

ANÁLISE MULTIDIMENSIONAL
(Multidimensional Analysis)

Tony Berber SARDINHA
(Pontifícia Universidade Católica de São Paulo)

ABSTRACT: This article presents an overview of the approach introduced by Douglas Biber known as 'Multidimensional Analysis'. The article introduces the main concepts in this approach and reviews major studies using this methodology. A survey of unpublished studies carried out in Brazil is presented as well. The article also discusses the key criticisms levelled at multidimensional studies.

KEY WORDS: Multidimensional Analysis, Corpus Analysis, Text Analysis, Register

RESUMO: Este trabalho oferece uma retrospectiva da abordagem conhecida por 'Análise Multidimensional', iniciada por Douglas Biber. O trabalho apresenta os conceitos centrais relacionados a esta abordagem e resenha os estudos principais que a utilizaram. Trabalhos ainda inéditos, realizados no Brasil, também são resenhados. O artigo ainda inclui uma discussão acerca das principais críticas relacionadas à Análise Multidimensional.

PALAVRAS-CHAVE: Análise Multidimensional, Análise de Corpus, Análise Textual, Registro.

Introdução

O ano de 1998 marca dez anos do lançamento do livro 'Variation Across Speech and Writing' de Douglas Biber. Nesta obra, o autor apresenta para o grande público a abordagem conhecida por Análise Multi-traço e Multidimensional de Variação de Registro (Multi-feature Multi-dimensional Analysis of Register Variation), ou simplesmente Análise Multidimensional. É, portanto, recente, mas já informa um grande número de investigações, muitas delas publicadas em veículos de renome na linguística como os periódicos *Language* e *Text*. Além disso, está presente em várias coletâneas dedicadas a áreas diversas, como a Linguística do Corpus e a Retórica Contrastiva. Ela também informa a produção de uma nova gramática da língua ingle-

sa, cujo colaborador principal é Douglas Biber. A Análise Multidimensional se apresenta, então, como uma metodologia de peso na linguística atual, entretanto no Brasil ela ainda é virtualmente desconhecida. A Análise Multidimensional merece um relato em um periódico da língua portuguesa.

Para iniciar a apresentação da Análise Multidimensional, discutiremos primeiramente o que se considera ser sua maior contribuição para o estudo da linguagem: a possibilidade de se utilizar concomitantemente uma variedade de traços linguísticos empregados na análise textual e de se aplicar a codificação desses traços a um número de textos maior do que se poderia fazer manualmente (e.g. Hoey, 1983; Swales, 1990; van Dijk, 1980), por meio do emprego de computadores e técnicas estatísticas. E para finalizar a introdução, se fará uma breve exposição acerca do conceito fundamental deste tipo de análise, que é justamente o construto ‘dimensão’.

Um ponto de partida para o entendimento da Análise Multidimensional é compreender o contexto no qual ela foi criada. Basicamente, ela foi desenvolvida tendo em vista o fato indesejável de que, na visão de Biber, o estudo da variação entre textos ter sido feita com pouca abrangência, fazendo-se valer de poucas características ou traços linguísticos. Segundo ele, é comum descrever-se uma variedade textual por meio de poucas características, como a ‘formalidade’, ‘impessoalidade’ ou ‘oralidade’ (as quais ele define como ‘características situacionais’), ou como a frequência de palavras, emprego de marcadores, ou traços coesivos (as quais ele chama de ‘características linguísticas’).

Entretanto, para Biber a variação entre variedades linguísticas (gêneros, meios, períodos históricos, etc.) deve levar em conta uma maior quantidade de traços, já que nenhum traço isolado é suficiente para uma descrição adequada. Tampouco o estudo baseado em características situacionais somente, ou linguísticas apenas, é adequado. O ideal, segundo Biber, seria combinar-se as duas perspectivas, isto é, a descrição firmada em características situacionais com a descrição baseada em traços linguísticos. A Análise Multidimensional se propõe justamente a isso, ou seja, fornecer o instrumental para a identificação de padrões de co-ocorrências dos dois tipos de características, visando uma caracterização de uma língua, ou de um conjunto de tipos de texto, de modo abrangente.

O nome da abordagem deriva do conceito de ‘dimensão’ de variação. Uma dimensão é um conjunto de traços que subjazem a um *corpus*. O *corpus* pode consistir de uma seleção de textos, de um conjunto de gêneros ou até mesmo de amostras relativas a um idioma inteiro. Este conceito será discutido mais abaixo, na seção terminologia.

Pressupostos da Análise Multidimensional

A Análise Multidimensional possui um caráter essencialmente quantitativo e computacional. Ela permite a descrição de línguas e tipos de textos por meio de uma grande quantidade de características linguísticas.

Como dito acima, a análise Multidimensional foi criada por Douglas Biber com o objetivo de permitir uma descrição rica e complexa de *corpora* inteiros de textos por meio estatísticos bem como a extração precisa de características textuais em comum entre *corpora*. Anteriormente à Análise Multidimensional, a tendência era de que se estudasse a co-ocorrência de poucos traços e que se fizesse a interpretação de modo intuitivo. A variação entre registros era investigada comumente por meio de poucos parâmetros, por exemplo, ‘formalidade’ ou ‘planejamento’ e por conseguinte a distinção que se fazia entre textos era incompleta pois privilegiava apenas uma das muitas diferenças que podem existir entre os textos. Essa ênfase no emprego de poucos parâmetros também tinha o efeito de polarizar a descrição de cada parâmetro; assim, havia uma tendência para se descrever textos através de dois opostos, por exemplo, formal x informal, ou planejado x espontâneo.

A análise Multidimensional propõe alternativas para esse cenário. Em primeiro lugar, ela reconhece que a análise deve utilizar-se de uma quantidade maior de parâmetros para permitir uma comparação mais abrangente. Características como formalidade, planejamento, oralidade, referencialidade, entre outras, devem ser possíveis de serem incluídas. Além disso, o analista deve dispor de um arsenal que inclua o maior número possível de características linguísticas, já que o aumento da quantidade parâmetros implica em número maior de traços linguísticos necessários para cobrir a maior gama de parâmetros.

Segundo, as categorias a serem empregadas não necessitam ser limitadas. As categorias podem exibir uma gama maior de possibilidades de variação, por exemplo através de uma escala que vá, no caso da ‘formalidade’, de ‘mais formal’ para ‘menos formal’. Por fim, a descrição dos textos dentro destes aspectos mais abrangentes é feita por meio de técnicas estatísticas multivariadas, o que evita erros na especificação dos traços co-ocorrentes.

Para exemplificar os dois pontos acima, basta examinar algumas das categorias lingüísticas empregadas na descrição do inglês (Biber, 1995: 94): especialização verbal; subordinação; estruturas reduzidas; pronominalização; classes lexicais específicas; nominalização. Para ilustrar essas categorias, vide o trecho abaixo, parte de uma conversa informal à mesa de jantar:

- A: Well I must admit I feel [pause] I mean Edward’s mother and his great [pause] and his grandfather
 B: mm
 A: will come up on Christmas Day
 B: yes
 A: but I feel somehow (pause) the sheer fact of not having to have [pause] to have [pause] this [pause] really sort of [long pause] it’s for one thing it does nark me that
 B: [unintelligible]
 A: it’s so bloody expensive that he won’t eat anything except the largest most splendid pieces of meat you know
 B: [unintelligible] how annoying
 A: and mm it upsets me you know if he needed it I wouldn’t mind [pause]
 B: mm
 D: come to think of it he’s also he’s also [pause] an extremely greedy individual who
 A: yeah [pause] so that if you buy enough
 D: he isn’t satisfied thank you
 B: does anyone want ham¹
 (Biber, 1995: 145-147)

¹ A metodologia de transcrição adotada é simples, não incluindo aspectos acústicos. Normaliza-se a fala para a norma escrita, com exceção de comentários de pausa (‘[pause]’), trechos incompreensíveis (‘[unintelligible]’) e falante (‘A:’). Os textos provenientes de *corpora* já existentes (como o London-Lund, de inglês) seguem a convenção original do *corpus*, que inclui muitos outros elementos além desses mencionados aqui. Contudo, as referências de Análise Multidimensional não deixam claro até que ponto essa anotação original elaborada foi aproveitada ou não.

O trecho na página anterior inclui: na categoria especialização verbal, verbos ‘particulares’ (‘private’): ‘admit’, ‘feel’, etc; subordinação: apagamento de ‘that’ (‘I must admit [*that*] I feel [*that*]...’); estruturas reduzidas: contrações (it’s, he’s, isn’t); pronominalização: pronomes demonstrativos (‘I don’t like *this*’); classes lexicais específicas: ‘hedges’ (‘sort of’) e enfatizadores / amplificadores (‘really’, ‘extremely’); e nominalização: escassez de substantivos.

A abordagem Multidimensional possui várias características que no seu conjunto distinguem essa metodologia de outros sistemas analíticos de descrição. Primeiramente, ela baseia-se em *corpora*, isto é, ela pretende descrever um grande número de textos autênticos. Outra característica é que ela é essencialmente computacional, fazendo uso de ferramentas automáticas e semi-automáticas para rotulação das características de interesse nos textos. Além disso, ela se presta à descrição de conjuntos de textos ou registros, ao invés de textos individuais. Ela também tem um caráter essencialmente comparativo, pois promove o contraste entre os textos ou registros. Como diz seu rótulo, ela é multidimensional, ao reconhecer que a variação entre textos e registros pode ser mais adequadamente descrita por meio de múltiplos parâmetros. Uma outra marca característica é a utilização de um aparato quantitativo de descrição, o qual permite a especificação da co-ocorrência dos traços linguísticos de modo preciso. Porém, a abordagem não descarta a utilidade de técnicas qualitativas de interpretação, pois as dimensões são rotuladas seguindo a interpretação qualitativa dos fatores. Por fim, a abordagem Multidimensional combina análises de nível macro com análises de nível micro, já que a micro-descrição dos traços de cada texto permite a indução dos macro-agrupamentos textuais ou genéricos.

Outra característica importantíssima da abordagem Multidimensional que vale ressaltar é seu caráter cumulativo. A descrição de um *corpus* de uma certa variedade multidimensionalmente permite a comparação desta descrição com a descrição de outras variedades posteriormente, por exemplo, o contraste entre textos de negócio e textos jornalísticos. A comparação posterior é feita por meio do contraste das dimensões finais obtidas em cada análise separadamente, não havendo necessidade de se promover uma outra análise por completo para fins comparativos. Além da comparação de variedades diferentes, as análises individuais também se prestam para a comparação da descrição de *corpora* de línguas diferentes, por exemplo, a comparação das características multidimensionais de textos de ne-

gócios escritos em português e inglês. Desse modo, é possível empreender-se uma análise de larga escala de um *corpus* fazendo-se descrições individuais ao longo do tempo, combinando-se posteriormente as análises para fins comparativos. Por isso, a abordagem Multidimensional presta-se perfeitamente a projetos de descrição de bancos de dados em crescimento, ou seja, aquelas bases de dados linguísticos que estão em processo de coleta.

Por fim, a abordagem Multidimensional é flexível, podendo acomodar diversos tipos de traços linguísticos. Tradicionalmente, tem-se utilizado características lexicais e gramaticais (Biber 1988, 1995), entretanto é possível incluir características de cunho mais discursivo (Pacheco 1997) e funcionais (Shimazumi, 1998). Além disso, as próprias dimensões não são definitivas, podendo ser modificadas com a inclusão de novas características linguísticas.

Terminologia

Nesta seção são apresentados alguns conceitos centrais empregados na Análise Multidimensional.

Traços

São elementos linguísticos pertinentes à análise que se quantificam. Por exemplo, número de infinitivos, gerúndios e substantivos. Os traços a serem levados em conta em uma análise multidimensional são escolhidos mediante pesquisa na literatura disponível e devem representar um aspecto funcional no nível do texto (Biber 1988, p.72). Os traços são chamados de ‘variáveis’ quando da feitura da análise fatorial.

Características

Distingue-se entre dois tipos de características analíticas na Análise Multidimensional: linguísticas e não linguísticas (também conhecidas por ‘situacionais’). As características linguísticas são traços que se escolheu quantificar. Por exemplo, a quantidade de substantivos, ou de marcadores discursivos, ou a densidade lexical. As caracte-

terísticas situacionais são aquelas preexistentes e de uso não estritamente acadêmico que descrevem as características de uso de uma variedade. Por exemplo, propósito, tópico, meio, formalidade, etc.

Registro e gênero

Os termos são empregados com sentidos idênticos na Análise Multidimensional. Na Análise Multidimensional, registro ou gênero é um termo empregado para definir uma variedade linguística, geral ou específica, como discurso acadêmico ou cartas pessoais. Em seu trabalho mais antigo, Biber (1988) emprega o termo ‘gênero’ e mais recentemente tem privilegiado ‘registro’ (Biber, 1995).

Um registro ou gênero é uma variedade definida por variáveis situacionais, isto é, não linguísticas (vide tipo de texto, abaixo) e cujos rótulos são empregados corriqueiramente por falantes nativos da língua no dia a dia. Este é um conceito assumidamente impreciso (‘cover term’), podendo definir tanto variedades de caráter geral, como ‘conversaçoão’, quanto específicas, como por exemplo, os textos escritos por um certo autor. Assim, prosa acadêmica, conversaçoão espontânea e editoriais jornalísticos seriam três registros diferentes. Não é feita uma diferenciação formal entre ‘registro’ e ‘gênero’, ao contrário de outras áreas (e.g. linguística sistêmico-funcional). O termo ‘registro’ é o mais empregado.

Tipo de texto

O conceito de tipo de texto é diferente do de registro e gênero. Ele designa um conjunto de textos formado exclusivamente com base em critérios linguísticos. Chega-se à descrição de tipos de texto somente em estágios avançados da Análise Multidimensional, quando já se descreveram as dimensões e se mapearam os registros participantes em cada uma. Por isso, não é uma meta de toda análise multidimensional descrever os tipos de texto de uma língua ou variedade.

Fator

Um fator é um grupo de variáveis que co-ocorrem significativamente do ponto de vista estatístico. Os fatores são extraídos mediante

a aplicação do procedimento estatístico conhecido como ‘Análise Fatorial’. Este procedimento baseia-se na extração de conjuntos de variáveis que se correlacionam. O procedimento estatístico retorna várias ‘soluções’, isto é, conjuntos com vários números de fatores (por exemplo, uma solução com dois fatores, três fatores, etc). Fica a cargo do analista escolher a solução mais apropriada. Não há métodos estatísticos comprovadamente eficazes para a identificação do número exato de fatores subjacentes a um conjunto de dados. Por isso, o analista deve recorrer a um rol de técnicas para aferir qual seria a melhor solução para os seus dados. Da criteriosa identificação do número de fatores depende a correta localização das dimensões.

Dimensão

Dimensão é o estatus que um fator assume assim que ele é interpretado do ponto de vista de sua função comunicativa. Uma dimensão permite visualizar características em comum partilhadas por uma porção significativa dos dados. A interpretação do fator leva em conta tanto as características linguísticas quanto as características partilhadas pelos registros que estão representados no fator. As dimensões permitem redefinir o quadro de registros inicial. Este conceito ficará mais claro a partir dos exemplos que se oferecerá mais abaixo.

Etapas na feitura de uma Análise Multidimensional

Agora que foram definidos os conceitos centrais da Análise Multidimensional, torna-se possível fazer-se uma breve apresentação de quais são os passos envolvidos na consecução de uma investigação dentro desta linha. O propósito aqui é meramente situar o leitor dentro dos passos mais gerais, para que ele possa seguir a discussão a seguir com mais subsídios. Obviamente, os interessados em levar a cabo uma Análise Multidimensional devem consultar fontes mais específicas.

De modo geral, a execução de uma análise multidimensional envolve três etapas básicas. A primeira, de caráter preliminar, compreende a revisão da literatura em busca de traços linguísticos relevantes a serem levados em conta na análise, a coleta do *corpus* e a codificação dos textos de acordo com o elenco de características linguísticas

selecionadas para análise. A segunda fase da análise refere-se à análise fatorial, durante a qual se faz um agrupamento das características linguísticas em fatores bem como a interpretação destes fatores funcionalmente a fim de se descobrir um traço comunicativo dominante subjacente ao fator, dando origem às dimensões. Por fim, na terceira etapa faz-se o cálculo de escores de cada texto em relação a cada fator e interpreta-se as dimensões à luz dos textos que as compõem.

Mais especificamente, os passos a serem seguidos numa análise multidimensional são apresentado abaixo. A listagem não é exaustiva, mas compreende as etapas principais:

1. Levantamento das características linguísticas relevantes para análise por meio de ampla consulta à literatura disponível.
2. Coleta ou adoção de um *corpus* de dados linguísticos representativo e compatível com as metas da análise.
3. Transformação das características linguísticas em variáveis quantificáveis.
4. Codificação dos dados baseada nas variáveis selecionadas, usando-se ferramentas computacionais para análise automática, semi-automática (interativa), ou manual.
5. Conferência manual da codificação feita por computador para se assegurar de sua exatidão.
6. Computação de frequência médias de cada variável.
7. Padronização das frequências (em geral por 1000 palavras), para permitir a comparação entre variedades (textos, registros ou *corpora*) de extensões diferentes.
8. Análise Fatorial inicial, a fim de se obter os pesos ('loadings') de cada variável em cada variedade.
9. Determinação do número de fatores, por meio da aplicação de técnicas como observação dos valores eigen ('eigenvalues') em um gráfico 'scree' ('scree plot').
10. Análise Fatorial posterior, fazendo-se a rotação dos fatores para se evitar sobrecarga de um fator.
11. Interpretação de cada fator e rotulação das dimensões.
12. Cálculo de escores de cada texto por fator, através da padronização dos escores com base na média e desvio padrão.
13. Cálculo de escores médios de cada variedade por fator.

Em resumo, uma pesquisa multidimensional envolve análises macroscópicas e microscópicas. As primeiras são efetuadas quando

da computação dos fatores. Neste momento, as várias análises de cada texto são agrupadas de modo que se possa perceber a variação das mesmas em nível macro. As análises microscópicas, por outro lado, se dão quando da interpretação dos fatores de modo funcional. Neste nível leva-se em conta cada texto e cada registro individualmente.

Descrição Multidimensional da língua inglesa

O trabalho pioneiro de Douglas Biber se baseou na descrição das dimensões subjacentes à língua inglesa. Para tanto, inicialmente ele selecionou um *corpus* de textos que representasse a variedade de registros encontrada no inglês. Os *corpora* escolhidos foram o LOB, de textos escritos em inglês britânico e o London-Lund, de transcrições de eventos falados, também da variedade britânica. Foram retiradas porções destes *corpora* e adicionados outros dois registros (variedades de cartas) e obteve-se um total de 481 textos, somando 960 mil palavras.

A seguir, foi feito um levantamento das principais características que, segundo a literatura existente na época, seriam relevantes para a descrição da língua inglesa. Foram elencadas 67 características, todas de cunho lexical e estrutural (vide seção ‘críticas’ abaixo). Todos os 481 textos foram codificados levando-se em conta estas variáveis. A codificação foi feita tanto manual quanto automaticamente. Para a análise manual foram desenvolvidos programas de computador interativos para facilitar a codificação.

Em seguida, partiu-se para a Análise Fatorial, a qual identificou sete fatores como sendo a melhor solução. Fez-se então o mapeamento de quais textos estavam presentes em cada fator. Os fatores foram inspecionados um por um e decidiu-se eliminar o sétimo fator porque era composto de variáveis cujo peso era maior em outros fatores. Cada fator era composto de variáveis que ocorriam positiva e negativamente. As primeiras são aquelas que tendiam a co-ocorrer, enquanto as últimas são aquelas que tendiam a não ocorrer caso as primeiras estivessem presentes. Por exemplo, as variáveis positivas de maior peso do primeiro fator são: verbos ‘particulares’ (‘private verbs’, e.g. *doubt*, *forget*, *guess*), apagamento de ‘that’ e contrações. E as principais negativas são: substantivos, palavras longas e posições. Desse modo,

nos textos em que ocorriam verbos ‘particulares’, e apagamento de ‘that’, havia uma tendência de aparecimento também de contrações. Por outro lado, nos textos em que havia verbos ‘particulares’, apagamento de ‘that’ e contrações, havia uma tendência de escassez de substantivos, palavras longas e proposições.

A seguir, o conjunto de características linguísticas de cada fator foi interpretado funcionalmente, levando ao estabelecimento das dimensões. Isto permitiu que se fizesse a rotulação dos fatores de acordo com a função comunicativa que as características linguísticas pareciam exprimir. Os rótulos indicam uma polaridade entre dois traços comunicativos complementares. Por exemplo, em relação ao fator 1, decidiu-se que as características linguísticas com peso positivo (vide tabela 1) tinham subjacente um traço comunicativo que indicava ‘produção interativa’. Já o conjunto de características com peso negativo revelavam um traço comunicativo comum que se poderia chamar de ‘produção informacional’. Por isso, o rótulo adotado para a dimensão 1 foi ‘produção com interação ou produção informacional’. O mesmo procedimento foi levado a cabo em relação aos outros fatores.

Mas embora as dimensões tenham rótulos que indiquem uma polaridade entre duas características comunicativas básicas, as dimensões não são dicotômicas. Pelo contrário, elas exprimem um contínuo. Em cada dimensão, os textos podem se situar ao longo de uma escala que vai de ‘mais’ a ‘menos’ em relação a cada traço comunicativo. Assim, na dimensão 1 pode-se encontrar textos que se situem em qualquer posição ao longo da escala entre ‘mais interativo’ e ‘mais informacional’ (ou ‘menos informacional’ e ‘menos interativo’).

O resultado da interpretação e nomeação dos fatores foi a identificação de seis dimensões. As seis dimensões apontadas foram:

1. produção com interação vs. produção informacional;
2. preocupações narrativas vs. não-narrativas;
3. referências explícitas vs. referências dependente do contexto;
4. expressão explícita de persuasão vs. não-explícita;
5. informação abstrata vs. não-abstrata;
6. elaboração informacional ‘on-line’².

² Termos traduzidos por Pacheco (1997: 66).

**Tabela 1: Peso das variáveis que compõem o fator 1
– as de maior peso aparecem primeiro**

Peso	
POSITIVO	NEGATIVO
Verbos pessoais	Substantivos
Apagamento de 'that'	Tamanho das palavras
Contrações	Preposições
Verbos no tempo presente	Razão forma/palavra
Pronomes pessoais da 2ª pessoa	Adjetivos atributivos
'Do' como verbo auxiliar	
Negação analítica	
Pronomes demonstrativos	
Enfatizadores em geral	
Pronomes de 1ª pessoa	
Pronome 'it'	
'Be' como verbo principal	
Subordinação causativa	
Partículas discursivas	
Pronomes indefinidos	
'Hedges' em geral	
Amplificadores	
Pronomes relativos	
Perguntas WH	
Modais de possibilidade	
Coordenação não-frasal	
Orações WH	
Preposições finais	

Fez-se, então, a computação dos escores de cada texto em cada dimensão. Os escores consistiam de somas relativas às quantidades das variáveis existentes em cada fator³. Desse modo, cada texto possuía um valor que indicava sua participação em cada dimensão. Depois fez-se o cálculo dos escores médios para cada registro em cada

³ Para exemplificar o método de cálculo, tomemos o fator 2. Este fator inclui como variáveis de peso positivo as seguintes características: verbos no tempo passado, verbos no aspecto perfeito, pronomes pessoais de terceira pessoa, verbos 'públicos', orações reduzidas, e negações sintéticas. Supondo-se que um dos textos tenha a seguinte contagem destas características: 113 verbos no tempo passado, 124 verbos no aspecto perfeito, 30 pronomes pessoais de terceira pessoa, 14 verbos 'públicos', 5 orações reduzidas, e 3 negações sintéticas, seu escore no fator 2 seria 289, isto é, a soma de $113 + 124 + 30 + 14 + 5 + 3$. Na verdade, a computação dos escores não foi feita por meio das contagens brutas, mas sim através de contagens padronizadas com base na média e desvio padrão, a fim de se evitar que o tamanho diferente dos textos influísse nos escores. Estes valores padronizados podem assumir valores negativos, pois indicam quanto acima ou abaixo da média cada valor está. Por isso, os escores dos textos podem ser negativos também.

dimensão⁴. Os escores médios de cada registro nas seis dimensões é apresentado no anexo. Abaixo é apresentado o elenco de registros ordenados de acordo com seus escores em cada dimensão; os registros mais acima são os que obtiveram escore mais alto.

Tabela 2: Ordenação descendente dos registros de acordo com seus escores médios nas seis dimensões

Dimensões					
1	2	3	4	5	6
Tel	Rom	Doc	Cpr	Aca	PPr
Con	Mis	Cpr	Edi	Doc	Ent
Cpe	Fic	Rjo	Rom	Rel	Pal
Pal	Sfi	Aca	Pas	Pas	Cpr
Ent	Ave	Rel	Cpe	Rjo	Edi
Rom	Bio	Pop	Ent	Jor	Rel
PPr	Pal	Edi	Fic	Cpr	Aca
Ave	Hum	Bio	Tel	Edi	Con
Mis	PPr	Pal	PPr	Pop	Bio
Fic	Jor	PPr	Pal	Hum	Pas
Cpr	Cpe	Pas	Rel	Bio	Pop
RTV	Pop	Jor	Doc	RTV	Doc
Sfi	Con	Ent	Pop	PPr	Jor
Rel	Rel	Hum	Hum	Ent	Tel
Hum	Edi	Sfi	Con	Fic	Rjo
Pop	Ent	Fic	Aca	Sfi	Rom
Edi	Rjo	Mis	Bio	Ave	RTV
Pas	Tel	Cpe	Jor	Pal	Cpe
Bio	Cpr	Ave	Sfi	Cpe	Hum
Rjo	Aca	Con	Mis	Mis	Fic
Aca	Pas	Rom	Ave	Rom	Sfi
Jor	Doc	Tel	Rjo	Con	Ave
Doc	RTV	RTV	RTV	Tel	Mis

Legenda: Bio: Biografias; Cpe: Cartas pessoais; Cpr: Cartas profissionais; Con: Conversação face a face; Tel: Conversação telefônica; Pop: Cultura popular; Doc: Documentos oficiais; Edi: Editoriais jornalísticos; Ent: Entrevistas; Sfi: Ficção científica; Ave: Ficção de aventura; Mis: Ficção de mistério; Fic: Ficção geral; Rom: Ficção romântica; Hum: Humor; Pal: Palestras espontâneas; PPr: Palestras preparadas; Pas: Passatempos; Aca: Prosa acadêmica; RTV: Rádio e TV; Rel: Religião; Jor: Reportagem jornalística; Rjo: Resenhas jornalísticas.

⁴ Esse cálculo é feito por uma média aritmética. Por exemplo, se houvesse três textos de um registro específico na dimensão 2, e eles tivessem os escores 16, 12 e 11, somariam-se os três valores, o que resultaria em 39, e dividiria-se este total por 3, o que daria 13. O valor 13 seria então o escore médio deste registro na dimensão 2. É possível ter-se escores médios negativos. Isto acontece quando há uma maioria de escores negativos de cada texto individual.

A ordenação dos registros de acordo com seus escores médios em cada dimensão permite que se trace paralelos entre eles. Conforme dito antes, as dimensões são caracterizadas por uma polaridade. Os registros com escores positivos representam melhor um dos pólos, enquanto os registros com os maiores escores negativos ilustram o outro pólo:

Dimensão 1 – produção com interação vs. produção informacional: os registros que melhor representam o modo de produção com interação são as conversas, tanto ao telefone quanto cara a cara; os registros que melhor representam a produção informacional são documentos oficiais, reportagem jornalística e prosa acadêmica.

Dimensão 2 – preocupações narrativas vs. não-narrativas: os registros que melhor demonstram uma preocupação com a narração são os registros de ficção, enquanto que os que melhor exprimem uma orientação não narrativa são os registros de rádio e TV, passatempos e documentos oficiais.

Dimensão 3 – referências explícitas vs. referências dependente do contexto: os registros que apresentam referência explícita em maior grau são documentos oficiais, cartas profissionais, resenhas jornalísticas e prosa acadêmica. Já os registros de rádio e TV, conversas telefônicas e cara a cara e ficção romântica exprimem referência dependente da situação.

Dimensão 4 – expressão explícita de persuasão vs. não-explícita: os registros de caráter mais persuasivo são as cartas profissionais, os editoriais e a ficção romântica. Por outro lado, os registros nos quais a persuasão é menos explícita são os de rádio e TV, resenhas jornalísticas e ficção de aventura.

Dimensão 5 – informação abstrata vs. não-abstrata: os registros que veiculam informação mais abstrata são os acadêmicos, os documentos oficiais e os religiosos. Já as conversas telefônicas, face a face e ficção romântica apresentam informação menos abstrata.

Dimensão 6 – elaboração informacional 'on-line': os registros nos quais a elaboração da informação é mais imediata são palestras preparadas, entrevistas e palestras espontâneas, enquanto que os registros nos quais a informação é elaborada de antemão são os de ficção (mistério, aventura, científica e geral.)

As dimensões mostram uma inter-relação entre registros escritos e falados. Alguns registros escritos possuem características em comum com registros falados e vice-versa. Por exemplo, de acordo com a dimensão 1, cartas pessoais, palestras espontâneas e entrevistas possuem como característica comum o fato de serem produzidas com interação entre escritor ou falante de um lado e leitor ou ouvinte do outro.

Apesar disso, persiste uma diferenciação básica entre os registros falados e escritos na metade das dimensões. Nas dimensões 1, 3 e 5, os registros escritos ocupam majoritariamente um dos pólos e os registros falados o outro. O registro que predomina no pólo onde se concentram os textos escritos é a escrita acadêmica. Já os registros que se concentram no pólo falado destas dimensões são os conversacionais.

Em conclusão, a grande contribuição da Análise Multidimensional para a descrição textual da língua inglesa é permitir uma visão precisa da variação entre textos falados e escritos. A classificação foi obtida com base na presença, ausência e co-ocorrência de traços linguísticos. Até então, as inter-relações entre registros escritos e falados eram assumidas, mas não haviam sido descritas formalmente com tanto rigor e abrangência. Os esquemas classificatórios existentes baseados em categorias restritas como ‘fala’ e ‘escrita’ ou ‘oral’ e ‘letrado’. Uma análise multidimensional deixa claro quanto de ‘fala’ e ‘escrita’, por exemplo, um conjunto de textos possui, com base no exame de *corpora* inteiros de textos. A Análise Multidimensional também facilita a descoberta de similaridades entre gêneros e permite a descrição precisa dos elementos que contribuem para essa similaridade. Uma outra contribuição significativa da Análise Multidimensional é permitir a comparação rica e precisa de registros entre línguas diferentes, o que é discutido na seção a seguir.

Análise Multidimensional de outras línguas

O mesmo arcabouço empregado para a descrição multidimensional do inglês foi aplicado a uma série de outras línguas. Até o presente, foram descritas multidimensionalmente por Biber e outros pesquisadores os seguintes idiomas: nukulaelae tuvalan, coreano e somali.

Para a descrição dessas línguas usou-se um conjunto variado de *corpora*. A quantidade de traços linguísticos, registros e de textos em cada registro também variou consideravelmente. A tabela a seguir resume os elementos centrais de cada *corpus*.

Tabela 3: Dimensões dos *corpora*

Língua	Traços linguísticos	Registros (escritos / falados)	Textos	Textos por registro	Total de palavras
Inglês	67	23 (17/6)	481	6 a 80	960.000
Nukulaelae Tuvaluan	6	7 (2/5)	222	12 a 70	152.771
Coreano	42	22 (12/10)	150	5 a 10	135.500
Somali	58	33 (23/10)	604	3 a 49	600.000

A segunda língua a ser descrita por meio de dimensões foi o nukulaelae tuvalan. Os falantes de nukulaelae estão inseridos em uma cultura bastante distinta da européia. Esta é uma língua indígena que não possui discursos institucionalizados na esfera pública de conteúdo não religioso, sejam escritos ou falados. O emprego da Análise Multidimensional para sua descrição mostra que esta metodologia pode ser empregada para se conhecer línguas pouco documentadas.

O *corpus* usado para a descrição do nukulaelae tuvalan consistiu de pouco mais de 150 mil palavras, compreendendo 222 textos de sete registros diferentes. As três dimensões delineadas são apresentadas na tabela abaixo, juntamente com os registros mais característicos de cada uma.

Tabela 4: Dimensões do *corpus* de nukulaelae tuvalan

Dimensão	Registros mais característicos	
Discurso atitudinal ou abalizado	Atitudinal: Discursos em contextos particulares e comícios	Abalizado: Sermões escritos e cartas pessoais
Referência interpessoal ou informacional	Interpessoal: Cartas pessoais e conversações	Informacional: Programas de rádio e TV e sermões escritos
Construção textual em grupo ou monológica	Em Grupo: Conversação, sermões escritos	Monológica: Cartas pessoais e programas de rádio e TV

A terceira língua cuja descrição multidimensional foi publicada é o coreano. O *corpus* empregado possuía cerca de 135 mil palavras, incluindo 22 registros. As seis dimensões extraídas aparecem na tabela a seguir.

Tabela 5: Dimensões do *corpus* de coreano

Dimensão	Registros mais característicos	
Interação informal ou elaboração explícita	Interação: Conversas particulares e dramaturgia televisiva	Elaboração: Crítica literária, livros didáticos para faculdade
Coesão explícita ou implícita	Explícita: Contos e conversação	Implícita: Documentos legais e oficiais e notícias de rádio e TV
Expressão explícita de posicionamento interpessoal	Mais explícita: Dramaturgia televisiva e conversas particulares	Menos explícita: Reportagem jornalística e documentos legais e oficiais
Discurso narrativo ou não-narrativo	Narrativo: Dramaturgia televisiva e contos	Não-narrativo: Documentos legais e oficiais e crítica literária
Relato presente ('online') de eventos	Mais presente: Transmissões esportivas e discursos públicos preparados	Menos presente: Reportagem jornalística e notícias de rádio e TV
Honorificação	Mais honorífico: Cartas pessoais e conversas em público	Menos honorífico: Documentos legais e oficiais e crítica literária

A quarta língua descrita multidimensionalmente foi o somali. O *corpus* que serviu de base para a descrição possuía 33 registros, o que somava por volta de 600 mil palavras, bem mais do que as outras línguas (com exceção do inglês). À semelhança do coreano, foram extraídas seis dimensões, as quais aparecem na tabela a seguir.

Tabela 6: Dimensões do *corpus* de somali

Dimensão	Registros mais característicos	
Elaboração estrutural: envolvimento ou exposição	Envolvimento: conversações e reuniões de família	Exposição: Introduções de livros e editoriais
Elaboração lexical: produção em tempo real ('on-line') ou planejada	Em tempo real: Transmissões esportivas e palestras universitárias	Planejada: Discursos políticos publicados e editoriais
Apresentação argumentativa ou relatada de eventos	Argumentativa: reuniões de família e reuniões formais	Relatada: Reportagens jornalísticas e histórias populares
Organização discursiva narrativa ou não-narrativa	Narrativa: Histórias populares e histórias seriadas	Petições e anúncios
Interação distanciada e diretiva	Mais distanciada e diretiva: Cartas pessoais e reuniões de família	Menos distanciada e diretiva: Reportagem jornalística e transmissões esportivas
Persuasão pessoal	Mais persuasão: Petições e cartas pessoais	Menos persuasão: reportagem jornalística e transmissões esportivas

As dimensões de cada língua, quando comparadas, revelam muitas semelhanças. Todas exibem dimensões relacionadas com os seguintes traços comunicativos: oralidade e letramento, interatividade, circunstâncias de produção, posicionamento pessoal e narração. Há também diferenças entre os quatro idiomas. Por exemplo, a noção comunicativa de estilo abstrato, a qual só é exibida no inglês e a honorificação, exclusiva do coreano.

Tipos textuais da língua inglesa

Conforme dito antes, é possível também obter-se uma tipologia textual a partir de uma análise multidimensional. Isto é feito com base em agrupamentos de textos cujos escores sejam similares. Para se obter estes conjuntos, utiliza-se 'cluster analysis', uma técnica estatística voltada para o agrupamento de dados.

Biber (1989) obteve uma tipologia textual da língua inglesa a partir de uma análise multidimensional. Os *corpora* utilizados foram os mesmos da análise apresentada acima, mas há algumas diferenças entre os dados utilizados para esta tipologia e aqueles empregados para a análise multidimensional apresentada acima. Em primeiro lugar, a categorização dos registros foi mais detalhada. Houve uma maior discriminação dos registros. Embora os textos sejam os mesmos, para a tipologia, Biber empregou uma classificação que levou em conta um número maior de registros do que anteriormente. Em segundo lugar, a análise dimensional que serviu como base para a tipologia incluía 5 dimensões e não 6. Isto pode ter ocorrido devido à classificação genérica mais detalhada, ou à mudanças nos procedimentos de determinação do número de fatores.

A tipologia delimitou oito tipos de texto:

1. Interação interpessoal
2. Interação informacional
3. Exposição científica
4. Exposição culta
5. Narrativa imaginativa
6. Exposição narrativa em geral
7. Reportagem
8. Persuasão

Cada um destes tipos de texto é composto de vários registros. A composição de cada tipo textual define o rótulo que cada um recebeu. Por exemplo, no tipo 1, os registros constantes são conversas face a face, conversas telefônicas entre amigos (62%), entre estranhos (17%) e de negócios (13%). Deste modo, este agrupamento parece exprimir um tipo de texto cuja característica comum é a expressão da interação pessoal.

A quantidade de cada registro nos vários tipos textuais varia. No caso ainda do tipo textual 1, as percentagens de cada um dos registros são as seguintes: conversas face a face, 29%, conversas telefônicas entre amigos, 62%, entre estranhos, 17% e conversas telefônicas de negócios 13%. Por isso, o registro que mais contribui para a caracterização do agrupamento como um tipo textual onde predomina a interação interpessoal são as conversas telefônicas, já a maioria dos textos deste registro se encontra neste tipo de texto.

Em alguns tipos, há uma predominância de um ou mais registros. A tabela abaixo mostra quais os registros cuja maioria dos seus textos aparecem em cada tipo textual. Apenas os tipos exposição científica e persuasão (números 3 e 8) não possuem registros com mais de 50% de ocorrência. Em exposição científica, o registro predominante é prosa acadêmica e em persuasão, são discursos espontâneos, ambos com 44%.

Tabela 7: Tipos textuais do inglês

Tipo de texto	Registros majoritários
1 Interação interpessoal	Conversas telefônicas entre amigos
2 Interação informacional	Conversas face a face, entrevistas, palestras espontâneas, cartas pessoais e conversas telefônicas entre desconhecidos e de negócios
3 Exposição científica	-
4 Exposição culta	Documentos oficiais
5 Narrativa imaginativa	Ficção (romance, geral, mistério, aventura) e discursos preparados
6 Exposição narrativa geral	Reportagem, editoriais jornalísticos, biografias, humorísticos, religião, transmissões, ficção científica
7 Reportagem	Transmissões esportivas de rádio e TV
8 Persuasão	-

Esta tipologia textual tem duas características que a distinguem das demais. Primeiramente, é verdadeiramente indutiva. Ela se baseia na presença de características linguísticas e não em categorizações

abstratas. Os textos são analisados linguisticamente, agrupados em dimensões e depois reagrupados de acordo com similaridades entre eles em relação às dimensões. E em segundo lugar, a tipologia é baseada em dados autênticos e extensos. Os textos fazem parte de *corpora* que reúnem um grande número de textos de uma ampla gama de registros.

Descrição diacrônica

A Análise Multidimensional também foi empregada para se descrever a variação entre registros ao longo do tempo. Biber e Finegan (1989) descreveram a variação diacrônica de três registros da língua inglesa (ensaios, ficção e cartas), do século XVI até o presente. Mais tarde, Biber (1996) apresentou resultados de uma descrição mais ampliada do inglês (incluindo sete registros) e uma análise similar do somali.

A metodologia exata usada para a identificação das dimensões diacrônicas de variação não está explicitada. Entretanto, percebe-se que o ponto de partida foi diferente. Em vez de se iniciar com as características compartilhadas de cada texto e partir-se então para o agrupamento destas características em dimensões, iniciou-se com as dimensões extraídas do estudo sincrônico do inglês e ‘encaixou-se’ os textos históricos nestas dimensões. Em outras, palavras, comparou-se os textos históricos com as dimensões preexistentes relativas à descrição do inglês contemporâneo.

O achado geral, no caso do inglês, refere-se ao movimento dos registros na direção da oralidade. Nos séculos XIX e XX houve uma intensificação dos traços caracterizadores da oralidade de ensaios, ficção e cartas (Biber e Finegan 1989). A razão da mudança em direção à oralidade seria, segundo Biber e Finegan (1989), o aumento do público leitor das classes média e trabalhadora. A escrita teria se adaptado às exigências de um público com menos familiaridade com contextos estritamente letrados e que exibia uma ampla gama de interesses e competências linguísticas.

Há ainda outros estudos utilizando a metodologia da Análise Multidimensional no estudo diacrônico da variação. Atkinson (1992), por exemplo, investigou a variação de um registro específico ao lon-

go do tempo, nomeadamente os artigos de pesquisa de medicina publicados no periódico 'Edinburgh Medical Journal' e Atkinson (1993) descreveu as mudanças no 'Philosophical Transactions of the Royal Society of London' desde o século XVII.

Análise da escrita de aprendizes

Uma outra área da descrição textual em que a Análise Multidimensional tem sido empregada é o estudo da escrita de alunos de línguas estrangeiras. Estes estudos em geral são conduzidos sob a égide da retórica contrastiva.

Os trabalhos pioneiros na área da retórica contrastiva por meio de Análise Multidimensional foram desenvolvidos por William Grabe, Dantas-Whitney, Ulla Connor e pelo próprio Douglas Biber (vide discussão em Connor, 1996: 159-160). Há, no entanto, estudos feitos no Brasil sobre contrastes multidimensionais focalizando a escrita de alunos de inglês como língua estrangeira. A importância destes estudos vai além de seus achados. Sua relevância também reside em grande parte no fato de introduzirem e divulgarem no país a metodologia da Análise Multidimensional. Por isso, no restante desta seção serão apresentados em maior detalhe dois estudos multidimensionais envolvendo a escrita de aprendizes.

Um destes trabalhos é Pacheco (1997). Esta pesquisa recente está voltada ao contraste da escrita de falantes nativos e não-nativos por meio de um *corpus* expressivo de 270 textos, composto por 90 textos em inglês como L1, 90 em inglês como língua estrangeira e 90 em português como L1. Os textos em inglês como L1 foram escritos por alunos americanos e os demais por estudantes brasileiros. No estudo, controlou-se os temas das composições.

Foram codificados 38 traços linguísticos, agrupados em 15 variáveis. Estas variáveis refletiam 5 redes funcionais, quais sejam: representação do contexto situacional no texto; relação escritor/leitor e escritor/texto; padrões de desenvolvimento retórico no texto; conexões entre orações, entre períodos e partes do texto; e complexidade do texto e de suas unidades. A análise multidimensional revelou a presença de 5 fatores, nomeadamente: estilo textual reduzido ou ela-

borado, orientação discursiva interacional ou informacional, explicitação do contexto situacional e cultural, lógica coesiva e posicionamento pessoal, orientação acadêmica estrutural e informacional.

Os escores médios dos três grupos de textos em cada fator revelou diferenças entre os grupos. As composições escritas em inglês L1 tiveram maiores escores nas dimensões 1 e 5, 'estilo textual reduzido ou elaborado' e 'orientação acadêmica estrutural e informacional'. Já as composições de alunos de inglês como língua estrangeira alcançaram maiores escores nas dimensões 2, 3, 4: 'orientação discursiva interacional ou informacional', 'explicitação do contexto situacional e cultural' e 'lógica coesiva e posicionamento pessoal'. As composições em português como língua nativa não obtiveram escores maiores em nenhuma dimensão. Mas estas composições mostram um distanciamento maior em relação ao inglês como L1.

Os resultados indicaram que as composições em inglês como língua nativa privilegiam a organização retórica, enquanto os textos em inglês como língua estrangeira colocam em destaque aspectos interacionais. Os autores brasileiros produziram textos em inglês com características híbridas, incluindo tanto traços típicos do inglês como língua nativa (e.g. interação e coesão textual) quanto do português (e.g. elaboração textual e pouca elaboração acadêmica). Em geral, os resultados sugerem haver uma influência cultural na escrita.

Uma peculiaridade técnica da pesquisa de Pacheco (1997) é que as variáveis levadas em conta para a formação dos fatores eram na verdade grupos de traços. Por exemplo, o fator 2 é composto de 4 variáveis ('expressões interacionais', 'perguntas retóricas', 'reflexão pessoal' e 'padrões linguísticos de problema-solução'), mas na verdade elas representam 13 traços. A dúvida que fica é se os traços agrupados em cada variável realmente co-ocorrem.

Um outro estudo que contrastou a escrita de falantes nativos e não-nativos é Shimazumi (1998). Foram contrastados 30 textos: 10 de não-nativos, alunos brasileiros de inglês como língua estrangeira; e 20 escritos por nativos, subdivididos em 10 produzidos por alunos britânicos e 10 por jornalistas britânicos. Assim como em Pacheco (1997), os textos foram controlados por tópicos.

Uma diferença importante deste estudo para com os demais é que a análise se baseou em traços sistêmico-funcionais. Foram escolhidos 13 traços pertencentes às três metafunções sistêmico-funcionais: transitividade, modo e tema e ainda uma referente ao nível do registro (nominalização).

A vantagem deste arcabouço teórico para a Análise Multidimensional é que o aparato analítico é semântico e se aplica naturalmente à análise textual, ao contrário das categorias contempladas por Biber, as quais se referem à estrutura oracional (ou intra-frasal). Deste modo, o estudo de Shimazumi (1998) leva em conta as críticas feitas à falta de validade do aparato analítico empregado por Biber (vide seção abaixo).

A análise sugeriu três dimensões: ‘expressão de organização textual ou densidade de informação’, ‘expressão de posicionamento interpessoal’ e ‘relato de eventos’. Os três grupos de textos tiveram escores maiores em uma dimensão diferente. Os textos escritos por alunos nativos tiveram um escore médio maior na dimensão ‘expressão de organização textual ou densidade de informação’. Isto indica que estes textos privilegiaram marcadores de organização textual, como temas textuais, adjuntos textuais e modalização. Os textos dos alunos não-nativos, por sua vez, tiveram maior escore na dimensão ‘expressão de posicionamento interpessoal’. Estes textos tiveram uma preponderância de traços como temas interpessoais e adjuntos modais. Por fim, os textos escritos pelos jornalistas nativos se destacaram na dimensão ‘relato de eventos’, incluindo traços como processos materiais e verbais.

Os resultados sugerem uma influência cultural e didática entre os alunos nativos e não-nativos. Estes últimos demonstraram uma tendência para incluir expressões de um repertório restrito, privilegiado pelos materiais de ensino de língua estrangeira, tais como ‘personally’, ‘I think’. Os alunos nativos, por sua vez, exprimem a interpessoalidade por meio de um outro repertório que inclui, por exemplo, traços de oralidade. Tal repertório não faz parte, tipicamente, dos materiais de ensino de língua estrangeira e, portanto, não consta dos recursos de expressão do aluno não-nativo.

Há uma semelhança entre os achados de Shimazumi (1998) e Pacheco (1997). Esta semelhança é realçada pelo fato de ambos os

estudos terem sido conduzidos com alunos brasileiros de inglês como língua estrangeira. Isto torna os resultados mais diretamente comparáveis. Embora os dois estudos tivessem usado quadros analíticos diferentes, em ambos os estudos, os alunos não-nativos demonstraram privilegiar aspectos interpessoais na escrita. Parece haver, portanto, uma tendência de os alunos não-nativos brasileiros enfatizarem a relação interpessoal com o leitor.

Críticas à Análise Multidimensional

A Análise Multidimensional se propõe, na sua essência, a descrever aspectos textuais. Sendo assim, pode-se levantar duas objeções em relação ao modo como ela foi levada a cabo por Biber (1988, 1995). A primeira crítica se refere ao fato de Biber ter utilizado fragmentos de textos escritos e falados (retirados dos *corpora* LOB e London-Lund). A utilização de fragmentos causa uma perda de certas características constitutivas da textualidade. Aspectos textuais fundamentais para a formação textual como coesão (Hoey, 1991) e textura (Halliday e Hasan, 1989) necessitam do texto completo para serem analisados satisfatoriamente.

Em resumo, por um lado o trabalho de Biber enfoca a análise do texto, mas por outro lado o faz por meio de trechos de textos. Esta é uma limitação grave, visto que suas conclusões, embora extrapoladas para o universo de textos escritos em inglês, são na verdade baseadas em informação referente a textos incompletos.

A defesa desta crítica é óbvia: Biber não necessita de textos completos para analisar os aspectos que ele privilegia. Entretanto, este fato em si já demonstra que os aspectos contemplados por ele se referem mais a traços estruturais no nível da oração e da palavra do que propriamente textuais. Ou seja, estes aspectos poderiam prontamente ser empregados para a análise de orações isoladas.

A segunda crítica, na verdade, se refere justamente a este descompasso entre seu ponto de partida (a palavra e categorias estruturais) e suas conclusões (tipos de texto). Conforme argumenta Stubbs (1996, p.33), 'Biber's analysis is across representative samples of

genres and sub-genres, with no analysis of the discourse structure of individual instances of the genres.’ Uma análise multidimensional mais centrada no texto contemplaria traços discursivos, como por exemplo, relativos à organização textual:

‘Analysis of clause level relations such as cohesion, information structure, thematic organisation, etc. requires a unit that is fundamentally different from Biber’s units’ (Ghadessy, no prelo, p.4).

Em um outro trabalho, há evidências deste descomprometimento para com a investigação de componentes textuais. Biber (1995, p.131) enfoca especificamente a questão do tamanho mínimo de texto necessário para seu tipo de análise. Ele conclui que fragmentos de 1000 palavras em geral possuem as mesmas características de outros textos maiores. Isto deixa claro que o tipo de análise de texto empreendida por Biber é na verdade feita no nível da palavra e não do texto.

Uma resposta a esta crítica seria o fato de Biber (1995) deixar em aberto o conjunto de características linguísticas que podem ser incluídos em uma análise multidimensional. Tanto isto é verdade que outros autores (Pacheco 1997; Shimazumi 1998; vide discussão acima) utilizaram-se de um quadro analítico que contempla categorias semânticas, textuais e organizacionais. Um outro trabalho (Ghadessy, 1996) demonstra como podem ser obtidas dimensões bem diferentes das dimensões de Biber levando-se em conta o aspecto organizacional no nível do texto em um conjunto de textos completos.

Em resumo, estas críticas não se referem ao modo pelo qual a Análise Multidimensional foi concebida, mas à maneira como tem sido posta em prática. Como foi relatado, outros trabalhos além dos de Biber tem incorporado traços linguísticos mais intimamente relacionados a aspectos discursivos do que estruturais. Por um lado, estes estudos se defrontam com o problema da análise em larga escala, visto que muitas das características discursivas são interpretativas e requerem análise manual, não podendo ser analisadas automaticamente (nem mesmo interativamente) por computador. Por outro lado, há um retorno satisfatório em termos de um ganho de validade.

Comentários finais

Nos dez anos que se passaram desde o lançamento de ‘Variation Across Speech and Writing’ de Douglas Biber houve um crescimento do número de trabalhos que adotaram a Análise Multidimensional como arcabouço metodológico. Em grande parte, a disseminação desta abordagem se deve ao seu criador, Douglas Biber, o qual tem uma produção acadêmica das mais expressivas, com publicações em vários dos principais veículos da área. Mas há outros nomes importantes associados à abordagem, como William Grabe, Ed Finegan e Dwight Atkinson. O grupo de seus seguidores continua crescendo, inclusive no Brasil, onde a Análise Multidimensional entrou pela porta da pesquisa em escrita em língua estrangeira.

A diversidade das frentes de investigação que podem se beneficiar da Análise Multidimensional é ampla, como demonstra o painel traçado neste trabalho. Pode-se aplicar esta abordagem na descrição de qualquer idioma, tanto sincrônica quanto diacronicamente. As virtudes da abordagem são inegáveis em comparação com muitas das metodologias de análise de discurso tradicionais (e.g. Hoey, 1983; Swales, 1990; van Dijk, 1980): maior abrangência em relação ao número de traços incorporados à análise, maior cobertura no que tange à quantidade de textos que se pode investigar e um ganho de consistência na análise, visto que as dimensões são extraídas computacionalmente.

A abordagem não está imune à críticas, contudo, como mostrou o relato na seção anterior, principalmente no tocante à questão da validade dos traços incluídos na análise. Entretanto, conforme discutido acima, estas críticas não se referem a uma característica inerente da metodologia, antes ao modo como ela tem tipicamente sido levada a cabo.

Mesmo os críticos mais vorazes das abordagens quantitativas não deveriam ignorar a riqueza das descrições linguísticas apresentadas nos principais trabalhos que empregam a Análise Multidimensional, tampouco o esforço interpretativo investido nas análises. As descrições com base multidimensional podem se tornar uma fonte de referência para os estudiosos do texto de qualquer vertente.

REFERÊNCIAS BIBLIOGRÁFICAS

- ATKