone (informada por *corpus*)	**2370**	**2793**	**961**	**538**
Porcentagem	71.1%	83.8%	28.9%	16.2%

Tabela 13: Recontagem dos pacotes lexicais convergentes e divergentes do livro *Touchstone* comparado ao Google *Corpus*.

Como podemos constatar, com essa recontagem, o número de pacotes lexicais convergentes aumentou de 71.1% para 83.8%, aumentando, assim, o grau de convergência, ou de autenticidade, do livro *Touchstone* quando comparado ao Google *Corpus*.

3.1.2.1 Conclusão da recontagem

Percebemos, com essa recontagem, um aumento sutil, mas, mesmo assim, expressivo no grau de convergência de alguns livros didáticos. Dessa maneira, demostramos, com esses dados, como aquelas características específicas (nome próprio, numeral e pontuação) podem influenciar no resultado do grau de convergência.

Para uma melhor visualização, mostraremos um gráfico com as porcentagens “anteriores” e “posteriores” à recontagem dos pacotes lexicais convergentes. Primeiro, será feito o gráfico referente ao *BNC*.

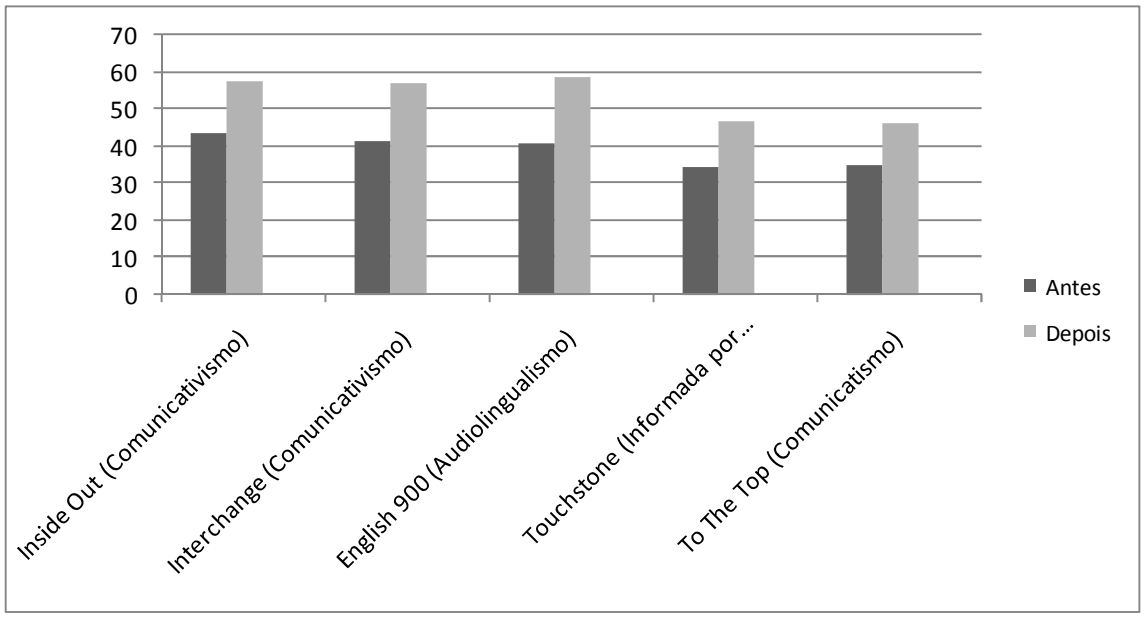


Figura 11: Gráfico após a recontagem dos pacotes lexicais convergentes comparados ao *BNC*.

Com a ajuda do gráfico, notamos que, antes da recontagem, o grau de convergência entre os livros era bem igualado. Mas, após a recontagem, notamos um expressivo aumento no grau de convergência entre todos os livros pesquisados. Isso demonstra que, se eliminarmos as características específicas, o grau de autenticidade tende a aumentar.

Como foi feito o teste estatístico do qui-quadrado, antes de fazer-se a recontagem dos pacotes lexicais, resolvemos refazê-lo para verificar se houve diferença no resultado após a recontagem.

As freqüências das ocorrências nos pacotes lexicais convergentes anteriores e posteriores à recontagem foram submetidas ao teste estatístico qui--quadrado. O valor de qui-quadrado foi 222.27, que corresponde a $p<.0001$. Isso indica que há diferença significativa nas quantidades de pacotes lexicais antes e depois da recontagem. Agora, faremos o mesmo procedimento de análise com o Google *Corpus*.

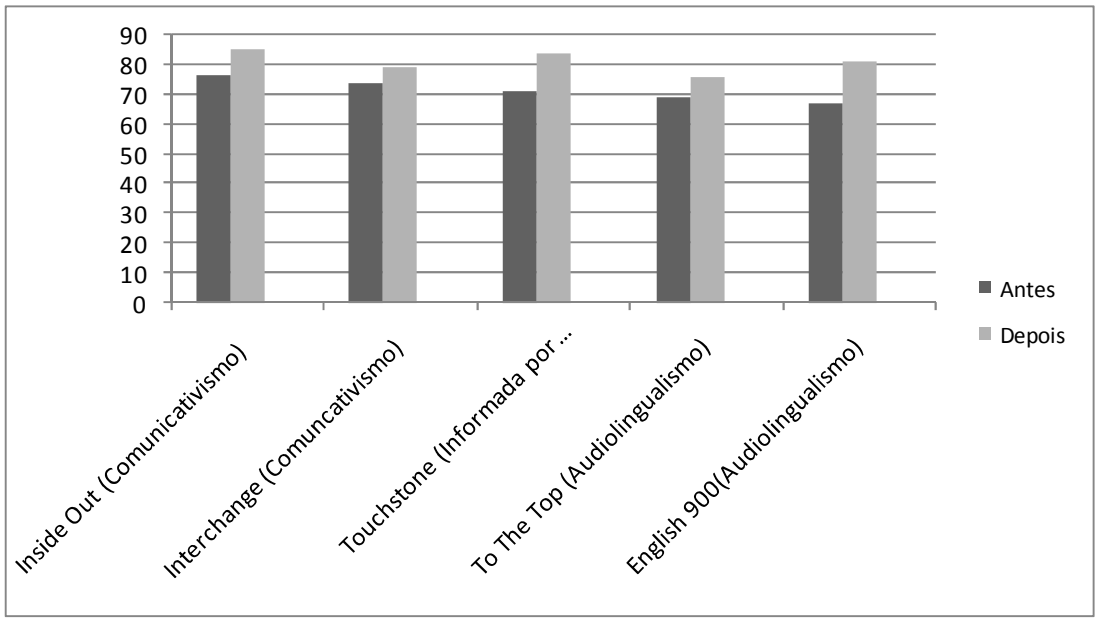


Figura 12: Gráfico após a recontagem dos pacotes lexicais convergentes comparados ao Google *Corpus*.

Com o auxílio do gráfico, notamos que, após a recontagem dos pacotes lexicais convergentes, o grau de autenticidade aumentou, mas esse aumento foi estatisticamente pouco.

As freqüências das ocorrências nos pacotes lexicais convergentes, antes e depois da recontagem, foram submetidas ao mesmo teste estatístico qui--quadrado feito para as outras contagens. O valor de qui-quadrado foi 165.82, que corresponde a $p<.0001$. Isso indica que há diferença significativa nas quantidades de pacotes lexicais antes e depois da recontagem

Chama a atenção o fato de que, mesmo após a recontagem, os livros cujo grau de convergência é superior são os pertencentes à abordagem comunicativa (*Inside Out*), enquanto o livro cujo grau de divergência é superior é o pertencente à abordagem audiolingual. Coincidentemente, é o mesmo livro cujo grau de divergência era superior ao dos demais antes da recontagem: *To the Top*.

Já que analisamos as características dos pacotes lexicais divergentes nos *corpora* pesquisados, resolvemos fazer a análise quantitativa de cada texto para podermos depreender, texto a texto, as características dos pacotes lexicais convergentes.

3.1.3 Análise quantitativa de variação de convergência texto por texto

Um problema da análise apresentada até aqui é que o cálculo de convergência e divergência foi feito com base no *subcorpus* completo de cada livro. É concebível, porém, que haja variação entre os textos de cada livro, de tal modo que alguns sejam mais convergentes do que outros, o que passaria despercebido pelos valores médios. Por isso, é importante que vejamos quais são os textos com mais e menos convergência em cada livro. Para tanto, cada texto teve os seus pacotes convergentes contados (por meio de um programa especialmente criado pelo professor-orientador) e classificados dentro de quatro faixas de convergência:

- alta: convergência = 80%
- média: entre 60 e 79%
- baixa: entre 50 e 59%
- muito baixa: abaixo de 49%

Esta análise tem como objetivo demonstrar a variação entre os textos constantes de cada livro no tocante à convergência. A meta, portanto, não é caracterizar o *subcorpus* completo de cada livro, mas, sim, texto por texto.

Os resultados aparecem nas tabelas abaixo. Para se obterem esses resultados, foi preciso consultar a tabela descrita como faixas de porcentagem de convergência (no anexo 2). Dessa tabela, constam faixas que contêm o número do texto, o número total de pacotes lexicais existentes em cada texto e quantos desses pacotes são convergentes ou divergentes:

Livros	**Faixa de convergência**				
	Alta (≥80%)	**Média (60 a 79%)**	**Baixa (50 a 59%)**	**Muito baixa (0 a 49%)**	**Total**
Inside Out	**5**	**28**	**2**	**1**	**36**
Way Ahead	**3**	**21**	**9**	**0**	**33**
Interchange	**4**	**17**	**4**	**4**	**29**

Touchstone	**3**	**18**	**5**	**2**	**28**
English 900	**2**	**4**	**5**	**0**	**11**

Tabela 14: Quantidade absoluta de pacotes convergentes em cada livro, por faixa de convergência.

Na tabela abaixo, demonstraremos esses dados em porcentagem:

Livros	**Faixa de convergência**				
	Alta (≥80%)	**Média (60 a 79%)**	**Baixa (50 a 59%)**	**Muito baixa (0 a 49%)**	**Total**
Inside Out	**13%**	**78%**	**6%**	**3%**	**100%**
Way Ahead	**9%**	**64%**	**27%**	**0%**	**100%**
Interchange	**13%**	**59%**	**14%**	**14%**	**100%**
Touchstone	**11%**	**64%**	**18%**	**7%**	**100%**
English 900	**18%**	**36%**	**45%**	**0%**	**100%**

Tabela 15: Quantidade relativa, em porcentagem, de pacotes convergentes em cada livro, por faixa de convergência.

E, mais abaixo, os mesmos dados em apenas duas faixas de convergência:

Livros	**Faixa de convergência**		
	Alta e Média (≥60%)	**Baixa ou muito baixa (0 a 59%)**	**Total**
Inside Out	**81%**	**9%**	**100%**
Way Ahead	**73%**	**27%**	**100%**
Interchange	**72%**	**28%**	**100%**
Touchstone	**75%**	**25%**	**100%**
English 900	**55%**	**45%**	**100%**

Tabela 16: Quantidade relativa, em porcentagem, de pacotes convergentes em cada livro, em duas faixas principais.

O quadro acima demonstra que o livro com mais convergência é o *Inside Out*, com 81% de convergência, e o livro com menos convergência também é o *Inside Out*; ou seja, se somarmos as colunas de *'baixa'* e 'm*uito baixa'* convergência, o *Inside Out* fica com 9%. Assim, entendemos que mesmo livros inventados podem conter um alto número de textos convergentes.

Em segundo lugar, o livro com mais convergência é o *Touchstone*. Se nos basearmos na soma das colunas *'alta'* e *'média'*, o *Touchstone* fica com 75% de convergência.

Em terceiro lugar, o livro com mais convergência, com base nas colunas *'alta'* e '*média'*, é o *Way Ahead*, que tem 73% de convergência.

Em quarto lugar, o livro com mais convergência, com base nas colunas *'alta'* e *'média'*, é o *Interchange*, com 72% de convergência

E, finalmente, em último lugar, o livro com menos convergência, com base nas colunas *'alta'* e *'média'*, é o *English 900*, com 55% de convergência. A tabela 16 deixa bem claros os números e revela o contraste existente entre o grau de convergência dos livros entre si, de tal modo que podemos propor a seguinte classificação deles:

+	***RANKING* OBTIDO**
convergente	***Inside Out* (abordagem comunicativa)**
	***Touchstone* (abordagem informada por corpus)**
	***Way Ahead*: *To the Top* (abordagem audiolingual)**
	***Interchange* (abordagem comunicativa)**
	***English 900* (abordagem audiolingual)**
-	

Tabela 17: *Ranking* de convergência de acordo com a tabela 15 (resumida acima).

Baseados no *ranking* acima, apontaremos o texto mais convergente e o menos convergente de todo o *corpus* de estudo. Para melhor entender o texto, ele será apresentado, a seguir, com a seguinte notação:

- Cada pacote lexical é colocado numa linha de texto;

- Todas as linhas são numeradas seqüencialmente, e tal número aparece no início de cada linha;
- Ao final de cada linha, aparece o código *M* (incial do inglês '*match*'), quando o pacote lexical é convergente, e *NM* (inciais do inglês *'no match*'), quando o pacote lexical não é convergente.

Abaixo, demonstramos as 50 primeiras linhas do texto mais convergente de todo o *corpus* de estudo, que foi o de **número 18**, do livro *Inside Out*, com 88,4% de convergência (161 pacotes convergentes, de 182 pacotes no total). Esse texto aparece, na íntegra, no anexo 3, com a notação explicitada acima, ou seja, com a identificação dos pacotes convergentes e não-convergentes.

000001 my_second_and (M)
000002 second_and_only (M)
000003 and_only_other (M)
000004 only_other_memory (M)
000005 other_memory_of (M)
000006 memory_of_llandaff (NM)
000007 of_llandaff_cathedral (M)
000008 llandaff_cathedral_school (M)
000009 cathedral_school_is (M)
000010 school_is_extremely (M)
000011 is_extremely_bizarre (M)
000012 extremely_bizarre_it (NM)
000013 bizarre_it_happened (NM)
000014 it_happened_a (M)
000015 happened_a_little (M)
000016 a_little_over (M)
000017 little_over_a (M)
000018 over_a_year (M)
000019 a_year_later (M)
000020 year_later_when (M)
000021 later_when_i (M)
000022 when_i_was (M)
000023 i_was_just (M)
000024 was_just_nine (M)
000025 just_nine_by (M)
000026 nine_by_then (NM)
000027 by_then_i (M)
000028 then_i_had (M)
000029 i_had_made (M)
000030 had_made_some (M)
000031 made_some_friends (M)
000032 some_friends_and (M)
000033 friends_and_when (M)
000034 and_when_i (M)
000035 when_i_walked (M)
000036 i_walked_to (M)
000037 walked_to_school (M)
000038 to_school_in (M)
000039 school_in_the (M)
000040 in_the_mornings (M)
000041 the_mornings_i (M)

000042 mornings_i_would (M)
000043 i_would_start (M)
000044 would_start_out (M)
000045 start_out_alone (M)
000046 out_alone_but (M)
000047 alone_but_would (M)
000048 but_would_pick (M)
000049 would_pick_up (M)
000050 pick_up_four (M)

O texto menos convergente de todo o *corpus* foi o de **número 13**, do livro *Inside Out*, com apenas 34% de convergência (51 de 150 pacotes). Ele também está incluso, na íntegra, no anexo 4. Nele, também encontramos um dado interessante: se prestarmos atenção, perceberemos, novamente, que os textos mais e menos convergentes pertencem ao mesmo livro. Assim, reforçamos a nossa constatação de que o mesmo livro pode conter textos totalmente autênticos ou totalmente inventados.

Os dados apresentados acima são muito interessantes porque demonstram que mesmo textos altamente simplificados e artificiais podem, por um lado, conter lexicogramática autêntica, isto é, pacotes lexicais convergentes. Por outro, demonstram também que são poucos os textos inventados lexicogramaticalmente autênticos; ou seja, textos inventados e, ao mesmo tempo, lexicogramaticalmente autênticos são a exceção, e não a norma. A autencidade parece andar de mãos dadas com a lexicogramática: um texto autêntico exibe, naturalmente, lexicogramática autêntica. Assim, entendemos, mais uma vez, que nenhum livro entre os pesquisados é totalmente autêntico nem totalmente inventado.

3.2 Análise qualitativa dos *corpora*

Notamos que, à luz do Google *Corpus*, os livros têm um grau de autenticidade superior àquele que têm quando são comparados ao *British National Corpus*; essa discrepância pode ser atribuída a duas características:

- O Google *Corpus* contém 1 (um) trilhão de palavras, enquanto o *BNC* contém 100 (cem) milhões delas. Para esse tipo de análise, quanto maior e mais variado for o *corpus* de referência, melhor será o resultado obtido. E mesmo um *corpus* tão extenso quanto o *BNC* não abarca todos os textos produzidos em língua inglesa;

- O *BNC* é composto de inglês britânico (escrito e falado); já o *corpus* de referência (os livros didáticos) desta pesquisa é composto de inglês norte-americano, e esse detalhe pode ter influenciado na obtenção do resultado referente ao grau de autenticidade.

Esta análise será dividida em duas partes: análise qualitativa dos pacotes lexicais divergentes e análise qualitativa dos pacotes lexicais convergentes.

3.2.1 Análise qualitativa dos pacotes lexicais divergentes

Já que os resultados indicaram diferença entre os livros, passamos a verificar os pacotes divergentes, de cada um deles, para tentarmos entender melhor esses resultados. Assim, apontaremos algumas características importantes existentes nos livros didáticos:

a) Nome próprio:

Nesta categoria, apresentamos pacotes que não constavam dos *corpora* de referência e eram formados por nomes próprios de pessoa ou de lugar. Tais pacotes não são, na verdade, marcadores de falta de autenticidade. São pacotes autênticos, mas não representados nos *corpora* de referência por causa da presença de nomes próprios específicos.

Exemplos retirados dos livros didáticos:

1) Bundle **(NO_ZAHRA_WHY)**
"…MR. NIKZAD: *No, Zahra. Why*?" (English 900)
2) Bundle (**DENISE_DOESNT_HAVE)**
"…is that *Denise doesn't have* a clue that we're doing this." (To the Top)
3) Bundle (**JEALOUS_MEG_NO)**
"…You're just *jealous. Meg: No*, I'm not. I don't want him." (Inside Out)
4) Bundle (**TOKYO_AND_ONE)**
"…I love shopping in *Tokyo. And one* of my favorite stores…" (Interchange)
5) Bundle (**ALICIA_TO_CONTINUE)**

"... very supportive and always encouraged *Alicia to continue* playing." (Touchstone)

b) Pontuação:

Nesta categoria, apresentamos pacotes que não constavam dos *corpora* de referência por causa da falta de pontuação. Essa pontuação foi retirada pelo *script* de comparação dos dados, o que apontou divergência entre os *corpora* (de estudo e de referência), mas, na verdade, ela não significa ausência de autenticidade.

Exemplos retirados dos livros didáticos:

1) Bundle **(AGAIN_ILL_MISS)**
"I hope to see them *again. I'll miss* you…" (English 900)
2) Bundle **(LETS_BE_HONEST)**
"Yes, *let's be honest*, people from English-speaking countries…" (To the Top)
3) Bundle **(YES_HES_GOOD)**
"*Yes, he's good-looking*, interesting, intelligent, and athletic." (Inside Out)
4) Bundle **(ITS_ALWAYS_FULL)**
"…and *it's always full* of interesting people, which is why I like to go there." (Interchange)
5) Bundle **(COULDNT_CALL_IT)**
"We *couldn't "call it"* while we were still inside." (Touchstone)

c) Numeral:

Nesta categoria, apresentamos pacotes que não constavam dos *corpora* de referência e eram formados por números; como eles são, no *corpus* de estudo, características específicas, o número de pacotes lexicais divergentes aumenta.

Exemplos retirados dos livros didáticos:

1)Bundle (**FLIGHT_629_FOR)**

"Last call for Flight 629 for Bogotá, leaving from Gate 10." (English 900)

2)Bundle (**SINCE_1992_AND)**

"… has been active since 1992 and has had a nearly 100% adoption rate." (To the Top)

3)Bundle (**YEAR_1923_WAS)**

"The sweet-shop in Liandaff in the year 1923 was the very centre of our lives…" (Inside Out)

4)Bundle **(OCTOBER_12TH_HE)**

"...and on October 12th, he arrived, in the Indies…" (Interchange)

5) Bundle **(HAS_96_OF)**

"It has 96 of the world's 109 peaks over 7,317 meters…" (Touchstone)

Todas essas características tornam os pacotes lexicais oriundos do *corpus* de estudo muito específicos e acabam dificultando a ligação deles aos dos *corpora* de referência. Por isso, há um número consideravelmente alto de pacotes lexicais divergentes tanto no Google *Corpus* quanto no *BNC*, inclusive porque essas características são compartilhadas por ambos os *corpora*. Na próxima subseção, faremos a análise dos pacotes lexicais convergentes, a fim de entendermos as suas características.

3.2.2 Análise qualitativa dos pacotes lexicais convergentes

Passamos, agora, a analisar os pacotes lexicais convergentes somente no Google *Corpus*, já que o número de pacotes lexicais convergentes nele é superior ao dos convergentes no *BNC*. Para tanto, analisaremos as características gramaticais dos pacotes convergentes mais freqüentes —de cada livro— a partir da classificação de pacotes lexicais oferecida por Biber *et al* (1999) na *Longman Grammar of Spoken and Written English*. O objetivo é saber qual é o registro mais típico dos pacotes de cada livro, bem como a sua estrutura gramatical, a fim de termos um perfil específico de cada livro a partir desses quesitos. Vale ressaltar que Biber *et al* (1999) concentram a sua investigação nos registros acadêmico e conversacional. Buscamos, portanto, entender até que ponto as propostas didáticas do livro e da abordagem cujos pressupostos teóricos nortearam o seu preparo são condizentes com a língua

constante deles, considerada autêntica, uma vez que se trata de pacotes convergentes.

Essa análise foi feita em relação a todos os livros didáticos do nosso *corpus* de estudo; iniciaremos, como temos feito até aqui, com os livros da abordagem audiolingual.

a) Abordagem audiolingual

Começamos esta análise pelo livro *English 900* (1961), que é o mais antigo no *corpus* de estudo e, segundo os resultados da recontagem, na análise quantitativa (item 3.1.2.2), tem 81.% de pacotes lexicais convergentes.

Abaixo, exibimos a tabela dos 20 pacotes lexicais convergentes mais freqüentes no *corpus* de referência Google e a sua classificação, que foi feita com base na *Longman Grammar of Spoken and Written English*, de Biber *et al* (1999). Nessa obra, os autores apresentam os principais pacotes lexicais dos registros 'conversação' e 'acadêmico' (como são chamadas essas variedades) do inglês e os classificam de acordo com a sua composição gramatical. A nossa análise, então, foi norteada pela busca do enquadramento de cada um dos nossos pacotes lexicais na classificação apresentada em Biber *et al* (1999). A correspondência entre os nossos dados e os de Biber *et al* (1999), todavia, não foi, muitas vezes, exata. Em primeiro lugar, devemos salientar que a classificação de Biber *et al* (1999) foi feita com pacotes de 4 e 5 palavras, enquanto os nossos dados são referentes a pacotes de 3 palavras. Dessa maneira, tivemos de enquadrar o nosso pacote lexical como parte dos pacotes lexicais maiores. Em segundo lugar, alguns dos nossos pacotes lexicais não constavam da lista de modo exato. Nesses casos, nós os analisamos gramaticalmente e os encaixamos na categoria mais adequada de Biber *et al* (1999). Abaixo, expomos, enfim, a tabela dos 20 pacotes lexicais convergentes mais freqüentes no *corpus* de referência Google:

Seq.	Pacotes lexicais	Freq. no livro didático	Freq. no google	Registro	Classificação gramatical
1	ONE OF THE	1	119.798.445	Acadêmico	Noun-phrase with of-phrase fragment
2	BE ABLE TO	2	52.502.845	Conversação	To-clause fragments
3	THERE IS A	1	47.759.169	Conversação	Pronoun/noun-phrase + be
4	A LOT OF	4	47.447.438	Conversação	Noun-phrase expressions

5	THERE IS NO	2	39.911.455	Conversação	Pronoun/noun-phrase + be
6	WOULD LIKE TO	1	39.695.800	Conversação	To-clause fragments
7	TO BE A	1	37.583.317	Conversação	To-clause fragments
8	MOST OF THE	1	24.726.337	Conversação	Noun-phrase expressions
9	I WANT TO	2	21.202.830	Conversação	Personal pronoun + lexical phrase
10	OF THE YEAR	1	19.344.874	Acadêmico	Other prepositional phrase fragment
11	TO BE THE	2	19.330.211	Conversação	To-clause fragments
12	THE FIRST TIME	1	18.824.271	Conversação	Noun-phrase expressions
13	FOR THE FIRST	1	18.723.477	Acadêmico	Other prepositional phrase fragment
14	GOING TO BE	4	15.992.589	Conversação	Verb phrase with active verb
15	TO SEE THE	1	15.647.257	Conversação	To-clause fragments
16	LOOKING FOR A	2	15.199.586	Conversação	Verb phrase with active verb
17	YOU HAVE ANY	1	14.806.787	Conversação	Personal pronoun + lexical phrase
18	IN FRONT OF	1	14.259.718	Acadêmico	Other prepositional phrase fragment
19	IN THE NEXT	1	14.041.720	Acadêmico	Other prepositional phrase fragment
20	THAT THERE IS	1	13.761.335	Conversação	(Verb +) that-clause fragments

Tabela 18: Os 20 primeiros *pacotes lexicais matches* do livro *English 900* comparados ao Google *Corpus*.

A maioria dos pacotes desse livro provém da conversação (75%), o que nos parece condizente com a proposta do livro, que é a de se basear na fala. A categoria gramatical mais recorrente é a dos 'To-clause fragments' (25%), que são partes de grupos verbais. Esses pacotes servem para ajudar o aluno a exprimir um elevado número de mensagens, pois podem ser 'afixados' muitos verbos diferentes.

Agora, demonstraremos a tabela referente ao livro *To the Top* (2006), que é o atual da abordagem audiolingual e, segundo os resultados da recontagem, na análise qualitativa (item 3.1.2.4), tem 75.8% de pacotes lexicais convergentes.

Abaixo, está a tabela dos 20 pacotes lexicais convergentes mais freqüentes no *corpus* de referência Google, que inclui a classificação de Biber (1999):

Seq.	Pacotes lexicais	Freq. no livro didático	Freq. no Google	Registro	Classificação gramatical
1	ONE OF THE	2	119.798.445	Acadêmico	Noun-phrase with of-phrase fragment

2	BE ABLE TO	2	52.502.845	Conversação	To-clause fragments
3	FOR MORE INFORMATION	1	49.590.413	Acadêmico	Other prepositional phrase fragment
4	THE END OF	1	48.454.724	Conversação	Noun-phrase with of-phrase fragment
5	A LOT OF	4	47.447.438	Conversação	Noun-phrase expressions
6	THE NUMBER OF	1	45.867.204	Acadêmico	Noun-phrase with of-phrase fragment
7	THERE IS NO	1	39.911.455	Conversação	Pronoun/noun-phrase + be
8	WOULD LIKE TO	1	39.695.800	Conversação	To-clause fragments
9	THIS IS THE	1	37.589.463	Conversação	Pronoun/noun-phrase + be
10	YOU WANT TO	1	36.907.022	Conversação	Personal pronoun + lexical phrase
11	END OF THE	1	31.758.942	Conversação	Noun-phrase expressions
12	THE FACT THAT	1	29.484.748	Acadêmico	Noun-phrase with other post-modifier frements
13	OF THE WORLD	1	29.222.319	Acadêmico	Other prepositional phrase fragment
14	IN ADDITION TO	1	26.502.408	Acadêmico	Other prepositional phrase fragment
15	MEMBER OF THE	1	24.262.402	Acadêmico	Noun-phrase with of-phrase fragment
16	IF YOU WANT	2	22.072.977	Conversação	Adverbial clause fragments
17	A MEMBER OF	1	21.886.455	Acadêmico	Noun-phrase with of-phrase fragment
18	YOU HAVE A	1	21.177.875	Conversação	Personal pronoun + lexical phrase
19	NEED TO BE	1	19.948.745	Conversação	Personal pronoun + lexical phrase
20	OF THE DAY	1	19.809.882	Acadêmico	Other prepositional phrase fragment

Tabela 19: Os 20 primeiros pacotes lexicais *matches* do livro *To the Top: Way Ahead* comparados ao Google *Corpus*.

A maior parte dos pacotes lexicais desse livro provém do registro da conversação, como é o caso do *English 900*, mas em grau inferior (55%). Isso sugere que, embora a ênfase ainda recaia sobre a conversação, o autor do livro já enfoca bastante da linguagem escrita. O tipo de pacote mais freqüente —entre os analisados nesse livro— é o de 'Noun-phrase with of-phrase segment', que serve para expressar informações compactadas no grupo nominal, o que nos parece condizente com a orientação para a língua escrita, que é mais compacta.

Percebemos também que, com o passar do tempo, mesmo nos livros que "pertenciam" à mesma abordagem, o grau de autenticidade teve uma ligeira diminuição, pois o mais antigo, *English 900* (81.1%), apresentou um grau de convergência superior ao do mais atual, *To the Top* (75.8%). Assim, entendemos que o livro atual da abordagem audiolingual não acompanhou as novas tendências de materiais didáticos: continua respeitando as características típicas da sua abordagem. Passaremos, então, a analisar os livros da abordagem comunicativa.

b) Abordagem comunicativa

Começamos esta análise com o livro *Interchange* (1991), que é o mais antigo da abordagem comunicativa e, segundo os resultados da recontagem, na análise quantitativa (item 3.1.2.8), tem 79.5% de pacotes lexicais convergentes.

Abaixo, está a tabela dos 20 pacotes lexicais convergentes mais freqüentes no *corpus* de referência Google, que inclui a classificação de Biber (1999):

Seq.	Pacotes lexicais	Freq. no livro didático	Freq. no google	Registro	Classificação gramatical
1	ONE OF THE	4	119.798.445	Acadêmico	Noun-phrase with of-phrase fragment
2	AS WELL AS	2	99.162.758	Conversação	Adverbial clause fragments
3	THE UNITED STATES	5	62.082.983	Conversação	Noun-phrase expressions
4	IN ORDER TO	1	59.281.305	Acadêmico	Other prepositional phrase fragment
5	BE ABLE TO	1	52.502.845	Conversação	To-clause fragments
6	SOME OF THE	2	51.600.982	Acadêmico	Noun-phrase with of-phrase fragment
7	THERE IS A	1	47.759.169	Conversação	Pronoun/noun-phrase + be
8	A LOT OF	1	47.447.438	Conversação	Noun-phrase expressions
9	THE NUMBER OF	2	45.867.204	Acadêmico	Noun-phrase with of-phrase fragment
10	THERE IS NO	1	39.911.455	Conversação	Pronoun/noun-phrase + be
11	A NUMBER OF	1	38.541.254	Acadêmico	Noun-phrase with of-phrase fragment

12	YOU WANT TO	1	36.907.022	Conversação	Personal pronoun + lexical phrase
13	OUT OF THE	3	33.706.412	Acadêmico	Prepositional phrase with embedded of-phrase fragment
14	GO TO THE	3	29.766.267	Conversação	Verb phrase with active verb
15	IN THE WORLD	2	27.932.952	Acadêmico	Other prepositional phrase fragment
16	IN ADDITION TO	1	26.502.408	Acadêmico	Other prepositional phrase fragment
17	OF THE MOST	2	24.786.212	Acadêmico	Other prepositional phrase fragment
18	ALL KINDS OF	1	22.969.254	Conversação	Quantifier expressions
19	I WANT TO	1	21.202.830	Conversação	Personal pronoun + lexical phrase
20	CAN BE USED	1	19.769.168	Conversação	Passive-verb (+prepositional phrase segment)

Tabela 20: Os 20 primeiros pacotes lexicais *matches* do livro *Interchange* comparados ao Google *Corpus*.

O livro *Interchange* também tem a maioria dos pacotes característica da conversação (55%). Como a proposta didática do comunicativismo engloba tanto a linguagem escrita quanto a falada, parece-nos que essa proporção é adequada, pois reflete a atribuição —a cada modalidade (escrita e fala)— de pesos quase iguais. As características gramaticais predominantes nos pacotes lexicais são '*Noun-phrase with of-phrase fragment*' (com 4 ocorrências) e '*Other prepositional phrase segments*' (também com 4 ocorrências); juntas, correspondem a 67% do total. Isso indica que o livro fornece ao aluno os meios para se comunicar, também na modalidade escrita, por meio desses pacotes lexicais, que são usados para compactar informação nominal, o que, aliás, é típico da modalidade escrita.

Agora, mostraremos a análise do livro *Inside Out* (2004), que é o atual da abordagem comunicativa e, segundo os resultados da recontagem, na análise qualitativa (item 3.1.2.6), tem 85.3% de pacotes lexicais convergentes.

Abaixo, está a tabela dos 20 pacotes lexicais convergentes mais freqüentes no *corpus* de referência Google, que inclui a classificação de Biber (1999):

Seq.	Pacotes lexicais	Freq. no livro didático	Freq. no Google	Registro	Classificação gramatical

1	ONE OF THE	5	119.798.445	Acadêmico	Noun-phrase with of-phrase fragment
2	TO A FRIEND	1	63.393.151	Conversação	To-clause fragments
3	BE ABLE TO	2	52.502.845	Conversação	To-clause fragments
4	THE END OF	3	48.454.724	Conversação	Noun-phrase with of-phrase fragment
5	A LOT OF	2	47.447.438	Conversação	Noun-phrase expressions
6	WOULD LIKE TO	2	39.695.800	Conversação	To-clause fragments
7	THIS IS THE	1	37.589.463	Conversação	Pronoun/noun-phrase + be
8	IT IS NOT	1	33.240.338	Conversação	Pronoun/noun-phrase + be
9	THE UNIVERSITY OF	1	32.785.808	Acadêmico	Noun-phrase with of-phrase fragment
10	END OF THE	1	31.758.942	Conversação	Noun-phrase expressions
11	ALL OF THE	1	31.410.105	Conversação	Quantifier expressions
12	IT IS A	1	31.156.795	Conversação	Pronoun/noun-phrase + be
13	AT THE END	1	25.829.889	Conversação	Prepositional phrase expressions
14	MOST OF THE	1	24.726.337	Conversação	Noun-phrase expressions
15	ACCORDING TO THE	1	23.837.870	Acadêmico	Other prepositional phrase fragment
16	AT THE TIME	1	23.055.984	Conversação	Prepositional phrase expressions
17	THE TOP OF	1	22.853.810	Acadêmico	Noun-phrase with of-phrase fragment
18	AT THE SAME	4	22.247.669	Conversação	Prepositional phrase expressions
19	I WANT TO	7	21.202.830	Conversação	Personal pronoun + lexical phrase
20	YOU HAVE A	5	21.177.875	Conversação	Personal pronoun + lexical phrase

Tabela 21: Os 20 primeiros pacotes lexicais *matches* do livro *Inside Out* comparados ao Google *Corpus*.

Ao contrário dos do *Interchange*, a maioria dos pacotes lexicais do *Inside Out* é característica da conversação (80%). Chama a atenção a proporção de pacotes dos tipos '*To-clause fragments*' (com 3 ocorrências) e '*Pronoun/noun-phrase +be*' (também com 3 ocorrências), que, juntos, respondem por 30% dos pacotes. Esses pacotes lexicais são típicos da conversação, o que reflete a orientação tendente a esse registro notada no cômputo geral dos pacotes. Há, portanto, dois livros —do comunicativismo— bastante diferentes, e o mais recente parece ter-se aproximado da conversação mais do que o seu antecessor.

Percebemos também que, com o passar do tempo, mesmo nos livros pertencentes à mesma abordagem, o grau de convergência teve um aumento considerável, pois, com base nos dados acima, o livro mais antigo, *Interchange* (79.5%), tem grau de convergência inferior ao do mais atual, *Inside Out* (85.3%). Entendemos, por conseguinte, que os livros mais atuais parecem trazer mais estruturas de conteúdo e, desse modo, mais realidade lingüística para os alunos.

c) Abordagem informada por *corpus*

Esta análise foi feita com o livro *Touchstone* (2006), que é o único desta pesquisa pertencente à abordagem informada por *corpus*, que, segundo os resultados da recontagem, na análise qualitativa (item 3.1.2.10), tem 83.8% de pacotes lexicais convergentes.

Abaixo, está a tabela dos 20 pacotes lexicais convergentes mais freqüentes no *corpus* de referência Google, que inclui a classificação de Biber (1999):

Seq.	Pacotes lexicais	Freq. no livro didático	Freq. no Google	Registro	Classificação gramatical
1	ONE OF THE	2	119.798.445	Acadêmico	Noun-phrase with of-phrase fragment
2	THE UNITED STATES	1	62.082.983	Conversação	Noun-phrase expressions
3	SOME OF THE	1	51.600.982	Acadêmico	Noun-phrase with of-phrase fragment
4	A LOT OF	3	47.447.438	Conversação	Noun-phrase expressions
5	THIS IS THE	1	37.589.463	Conversação	Pronoun/noun-phrase + be
6	YOU WANT TO	1	36.907.022	Conversação	Personal pronoun + lexical phrase
7	OUT OF THE	2	33.706.412	Acadêmico	Prepositional phrase with embedded
					of-phrase fragment
8	OF THE WORLD	1	29.222.319	Acadêmico	Other prepositional phrase fragment
9	IN THE WORLD	6	27.932.952	Acadêmico	Other prepositional phrase fragment
10	ACCORDING TO THE	1	23.837.870	Acadêmico	Other prepositional phrase fragment

11	IN THE UNITED	1	22.361.613	Acadêmico	Other prepositional phrase fragment
12	I WANT TO	2	21.202.830	Conversação	Personal pronoun + lexical phrase
13	THAT IT IS	1	20.599.619	Conversação	(Verb +) that-clause fragments
14	IT IS THE	2	19.115.082	Conversação	Pronoun/noun-phrase + be
15	NEW YORK CITY	3	18.799.789	Conversação	Noun-phrase expressions
16	IT WAS A	1	18.360.734	Conversação	Pronoun/noun-phrase + be
17	A COUPLE OF	2	18.150.776	Acadêmico	Noun-phrase with of-phrase fragment
18	TO GO TO	2	17.373.392	Conversação	To-clause fragments
19	AROUND THE WORLD	2	16.975.079	Acadêmico	Other prepositional phrase fragment
20	MANY OF THE	1	15.854.065	Acadêmico	Noun-phrase with of-phrase fragment

Tabela 22: Os 20 primeiros pacotes lexicais *matches* do livro *Touchstone* comparados ao Google *Corpus*.

Esse livro demonstra, por meio dos pacotes analisados, estar composto por exatamente a mesma proporção de unidades provenientes de cada registro (50% da conversação e 50% do registro acadêmico). Isso parece indicar que o autor do livro quer preparar o aluno tanto para lidar com a língua falada quanto com a escrita. As principais características gramaticais dos pacotes são 'Noun-phrase with of-phrase fragment' (com 4 ocorrências, no registro acadêmico), 'Noun-phrase expressions' (com 3 ocorrências, no registro conversação), e 'Other prepositional phrase fragment' (com 5 ocorrências, no registro acadêmico), e, juntas, respondem por 60% dos pacotes. Essas estruturas, conforme dissemos acima, auxiliam na expressão de conteúdos informativos, pois permitem a criação tanto de grupos nominais mais extensos quanto de outros, mais compactos.

Assim, se analisarmos o livro pela porcentagem indicativa do grau de convergência, concluiremos que ele é praticamente tão autêntico quanto os livros da abordagem comunicativa; mas, se observarmos melhor os seus pacotes lexicais de conteúdo, perceberemos que o número deles é superior ao dos pesquisados anteriormente. Essa característica dele deve ser atribuída à sua abordagem, que tem por pressuposto fundamentador a utilização de textos autênticos no desenvolvimento dos seus livros didáticos, de modo a construir

uma ponte entre o conteúdo lingüístico oferecido ao aluno —dentro da sala de aula— e a língua falada —fora dela— pelos nativos.

4 Considerações finais

Constatamos, por meio dessa análise, que a Lingüística de *Corpus* pode auxiliar o pesquisador a encontrar resultados que vão de encontro com a sua intuição. No início desta pesquisa, acreditávamos que o *ranking* dos livros mais autênticos seria bem diferente do que os resultados dela acabaram comprovando. Naquele momento, guiados ainda apenas pela coleta do *corpus* de estudo, acreditávamos que os resultados do grau de autenticidade seriam estes:

+	***RANKING* ESPERADO**	+
convergente	***Touchstone* (abordagem informada por corpus)**	**atual**
	***Inside Out* (abordagem comunicativa)**	
	***Interchange* (abordagem comunicativa)**	
	***Way Ahead: To the Top* (abordagem audiolingual)**	
	***English 900* (abordagem audiolingual)**	
-		-

Tabela 23: Quadro do *ranking* dos livros antes da pesquisa.

Após as pesquisas quantitativas e qualitativas, o *ranking* do grau de convergência teve algumas alterações em relação à nossa intuição.

Iniciaremos a exposição da classificação dos livros pela sua comparação com o *corpus* de referência *BNC*.

+	***RANKING* OBTIDO**
Convergente	***Interchange* (abordagem comunicativa)**
	***English 900* (abordagem audiolingual)**
	***Inside Out* (abordagem comunicativa)**
	***Touchstone* (abordagem informada por corpus)**
	***Way Ahead: To the Top* (abordagem audiolingual)**
-	

Tabela 24: Quadro do *ranking* dos livros após as análises à luz do *BNC*.

Esse *ranking* nos surpreendeu, pois livros que, na nossa intuição, não teriam um grau de convergência tão elevado acabaram ocupando os primeiros lugares. Agora, demonstraremos o *ranking* obtido por meio do *corpus* de referência Google *Corpus*.

+	***RANKING* OBTIDO**
Convergente	***Inside Out* (abordagem comunicativa)**
	***Touchstone* (abordagem informada por *corpus*)**
	***English 900* (abordagem audiolingual)**
	***Interchange* (abordagem comunicativa)**
	***Way Ahead: To the Top* (abordagem audiolingual)**
-	

Tabela 25: Quadro do *ranking* após as análises à luz do Google *Corpus*.

Esse *ranking* vem mais ao encontro da nossa intuição, pois acreditávamos, de início, que os livros mais atuais ocupariam as primeiras posições, como, de fato, ocuparam nele. Como, nesta pesquisa, foram feitos vários tipos de análise, resolvemos demonstrar, abaixo, o *ranking* obtido por meio da análise quantitativa de texto por texto, que foi de suma importância.

+	***RANKING* OBTIDO**
convergente	***Inside Out* (abordagem comunicativa)**
	***Touchstone* (abordagem informada por *corpus*)**
	***Way Ahead: To the Top* (abordagem audiolingual)**
	***Interchange* (abordagem comunicativa)**
	***English 900* (abordagem audiolingual)**
-	

Tabela 26: Quadro do *ranking* após a análise quantitativa de variação de convergência texto por texto.

Essas três perspectivas, como se percebe, são diferentes e fornecem resultados diferentes entre si. A fim de permitir a visualização integrada dos diferentes *rankings*, exibimos, abaixo, uma classificação de 1 (menos convergente) a 5 (mais convergente), a cada livro, nas três classificações: (1.ª) o *corpus* de estudo completo perante o *corpus* de referência *BNC*, (2.ª) o *corpus* de estudo completo perante o *corpus* de referência Google *Corpus*, e (3.ª) cada texto do *corpus* de estudo diante dele mesmo (texto por texto).

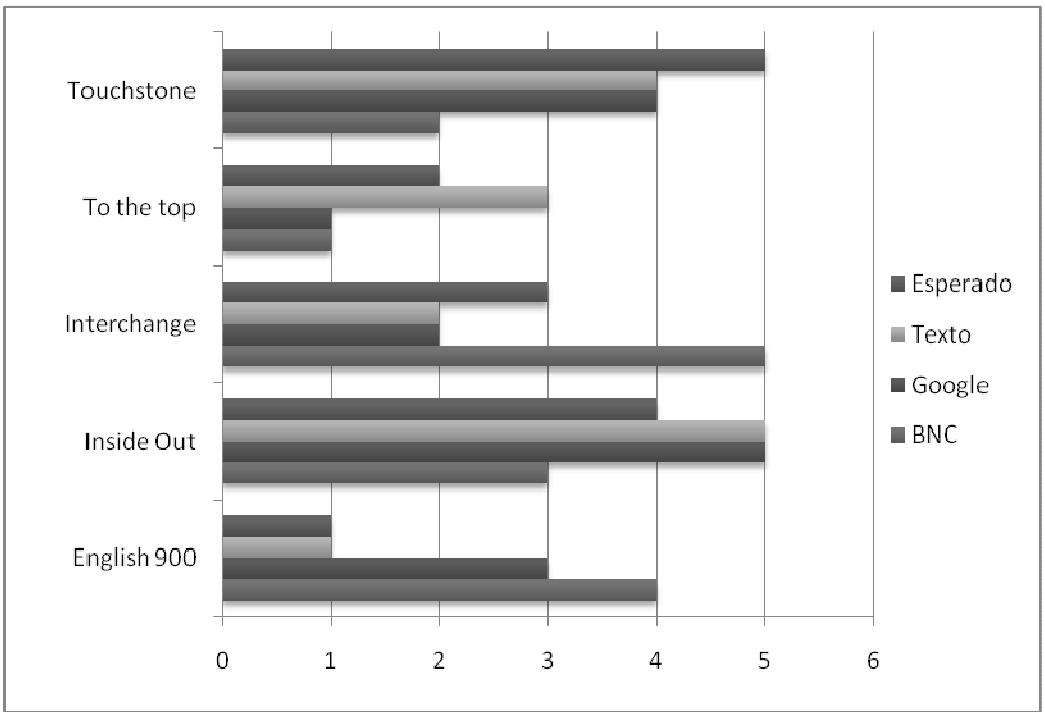


Figura 13: Gráfico de todos os *rankings* obtidos por meio da pesquisa.

O gráfico nos permite comparar a classificação de cada livro à luz de cada análise e ante a classificação anterior à pesquisa. Primeiro, podemos constatar que um dos livros (*Inside Out*) 'superou as expectativas', uma vez que, no gráfico, aparece com barras mais longas do que as esperadas. Ele desponta como aquele que tem uma proporção considerável de lexicogramática autêntica: mais do que era de se esperar de antemão (foi o mais convergente perante o *corpus* Google e na análise texto-a-texto). Também chama a atenção outro livro que superou as expectativas: *English 900*, que provou ter, no total, mais lexicogramática autêntica ante os *corpora* Google e *BNC*. Ele, porém, conforme demonstramos por meio da análise quantitativa de texto por texto, compõe-se de muitos textos claramente inventados, o que baixa a sua classificação quando se faz a verificação de cada um dos textos. Outro livro que chama a atenção é o *Touchstone*, que era tido como aquele que teria mais lexicogramática autêntica, e teve, de fato, perante o *corpus* Google e mediante a análise quantitativa de texto por texto. O *Touchstone* não foi bem classificado perante o *BNC*, o que nos parece refletir o fato de que é um livro recente, enquanto o *BNC* já tem cerca de 20 anos de existência. Sendo assim, muitos dos seus pacotes lexicais do *Touchstone* podem não ter correspondentes no *BNC* justamente porque

refletem temas ou assuntos que não estavam em voga ou não existiam há duas décadas. Essa é apenas uma suposição, pois não podemos fazer uma afirmação tão contundente, já que não fizemos uma análise nesse sentido. Por fim, dois livros tiveram classificação mais baixa do que a esperada: *Way Ahead: To the Top* e *Interchange*. Ambos tiveram classificações inferiores às esperadas; notadamente o *Way Ahead: To the Top*, que não foi bem--classificado perante nenhum dos *corpora* de referência.

Os livros da abordagem audiolingual tiveram classificações bastante diferentes. O *English 900* mostrou-se um livro que apresenta alguma lexicogramática autêntica em textos visivelmente não-autênticos, altamente simplificados, cujo foco está na conversação. Já com o *To the Top*, os resultados são praticamente inversos: os seus textos são um pouco menos simplificados, mas contêm muita lexicogramática pouco autêntica.

Em conclusão, dentre os estudados, os livros cuja lexicogramática é mais autêntica são o *Inside Out* e o *Touchstone*. Ambos são os mais atuais e, portanto, parecem ter seguido a tendência hodierna de apresentar aos estudantes textos autênticos, que, por natureza, constituem-se de lexicogramática também autêntica. Conforme consta do *webiste* do *Inside Out* (http://www.macmillanenglish.com/insideout/html/New_pages/New_insideout_ Course.htm), o livro visa a proporcionar aos estudantes a língua de que eles precisarão fora da sala de aula ('provide students with the key language they need outside the classroom'). Esses dois livros são exemplares da filosofia contemporânea, em TESOL/EFL, de apresentar ao aluno 'a língua real', não fabricada. Ao mesmo tempo, o fato de *Touchstone* não ter-se destacado claramente dos demais indica que o uso de *corpora* na preparação de material didático é importante, pois permite capturar a lexicogramática autêntica. Essa característica, no entanto, não é garantia de que outros livros, não baseados em *corpora* (como o *Inside Out*), não atingirão resultados semelhantes. Parece-nos que à medida que a didática com base em *corpora* , não será mais um 'diferencial', e os autores dos livros didáticos precisarão encontrar outros meios de diferenciá-los dos demais nesse mercado extremamente competitivo, que é o segmento editorial de livros didáticos voltados ao ensino de inglês como LE.

Referências

ALMEIDA FILHO, J. C. P. **Dimensões Comunicativas no Ensino de Línguas**. Campinas. Editora Pontes, 1993.

BERBER SARDINHA, A.P. Usando *WordSmith Tools* na investigação da linguagem**. DIRECT Papers**, v. 40, 1999.

_____. Lingüística de *Corpus*: histórico e problematização. **DELTA**, v.16, n.2, p.323-67, 2000.

_____. **Lingüística de *Corpus*.** Barueri, SP: Monole, p. 410. 2004.

_____. The book is NOT on the table: Autenticidade e idiomaticidade do texto para o ensino de inglês na perspectiva da Lingüística de Corpus. In: Maria Cristina Damianovic. (Org.). **Material Didático: Elaboração e Avaliação**. Taubaté: Cabral Editora, p. 273-286, 2007.

BIBER, D. **Variations across speech and writing**. Cambridge: Cambridge University Press, 1998.

BIBER, D. & REPPEN, R. **Corpus Linguistics: Investigating Language Structure and Use.** Cambridge: Cambridge University Press, 1998.

BIBER, D. et al. **Longman Grammar of Spoken and Written English**. Harlow, Essex: Pearson Education Ltda, 1999.

_____. **Longman Grammar of Spoken and Written English**. London: Longman, 2000.

_____. Using corpus-based methods to investigate grammar and use: some case studies on the use of verbs in English. In: SIMPSON, R.C & SWALES, J. M. (Eds.). *Corpus linguistics in North America: selections from the 1999 symposium USA*: The University of Michigan Press, p. 101-115, 2001.

BIBER, D.& CORTES, V. If you look at…Lexical bundles in university teaching and textbooks. **Applied Linguistics**, v. 25, p. 371-405, 2004.

BROWN, P. & LEVINSON. **Politeness: some universals in language usage**. Cambridge: Cambridge University Press,1987.

CANALE, M. & SWAIN, M. Theoretical bases of communicative approaches to second language teaching and testing. **Applied Linguistics**, | s. l. |, n. 1, p. 1-4, 1980.

CHOMSKY, Noam. **Aspectos da teoria da sintaxe**. Tradução José António Meireles e Eduardo Paiva Raposo. 2. ed. Coimbra: Arménio Amado, 1978. Título original **Aspects of the theory of syntax**.

CORACINI, Maria José Rodrigues Faria. **Interpretação, Autoria e Legitimação do livro didático – língua materna e língua estrangeira.** Campinas-SP: Pontes, 1999.

FLOWERDEW, J. Corpus-based Analyses in EAP. In: J. Flowerdew (ed.), **Academic Discourse**. London: Longman, p.95-114, 2002.

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

GASKELL & COBB. Can learners use concordance feedback for writing error? **System**, v. 32/3, p. 301-319, 2004.

GILMORE, A. A Comparasion of Textbook and Authentic Interaction. **ELT Journal**, V.58, n.4, 2004.

GUARIENTO, W. & MORLEY, J**.** Text and task authenticity in the EFL classroom. **ELT Journal**, v.55, n.4, p.347-353, 2001.

HARMER, Jeremy. **The practice of English Language Teaching**. London: Longman, 2001.

HOEY, M. **Patterns of Lexis in Text**. Oxford: Oxford University Press, 1991.

HUNSTON, S. & FRANCIS, G. **Pattern Grammar: a corpus-driven approach to the lexical grammar of English**. Amsterdã: Benjamins, 1999.

HUNSTON, S. Colligation, lexis, pattern and text. In: SCOTT, M & THOMPSON, G. **Patterns of text- In honour of Michael Hoey**. Amsterdã: Benjamins, p.13-33, 2000.

_____. Introduction to a corpus in use. In: **Corpora in Applied Linguistics**. Cambridge, p.1-24, 2002.

JACOBI, C. C. B. D. **Lingüística de Corpus e ensino de espanhol a brasileiros: Descrição de padrões e preparação de atividades didáticas (decir/hablar ; mismo; mientras /en cuanto/ aunque).** Dissertação de Mestrado, PUCSP, São Paulo, 2001.

JOHNS, T. 'Should you be persuaded: Two exemplos data- driving learning' in T. Johns, and P. King (eds). Classroom Concordancing, **ELR Journal 4**, vol.1, p. 16, 1991.

KENNEDY, G. **An Introduction to Corpus Linguistics**. Nova York, Longman, 1998.

KUMARAVADIVELU, B. The postmethod condition: emerging strategies for second/foreign language teaching. In: **TESOL Quarterly**, Washington, v. 28, n. 1, p. 27-49, 1994.

_____. **Beyond methods**. New Haven/ London: Yale University Press, 2003.

LEE, W. Y. Authenticity revisited: Text authenticity and learner authenticity. **ELT Journal**, 49, 4, 323-28, 1995.

LEECH, G. **Semantics**. Harmondsworth: Penguins, 1974.

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

_____**.** Introducing corpus annotation. In: R. GARSIDE et al (org**.). Corpus Annotation – Linguistic Information from Computer Text Corpora.** London and New York: Longman, 1997.

LEWIS, M. **Teaching Collocations: further developments in the Lexical Approach.** London: LTP, 2000.

LITTLE, D. & D. SINGLETON. 'Authentic texts, pedagogical grammar and language awareness in foreign language learning' in C. James and P. Garret (eds.). **Language Awareness in the Classroom***.* London: Longman: 123-32, 1991.

MAZZA, Luciene Novais. OS LEXICAL BUNDLES EM UM CORPUS ESPECÍFICO. In: SEMINÁRIO DO GEL, 57., 2009, **Programação...** Ribeirão Preto (SP): GEL, 2009. Disponível em: <http://www.gel.org.br/?resumo=5768-09>. Acesso em: 07/06/2009.

McCARTHY, M. **Spoken language and applied linguistics.** Cambridge: Cambridge University Press, 1998.

McCARTHY, M, MCCARTEN, J &SANDIFORD, H. **Touchstone 3.** Cambridge: Cambridge University Press, 2006.

McENERY, T. & WILSON. A. **Corpus Linguistics**. Edinburgh: Edinburgh University Press, 1996.

MELERO ABADÍA, P. **Métodos y Enfoques en la enseñanza, aprendizaje de español como lengua extranjera** .Madrid: Edelsa Grupo Didascalia, 2000.

MISHAN, F. Authenticating Corpora for Language Learning: a problem and its resolution. **ELT Journal**, v.58, n.3, p.219-227, 2004.

_____. **Designing Authenticity into Language Learning Materials.** Intellect, Bristol, 2005.

MORRISON, B. 'Using news broadcasts for authentic listening comprehension'. **ELT Journal** *, v.*43/1, p. 14-18, 1989.

NUNAN, D. **Introducing Discouse Analysis**. England: Penguin Books, 1993.

PAIVA, V.L.M.O.Como se aprende uma língua estrangeira? In: ANASTÁCIO, E.B.A.; MALHEIROS, M.R.T.L.; FIGLIOLINI, M.C.R. (Orgs). **Tendências contemporâneas em Letras.** Campo Grande: Editora da UNIDERP, p. 127-140, 2005.

PARTINGTON, A. **Patterns and meanings: using corpora for English language research and teaching.** Studies in Corpus Linguistics, 1998.

PEACOCK, M. The effect of authentic materials on the motivation of EFL Learners. **ELT Journal**, 51, 2, 144-53, 1997.

PEDREIRO, S.F. **O Movimento Comunicativo de Ensino de Língua Estrangeira no Brasil.** Dissertação de Mestrado, UnB, 2002.

PEREZ-LLANTADA, Carmen. Textual, genre and social features of spoken grammar. **A corpus-based approach**, v. 13/1, p.40-58, 2009.

RICHARDS, J. C. & RODGERS, T.S. **Approach and Methods in Language Teaching.** New York: Cambridge University Press, 1986.

_____. **Approaches and Methods in Language Teaching**. Cambridge: Cambridge University Press, 2001.

SAMPSON, G & McCARTHY, D. **Corpus Linguistics: readings in a widening discipline.** London and New York: Continuum International, p. 524, *2004.*

SANCHEZ PEREZ, A . **La Enseñanza de Idiomas**. Barcelona: Hora, 1982.

SANCHEZ, A. & P. CANTOS. El ritmo incremental de palabras nuevas en los repertorios de textos. Estudio experimental y comparativo basado en dos corpus linguisticos equivalentes de cuatro millones de palabras, de las lenguas inglesa y espanola y en cinco autores de ambas lenguas. **Atlantis**, v.19, n.2, p.1-27, 1997.

SANCHEZ, A. Definición e historia de los corpus. In: SANCHEZ,A. et al. (Orgs.). **Cumbre: corpus linguistico de Español contemporaneo.** Madri: SGEL, p. 7-24, 1995.

SCHÜTZ, Ricardo. **Communicative Approach - Abordagem Comunicativa**, 2006. Disponível em: <http://www.sk.com.br/sk-comm.html>. Acesso em: 12 jun. 2009.

SCOTT, M. **WordSmith Tools**. Oxford: Oxford University Press, 1996.

SCOTT, M.; TRIBBLE, C. Textual patterns: key words and corpus analysis in language education. In: TOGNINI-BOBELLI (Org.) **Studies in corpus linguistics**, vol. 22. Amsterdã: John Benjamins, 2006.

SHORTALL, Terry. The L2 syllabus: corpus or contrivance? **Corpora,** v.2, n.2, p.157-185, 2007.

SINCLAIR , J. **Corpus, Concordance, Collocation.** Oxford: Oxford University Press, 1991.

STUBBS, M**. Words and Phrases: Corpus Studies of Lexical Semantics**. Oxford: Blackwell, 2001

STUBBS, M .(With M Hoey, M Mahlberg & W Teubert) **Text, Discourse and Corpora.** London: Continuum, 2007.

TICKS, Luciane. O livro didático sob a ótica do genero. **Linguagem & Ensino,** v.8/1, p.15-49, 2005.

SWEET, H. **The Practical Study of Languages**. London: J.M. Dent and Co, 1899.

YOON, H., & HIRVELA, A. ESL student attitudes toward corpus use in L2 writing. **Journal of Second Language Writing**, v. 13, p. 257-283, 2004.

WIDDOWSON, H. G. **Teaching Language as Communication**. London: Oxford University Press, 1978.

_____. **Explorations in Applied Linguistic**. Oxford: Oxford University Press, 1979.

WILLIAMS, E. 'Communicative reading' in K. Johnson and D. Porter (eds.). **Perspectives in Communicative Language Teaching.** London: Academic Press: 171-88, 1983.

WILLIS, D. **The lexical syllabus: a new approach to language teaching**. Glasgow: Collins ELT, 1990.

ANEXOS

Anexo 1: Script criado pelo orientador.

```
#! /bin/sh

#script for whole books, in corpus folder

#check if files are true ascii
#head *.txt | less
#if not, convert with col -b < intext > outtext
#yes!

ascii () {

ls corpus | grep -v ascii > files

while read file
do
 col -b < corpus/$file > corpus/$file.ascii
done < files

}

normbnc () {

tr ' ' '_' < bnc_3gm.txt | gsort > bnc_3_norm.txt

}

makeclusters () {

ls corpus | grep ascii > files

while read text
do
 echo "..... doing corpus/"$text" "
 cat corpus/"$text" |  tr '\r' '\n' | tr '~' ' ' | tr '\n' '~' | tr -dc '[:print:]' | tr '~' '\n' | sed
"s/\(['']\)\([sdtlrmv]\)/APSTR\2/g" | tr '\n' ' ' | sed -e 's/-/ /g' -e 's/\$/ /g'  | sed 's/v.sa/v sa/g' |
sed  's/\([0-9]\)[,\.]\([0-9]\)/\1 \2/g'| tr -d '[:punct:]'  | tr -d '—“”…’’´——\t' | tr ' ' '\n' | tr -s '\n
'| sed "s/APSTR/'/g" | sed 's/^[ ]*//' | grep -v '^$' | tr '[a-záéíóúãõçëüâêôûà]'  '[A-
ZÁÉÍÓÚÃÕÇËÜÂÊÔÛÀ]'> "$text".token
done < files

ls *.token > files

while read file
do
 echo "..... doing $file ....."
 cat $file > 1
 tail +2 1 > 2
 tail +3 1 > 3
 gpaste -d'_' 1 2 3  | gsort | uniq -c | sed 's/^[ ]*//' | sed 's/\([0-9]*\) \(.*\)/\2       \1/' >
$file.bundles.norm
done < files
```

```
rm 1 2 3

}

matches () {

for f in `ls *.norm | grep ascii`
do
 gjoin  -1 1 -2 1 -o 1.1 -o 1.2 -o 2.2 $f bnc_3_norm.txt > $f.bnc
 echo "saved $f.bnc"
done

}

nomatches () {

for f in `ls *.norm | grep ascii`
do
 gjoin  -1 1 -2 1 -o 1.1 -o 1.2 -o 2.2 -a 1 -e 0 $f bnc_3_norm.txt | grep ' 0$' | grep -v '^[0-9]'
> $f.nomatch
 echo "saved $f.nomatch"
done

}


stats () {

ls *.bnc > files
ls *.nomatch >> files

#for f in `ls *.* | grep norm | grep ascii | grep -v -e 'bnc_3' -e 'results'`
while read f
do
 count=$(cat $f | grep -v '^$' | wc -l)
 echo "$f $count"
done < files > results_bnc_freq1orhigher.txt
#done > results.txt

}

google () {

ls *.norm | nl > files

while read n file
do
 tr '\t' ',' < $file | tr '_' ' '  | sed 's/^/,/' > $file.google.tmp
 sed "s/XXXFILEXXX/$file/" googlebundles_template.sql > googlebundles.$n.sql
 mysql -u root < googlebundles.$n.sql
done < files

}

countgoogle () {

rm results_google.txt

for file in `ls *.google`
do
```

```
 grep -v '^[0-9]' $file | grep '1,0$' > $file.nomatch
 grep -v '^[0-9]' $file | grep -v '1,0$' > $file.match
 countmatch=$(cat $file.match | grep -v '^$' | wc -l)
 countnomatch=$(cat $file.nomatch | grep -v '^$' | wc -l)
 echo "$file.match $countmatch" >> results_google.txt
 echo "$file.nomatch $countnomatch" >> results_google.txt
done

}


#ascii
###normbnc
#makeclusters
#matches
#nomatches
#stats
##google
countgoogle
```

Anexo 2: Faixas com % de convergência.

1	textmatches/inside_out.txt.ascii.fixed-18.txt.bundles.matches | bundles= 182 | matches=161 | 88.400 %
2	textmatches/way_ahead.txt.ascii.fixed-09.txt.bundles.matches | bundles= 189 | matches=165 | 87.300 %
3	textmatches/interchange.txt.ascii.fixed-09.txt.bundles.matches | bundles= 99 | matches=85 | 85.800 %
4	textmatches/inside_out.txt.ascii.fixed-15.txt.bundles.matches | bundles= 39 | matches=33 | 84.600 %
5	textmatches/touchstone.txt.ascii.fixed-14.txt.bundles.matches | bundles= 84 | matches=71 | 84.500 %
6	textmatches/interchange.txt.ascii.fixed-16.txt.bundles.matches | bundles= 76 | matches=64 | 84.200 %
7	textmatches/interchange.txt.ascii.fixed-17.txt.bundles.matches | bundles= 37 | matches=31 | 83.700 %
8	textmatches/english_900.txt.ascii.fixed-11.txt.bundles.matches | bundles= 164 | matches=137 | 83.500 %
9	textmatches/touchstone.txt.ascii.fixed-02.txt.bundles.matches | bundles= 97 | matches=80 | 82.400 %
10	textmatches/inside_out.txt.ascii.fixed-14.txt.bundles.matches | bundles= 509 | matches=415 | 81.500 %
11	textmatches/way_ahead.txt.ascii.fixed-27.txt.bundles.matches | bundles= 497 | matches=402 | 80.800 %
12	textmatches/inside_out.txt.ascii.fixed-33.txt.bundles.matches | bundles= 270 | matches=218 | 80.700 %
13	textmatches/touchstone.txt.ascii.fixed-11.txt.bundles.matches | bundles= 281 | matches=226 | 80.400 %
14	textmatches/interchange.txt.ascii.fixed-04.txt.bundles.matches | bundles= 122 | matches=98 | 80.300 %
15	textmatches/english_900.txt.ascii.fixed-04.txt.bundles.matches | bundles= 264 | matches=212 | 80.300 %
16	textmatches/way_ahead.txt.ascii.fixed-13.txt.bundles.matches | bundles= 283 | matches=227 | 80.200 %
17	textmatches/inside_out.txt.ascii.fixed-16.txt.bundles.matches | bundles= 298 | matches=239 | 80.200 %
18	textmatches/inside_out.txt.ascii.fixed-08.txt.bundles.matches | bundles= 253 | matches=202 | 79.800 %
19	textmatches/inside_out.txt.ascii.fixed-11.txt.bundles.matches | bundles= 84 | matches=67 | 79.700 %
20	textmatches/inside_out.txt.ascii.fixed-36.txt.bundles.matches | bundles= 117 | matches=92 | 78.600 %
21	textmatches/inside_out.txt.ascii.fixed-01.txt.bundles.matches | bundles= 504 | matches=396 | 78.500 %
22	textmatches/interchange.txt.ascii.fixed-08.txt.bundles.matches | bundles= 315 | matches=247 | 78.400 %
23	textmatches/english_900.txt.ascii.fixed-05.txt.bundles.matches | bundles= 186 | matches=145 | 77.900 %
24	textmatches/touchstone.txt.ascii.fixed-12.txt.bundles.matches | bundles= 69 | matches=53 | 76.800 %

25	textmatches/way_ahead.txt.ascii.fixed-07.txt.bundles.matches | bundles= 60 | matches=46 | 76.600 %
26	textmatches/interchange.txt.ascii.fixed-29.txt.bundles.matches | bundles= 47 | matches=36 | 76.500 %
27	textmatches/inside_out.txt.ascii.fixed-22.txt.bundles.matches | bundles= 128 | matches=98 | 76.500 %
28	textmatches/touchstone.txt.ascii.fixed-22.txt.bundles.matches | bundles= 85 | matches=65 | 76.400 %
29	textmatches/touchstone.txt.ascii.fixed-09.txt.bundles.matches | bundles= 242 | matches=185 | 76.400 %
30	textmatches/inside_out.txt.ascii.fixed-06.txt.bundles.matches | bundles= 34 | matches=26 | 76.400 %
31	textmatches/interchange.txt.ascii.fixed-20.txt.bundles.matches | bundles= 82 | matches=62 | 75.600 %
32	textmatches/interchange.txt.ascii.fixed-14.txt.bundles.matches | bundles= 695 | matches=526 | 75.600 %
33	textmatches/way_ahead.txt.ascii.fixed-12.txt.bundles.matches | bundles= 86 | matches=65 | 75.500 %
34	textmatches/interchange.txt.ascii.fixed-24.txt.bundles.matches | bundles= 336 | matches=254 | 75.500 %
35	textmatches/inside_out.txt.ascii.fixed-12.txt.bundles.matches | bundles= 356 | matches=269 | 75.500 %
36	textmatches/interchange.txt.ascii.fixed-28.txt.bundles.matches | bundles= 44 | matches=33 | 75.000 %
37	textmatches/touchstone.txt.ascii.fixed-03.txt.bundles.matches | bundles= 91 | matches=68 | 74.700 %
38	textmatches/inside_out.txt.ascii.fixed-09.txt.bundles.matches | bundles= 424 | matches=316 | 74.500 %
39	textmatches/way_ahead.txt.ascii.fixed-03.txt.bundles.matches | bundles= 72 | matches=53 | 73.600 %
40	textmatches/interchange.txt.ascii.fixed-25.txt.bundles.matches | bundles= 148 | matches=109 | 73.600 %
41	textmatches/inside_out.txt.ascii.fixed-30.txt.bundles.matches | bundles= 53 | matches=39 | 73.500 %
42	textmatches/inside_out.txt.ascii.fixed-05.txt.bundles.matches | bundles= 60 | matches=44 | 73.300 %
43	textmatches/inside_out.txt.ascii.fixed-03.txt.bundles.matches | bundles= 67 | matches=49 | 73.100 %
44	textmatches/touchstone.txt.ascii.fixed-27.txt.bundles.matches | bundles= 85 | matches=62 | 72.900 %
45	textmatches/touchstone.txt.ascii.fixed-24.txt.bundles.matches | bundles= 102 | matches=74 | 72.500 %
46	textmatches/interchange.txt.ascii.fixed-02.txt.bundles.matches | bundles= 216 | matches=155 | 71.700 %
47	textmatches/inside_out.txt.ascii.fixed-35.txt.bundles.matches | bundles= 141 | matches=100 | 70.900 %
48	textmatches/touchstone.txt.ascii.fixed-23.txt.bundles.matches | bundles= 72 | matches=51 | 70.800 %
49	textmatches/touchstone.txt.ascii.fixed-04.txt.bundles.matches | bundles= 68 | matches=48 | 70.500 %
50	textmatches/interchange.txt.ascii.fixed-19.txt.bundles.matches | bundles= 81 | matches=57 | 70.300 %
51	textmatches/inside_out.txt.ascii.fixed-10.txt.bundles.matches | bundles= 230 | matches=161 | 70.000 %
52	textmatches/interchange.txt.ascii.fixed-10.txt.bundles.matches | bundles= 123 | matches=86 | 69.900 %
53	textmatches/interchange.txt.ascii.fixed-27.txt.bundles.matches | bundles= 56 | matches=39 | 69.600 %
54	textmatches/inside_out.txt.ascii.fixed-31.txt.bundles.matches | bundles= 79 | matches=55 | 69.600 %
55	textmatches/interchange.txt.ascii.fixed-22.txt.bundles.matches | bundles= 276 | matches=192 | 69.500 %
56	textmatches/interchange.txt.ascii.fixed-21.txt.bundles.matches | bundles= 85 | matches=59 | 69.400 %
57	textmatches/way_ahead.txt.ascii.fixed-25.txt.bundles.matches | bundles= 52 | matches=36 | 69.200 %
58	textmatches/touchstone.txt.ascii.fixed-05.txt.bundles.matches | bundles= 143 | matches=99 | 69.200 %
59	textmatches/way_ahead.txt.ascii.fixed-31.txt.bundles.matches | bundles= 173 | matches=119 | 68.700 %
60	textmatches/inside_out.txt.ascii.fixed-28.txt.bundles.matches | bundles= 48 | matches=33 | 68.700 %
61	textmatches/inside_out.txt.ascii.fixed-04.txt.bundles.matches | bundles= 32 | matches=22 | 68.700 %
62	textmatches/way_ahead.txt.ascii.fixed-15.txt.bundles.matches | bundles= 54 | matches=37 | 68.500 %
63	textmatches/english_900.txt.ascii.fixed-08.txt.bundles.matches | bundles= 304 | matches=208 | 68.400 %
64	textmatches/touchstone.txt.ascii.fixed-08.txt.bundles.matches | bundles= 166 | matches=113 | 68.000 %
65	textmatches/inside_out.txt.ascii.fixed-34.txt.bundles.matches | bundles= 159 | matches=108 | 67.900 %
66	textmatches/interchange.txt.ascii.fixed-12.txt.bundles.matches | bundles= 56 | matches=38 | 67.800 %
67	textmatches/touchstone.txt.ascii.fixed-17.txt.bundles.matches | bundles= 62 | matches=42 | 67.700 %
68	textmatches/way_ahead.txt.ascii.fixed-28.txt.bundles.matches | bundles= 262 | matches=177 | 67.500 %
69	textmatches/interchange.txt.ascii.fixed-18.txt.bundles.matches | bundles= 92 | matches=62 | 67.300 %

70	textmatches/inside_out.txt.ascii.fixed-02.txt.bundles.matches | bundles= 46 | matches=31 | 67.300 %
71	textmatches/inside_out.txt.ascii.fixed-17.txt.bundles.matches | bundles= 681 | matches=458 | 67.200 %
72	textmatches/inside_out.txt.ascii.fixed-19.txt.bundles.matches | bundles= 145 | matches=97 | 66.800 %
73	textmatches/way_ahead.txt.ascii.fixed-23.txt.bundles.matches | bundles= 39 | matches=26 | 66.600 %
74	textmatches/touchstone.txt.ascii.fixed-19.txt.bundles.matches | bundles= 63 | matches=42 | 66.600 %
75	textmatches/inside_out.txt.ascii.fixed-23.txt.bundles.matches | bundles= 432 | matches=288 | 66.600 %
76	textmatches/inside_out.txt.ascii.fixed-20.txt.bundles.matches | bundles= 60 | matches=40 | 66.600 %
77	textmatches/inside_out.txt.ascii.fixed-27.txt.bundles.matches | bundles= 115 | matches=76 | 66.000 %
78	textmatches/way_ahead.txt.ascii.fixed-11.txt.bundles.matches | bundles= 44 | matches=29 | 65.900 %
79	textmatches/way_ahead.txt.ascii.fixed-08.txt.bundles.matches | bundles= 278 | matches=183 | 65.800 %
80	textmatches/inside_out.txt.ascii.fixed-24.txt.bundles.matches | bundles= 129 | matches=85 | 65.800 %
81	textmatches/touchstone.txt.ascii.fixed-16.txt.bundles.matches | bundles= 321 | matches=211 | 65.700 %
82	textmatches/interchange.txt.ascii.fixed-03.txt.bundles.matches | bundles= 124 | matches=81 | 65.300 %
83	textmatches/inside_out.txt.ascii.fixed-29.txt.bundles.matches | bundles= 49 | matches=32 | 65.300 %
84	textmatches/touchstone.txt.ascii.fixed-25.txt.bundles.matches | bundles= 118 | matches=77 | 65.200 %
85	textmatches/way_ahead.txt.ascii.fixed-04.txt.bundles.matches | bundles= 86 | matches=56 | 65.100 %
86	textmatches/english_900.txt.ascii.fixed-03.txt.bundles.matches | bundles= 201 | matches=131 | 65.100 %
87	textmatches/touchstone.txt.ascii.fixed-13.txt.bundles.matches | bundles= 174 | matches=113 | 64.900 %
88	textmatches/inside_out.txt.ascii.fixed-21.txt.bundles.matches | bundles= 91 | matches=59 | 64.800 %
89	textmatches/touchstone.txt.ascii.fixed-18.txt.bundles.matches | bundles= 88 | matches=57 | 64.700 %
90	textmatches/english_900.txt.ascii.fixed-07.txt.bundles.matches | bundles= 345 | matches=223 | 64.600 %
91	textmatches/way_ahead.txt.ascii.fixed-29.txt.bundles.matches | bundles= 152 | matches=98 | 64.400 %
92	textmatches/way_ahead.txt.ascii.fixed-26.txt.bundles.matches | bundles= 84 | matches=54 | 64.200 %
93	textmatches/interchange.txt.ascii.fixed-11.txt.bundles.matches | bundles= 355 | matches=228 | 64.200 %
94	textmatches/way_ahead.txt.ascii.fixed-19.txt.bundles.matches | bundles= 55 | matches=35 | 63.600 %
95	textmatches/way_ahead.txt.ascii.fixed-02.txt.bundles.matches | bundles= 77 | matches=49 | 63.600 %
96	textmatches/inside_out.txt.ascii.fixed-32.txt.bundles.matches | bundles= 270 | matches=171 | 63.300 %
97	textmatches/way_ahead.txt.ascii.fixed-05.txt.bundles.matches | bundles= 95 | matches=60 | 63.100 %
98	textmatches/way_ahead.txt.ascii.fixed-01.txt.bundles.matches | bundles= 92 | matches=58 | 63.000 %
99	textmatches/touchstone.txt.ascii.fixed-28.txt.bundles.matches | bundles= 105 | matches=66 | 62.800 %
100	textmatches/way_ahead.txt.ascii.fixed-10.txt.bundles.matches | bundles= 67 | matches=42 | 62.600 %
101	textmatches/inside_out.txt.ascii.fixed-25.txt.bundles.matches | bundles= 32 | matches=20 | 62.500 %
102	textmatches/way_ahead.txt.ascii.fixed-14.txt.bundles.matches | bundles= 188 | matches=117 | 62.200 %
103	textmatches/touchstone.txt.ascii.fixed-07.txt.bundles.matches | bundles= 371 | matches=229 | 61.700 %
104	textmatches/way_ahead.txt.ascii.fixed-20.txt.bundles.matches | bundles= 51 | matches=31 | 60.700 %
105	textmatches/way_ahead.txt.ascii.fixed-30.txt.bundles.matches | bundles= 146 | matches=88 | 60.200 %
106	textmatches/english_900.txt.ascii.fixed-09.txt.bundles.matches | bundles= 452 | matches=271 | 59.900 %
107	textmatches/touchstone.txt.ascii.fixed-26.txt.bundles.matches | bundles= 148 | matches=88 | 59.400 %
108	textmatches/english_900.txt.ascii.fixed-10.txt.bundles.matches | bundles= 370 | matches=220 | 59.400 %
109	textmatches/english_900.txt.ascii.fixed-06.txt.bundles.matches | bundles= 188 | matches=111 | 59.000 %
110	textmatches/interchange.txt.ascii.fixed-15.txt.bundles.matches | bundles= 51 | matches=30 | 58.800 %
111	textmatches/way_ahead.txt.ascii.fixed-32.txt.bundles.matches | bundles= 216 | matches=127 | 58.700 %
112	textmatches/way_ahead.txt.ascii.fixed-17.txt.bundles.matches | bundles= 46 | matches=27 | 58.600 %
113	textmatches/inside_out.txt.ascii.fixed-07.txt.bundles.matches | bundles= 99 | matches=58 | 58.500 %
114	textmatches/way_ahead.txt.ascii.fixed-18.txt.bundles.matches | bundles= 60 | matches=35 | 58.300 %

115	textmatches/way_ahead.txt.ascii.fixed-33.txt.bundles.matches | bundles= 115 | matches=67 | 58.200 %
116	textmatches/interchange.txt.ascii.fixed-13.txt.bundles.matches | bundles= 28 | matches=16 | 57.100 %
117	textmatches/touchstone.txt.ascii.fixed-21.txt.bundles.matches | bundles= 100 | matches=57 | 57.000 %
118	textmatches/way_ahead.txt.ascii.fixed-21.txt.bundles.matches | bundles= 30 | matches=17 | 56.600 %
119	textmatches/way_ahead.txt.ascii.fixed-22.txt.bundles.matches | bundles= 71 | matches=40 | 56.300 %
120	textmatches/interchange.txt.ascii.fixed-07.txt.bundles.matches | bundles= 93 | matches=52 | 55.900 %
121	textmatches/touchstone.txt.ascii.fixed-06.txt.bundles.matches | bundles= 104 | matches=58 | 55.700 %
122	textmatches/way_ahead.txt.ascii.fixed-16.txt.bundles.matches | bundles= 94 | matches=52 | 55.300 %
123	textmatches/way_ahead.txt.ascii.fixed-06.txt.bundles.matches | bundles= 58 | matches=32 | 55.100 %
124	textmatches/english_900.txt.ascii.fixed-01.txt.bundles.matches | bundles= 313 | matches=170 | 54.300 %
125	textmatches/way_ahead.txt.ascii.fixed-24.txt.bundles.matches | bundles= 35 | matches=19 | 54.200 %
126	textmatches/touchstone.txt.ascii.fixed-10.txt.bundles.matches | bundles= 110 | matches=58 | 52.700 %
127	textmatches/english_900.txt.ascii.fixed-02.txt.bundles.matches | bundles= 416 | matches=218 | 52.400 %
128	textmatches/touchstone.txt.ascii.fixed-01.txt.bundles.matches | bundles= 111 | matches=58 | 52.200 %
129	textmatches/inside_out.txt.ascii.fixed-26.txt.bundles.matches | bundles= 37 | matches=19 | 51.300 %
130	textmatches/interchange.txt.ascii.fixed-26.txt.bundles.matches | bundles= 49 | matches=25 | 51.000 %
131	textmatches/touchstone.txt.ascii.fixed-20.txt.bundles.matches | bundles= 45 | matches=22 | 48.800 %
132	textmatches/interchange.txt.ascii.fixed-23.txt.bundles.matches | bundles= 84 | matches=41 | 48.800 %
133	textmatches/interchange.txt.ascii.fixed-06.txt.bundles.matches | bundles= 70 | matches=34 | 48.500 %
134	textmatches/interchange.txt.ascii.fixed-01.txt.bundles.matches | bundles= 70 | matches=33 | 47.100 %
135	textmatches/interchange.txt.ascii.fixed-05.txt.bundles.matches | bundles= 94 | matches=44 | 46.800 %
136	textmatches/touchstone.txt.ascii.fixed-15.txt.bundles.matches | bundles= 121 | matches=53 | 43.800 %
137	textmatches/inside_out.txt.ascii.fixed-13.txt.bundles.matches | bundles= 150 | matches=51 | 34.000 %

Anexo 3: Texto número 18, do livro *Inside Out* (texto mais convergente do *corpus* de estudo).

000001 my_second_and (M)
000002 second_and_only (M)
000003 and_only_other (M)
000004 only_other_memory (M)
000005 other_memory_of (M)
000006 memory_of_llandaff (NM)
000007 of_llandaff_cathedral (M)
000008 llandaff_cathedral_school (M)
000009 cathedral_school_is (M)
000010 school_is_extremely (M)
000011 is_extremely_bizarre (M)
000012 extremely_bizarre_it (NM)
000013 bizarre_it_happened (NM)
000014 it_happened_a (M)
000015 happened_a_little (M)
000016 a_little_over (M)
000017 little_over_a (M)
000018 over_a_year (M)
000019 a_year_later (M)
000020 year_later_when (M)
000021 later_when_i (M)
000022 when_i_was (M)
000023 i_was_just (M)
000024 was_just_nine (M)

000025 just_nine_by (M)
000026 nine_by_then (NM)
000027 by_then_i (M)
000028 then_i_had (M)
000029 i_had_made (M)
000030 had_made_some (M)
000031 made_some_friends (M)
000032 some_friends_and (M)
000033 friends_and_when (M)
000034 and_when_i (M)
000035 when_i_walked (M)
000036 i_walked_to (M)
000037 walked_to_school (M)
000038 to_school_in (M)
000039 school_in_the (M)
000040 in_the_mornings (M)
000041 the_mornings_i (M)
000042 mornings_i_would (M)
000043 i_would_start (M)
000044 would_start_out (M)
000045 start_out_alone (M)
000046 out_alone_but (M)
000047 alone_but_would (M)
000048 but_would_pick (M)
000049 would_pick_up (M)
000050 pick_up_four (M)
000051 up_four_other (M)
000052 four_other_boys (M)
000053 other_boys_of (M)
000054 boys_of_my (M)
000055 of_my_own (M)
000056 my_own_age (M)
000057 own_age_along (NM)
000058 age_along_the (M)
000059 along_the_way (M)
000060 the_way_after (M)
000061 way_after_school (M)
000062 after_school_was (M)
000063 school_was_over (M)
000064 was_over_the (M)
000065 over_the_same (M)
000066 the_same_four (M)
000067 same_four_boys (NM)
000068 four_boys_and (M)
000069 boys_and_i (M)
000070 and_i_would (M)
000071 i_would_set (M)
000072 would_set_out (M)
000073 set_out_together (M)
000074 out_together_across (M)
000075 together_across_the (M)
000076 across_the_village (M)
000077 the_village_green (M)
000078 village_green_and (M)
000079 green_and_through (M)
000080 and_through_the (M)
000081 through_the_village (M)
000082 the_village_itself (M)
000083 village_itself_heading (NM)
000084 itself_heading_for (NM)

000085 heading_for_home (M)
000086 for_home_on (M)
000087 home_on_the (M)
000088 on_the_way (M)
000089 the_way_to (M)
000090 way_to_school (M)
000091 to_school_and (M)
000092 school_and_on (M)
000093 and_on_the (M)
000094 on_the_way (M)
000095 the_way_back (M)
000096 way_back_we (M)
000097 back_we_always (M)
000098 we_always_passed (M)
000099 always_passed_the (M)
000100 passed_the_sweet (M)
000101 the_sweet_shop (M)
000102 sweet_shop_no (NM)
000103 shop_no_we (NM)
000104 no_we_didnt (M)
000105 we_didnt_we (M)
000106 didnt_we_never (NM)
000107 we_never_passed (M)
000108 never_passed_it (M)
000109 passed_it_we (M)
000110 it_we_alwals (NM)
000111 we_alwals_stopped (NM)
000112 alwals_stopped_the (NM)
000113 stopped_the_sweet (NM)
000114 the_sweet_shop (M)
000115 sweet_shop_in (M)
000116 shop_in_llandaff (NM)
000117 in_llandaff_in (M)
000118 llandaff_in_the (M)
000119 in_the_year (M)
000121 year_1923_was (M)
000122 1923_was_the (M)
000123 was_the_very (M)
000124 the_very_centre (M)
000125 very_centre_of (M)
000126 centre_of_our (M)
000127 of_our_lives (M)
000128 our_lives_to (M)
000129 lives_to_us (M)
000130 to_us_it (M)
000131 us_it_was (M)
000132 it_was_what (M)
000133 was_what_a (M)
000134 what_a_bar (M)
000135 a_bar_is (M)
000136 bar_is_to (M)
000137 is_to_a (M)
000138 to_a_drunk (M)
000139 a_drunk_or (M)
000140 drunk_or_a (M)
000141 or_a_church (M)
000142 a_church_is (M)
000143 church_is_to (M)
000144 is_to_a (M)
000145 to_a_bishop (M)

000146 a_bishop_without (M)
000147 bishop_without_it (NM)
000148 without_it_there (M)
000149 it_there_would (M)
000150 there_would_have (M)
000151 would_have_been (M)
000152 have_been_little (M)
000153 been_little_to (M)
000154 little_to_live (M)
000155 to_live_for (M)
000156 live_for_but (M)
000157 for_but_it (M)
000158 but_it_had (M)
000159 it_had_one (M)
000160 had_one_terrible (M)
000161 one_terrible_drawback (NM)
000162 terrible_drawback_this (NM)
000163 drawback_this_sweet (NM)
000164 this_sweet_shop (M)
000165 sweet_shop_the (M)
000166 shop_the_woman (M)
000167 the_woman_who (M)
000168 woman_who_owned (M)
000169 who_owned_it (M)
000170 owned_it_was (M)
000171 it_was_a (M)
000172 was_a_horror (M)
000173 a_horror_we (M)
000174 horror_we_hated (NM)
000175 we_hated_her (M)
000176 hated_her_and (M)
000177 her_and_we (M)
000178 and_we_had (M)
000179 we_had_good (M)
000180 had_good_reason (M)
000181 good_reason_for (M)
000182 reason_for_doing (M)
000183 for_doing_so (M)

Anexo 4: Texto número 13, do livro *Inside Out* (texto menos convergente do *corpus* de estudo).

000001 i'm_not_crazy (NM)
000002 not_crazy_about (M)
000003 crazy_about_him (M)
000004 about_him_rose (NM)
000005 him_rose_what (NM)
000006 rose_what_do (M)
000007 what_do_you (M)
000008 do_you_think (M)
000009 you_think_of (M)
000010 think_of_jake (M)
000011 of_jake_meg (NM)
000012 jake_meg_he's (NM)
000013 meg_he's_all (NM)
000014 he's_all_right (NM)
000015 all_right_rose (NM)
000016 right_rose_you (NM)
000017 rose_you_don't (NM)
000018 you_don't_like (NM)

000019 don't_like_him (NM)
000020 like_him_do (M)
000021 him_do_you (M)
000022 do_you_meg (NM)
000023 you_meg_well (NM)
000024 meg_well_he (NM)
000025 well_he_was (M)
000026 he_was_unfriendly (M)
000027 was_unfriendly_rose (NM)
000028 unfriendly_rose_oh (NM)
000029 rose_oh_he's (NM)
000030 oh_he's_just (NM)
000031 he's_just_shy (NM)
000032 just_shy_that's (NM)
000033 shy_that's_all (NM)
000034 that's_all_meg (NM)
000035 all_meg_shy (NM)
000036 meg_shy_you (NM)
000037 shy_you_must (NM)
000038 you_must_be (M)
000039 must_be_joking (M)
000040 be_joking_five (NM)
000041 joking_five_minutes (NM)
000042 five_minutes_after (M)
000043 minutes_after_meeting (M)
000044 after_meeting_me (M)
000045 meeting_me_he (M)
000046 me_he_asked (M)
000047 he_asked_to (M)
000048 asked_to_borrow (M)
000049 to_borrow_five (M)
000050 borrow_five_dollars (M)
000051 five_dollars_that's (NM)
000052 dollars_that's_not (NM)
000053 that's_not_what (NM)
000054 not_what_i (M)
000055 what_i_call (M)
000056 i_call_shy (NM)
000057 call_shy_rose (NM)
000058 shy_rose_o (NM)
000059 rose_o_k (NM)
000060 o_k_that (NM)
000061 k_that_was (NM)
000062 that_was_rude (M)
000063 was_rude_but (M)
000064 rude_but_he's (NM)
000065 but_he's_broke (NM)
000066 he's_broke_meg (NM)
000067 broke_meg_well (NM)
000068 meg_well_i'm (NM)
000069 well_i'm_poor (NM)
000070 i'm_poor_myself (NM)
000071 poor_myself_and (M)
000072 myself_and_i'm (NM)
000073 and_i'm_trying (NM)
000074 i'm_trying_to (NM)
000075 trying_to_save (M)
000076 to_save_up (M)
000077 save_up_for (M)
000078 up_for_my (M)

000079 for_my_vacation (M)
000080 my_vacation_rose (NM)
000081 vacation_rose_all (NM)
000082 rose_all_right (NM)
000083 all_right_all (M)
000084 right_all_right (M)
000085 all_right_i'm (NM)
000086 right_i'm_sure (NM)
000087 i'm_sure_he (NM)
000088 sure_he_won't (NM)
000089 he_won't_ask (NM)
000090 won't_ask_you (NM)
000091 ask_you_again (M)
000092 you_again_he's (NM)
000093 again_he's_good (NM)
000094 he's_good_looking (NM)
000095 good_looking_though (M)
000096 looking_though_isnt (NM)
000097 though_isnt_he (M)
000098 isnt_he_meg (NM)
000099 he_meg_yes (NM)
000100 meg_yes_i (NM)
000101 yes_i_guess (M)
000102 i_guess_so (M)
000103 guess_so_but (M)
000104 so_but_he (M)
000105 but_he_knows (M)
000106 he_knows_it (M)
000107 knows_it_i (M)
000108 it_i_think (M)
000109 i_think_he's (NM)
000110 think_he's_really (NM)
000111 he's_really_bigheaded (NM)
000112 really_bigheaded_rose (NM)
000113 bigheaded_rose_you're (NM)
000114 rose_you're_just (NM)
000115 you're_just_jealous (NM)
000116 just_jealous_meg (NM)
000117 jealous_meg_no (NM)
000118 meg_no_i'm (NM)
000119 no_i'm_not (NM)
000120 i'm_not_i (NM)
000121 not_i_don't (NM)
000122 i_don't_want (NM)
000123 don't_want_him (NM)
000124 want_him_he's (NM)
000125 him_he's_mean (NM)
000126 he's_mean_bigheaded (NM)
000127 mean_bigheaded_and (NM)
000128 bigheaded_and_stupid (NM)
000129 and_stupid_rose (NM)
000130 stupid_rose_what (NM)
000131 rose_what_do (M)
000132 what_do_you (M)
000133 do_you_mean (M)
000134 you_mean_stupid (M)
000135 mean_stupid_you're (NM)
000136 stupid_you're_stupid (NM)
000137 you're_stupid_too (NM)
000138 stupid_too_meg (NM)

000139 too_meg_shut (NM)
000140 meg_shut_up (NM)
000141 shut_up_rose (NM)
000142 up_rose_no (NM)
000143 rose_no_you (NM)
000144 no_you_shut (M)
000145 you_shut_up (M)
000146 shut_up_meg (NM)
000147 up_meg_mom (NM)
000148 meg_mom_www (NM)
000149 mom_www_deadmike (NM)
000150 www_deadmike_com (NM)

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Administração
Baixar livros de Agronomia
Baixar livros de Arquitetura
Baixar livros de Artes
Baixar livros de Astronomia
Baixar livros de Biologia Geral
Baixar livros de Ciência da Computação
Baixar livros de Ciência da Informação
Baixar livros de Ciência Política
Baixar livros de Ciências da Saúde
Baixar livros de Comunicação
Baixar livros do Conselho Nacional de Educação - CNE
Baixar livros de Defesa civil
Baixar livros de Direito
Baixar livros de Direitos humanos
Baixar livros de Economia
Baixar livros de Economia Doméstica
Baixar livros de Educação
Baixar livros de Educação - Trânsito
Baixar livros de Educação Física
Baixar livros de Engenharia Aeroespacial
Baixar livros de Farmácia
Baixar livros de Filosofia
Baixar livros de Física
Baixar livros de Geociências
Baixar livros de Geografia
Baixar livros de História
Baixar livros de Línguas

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo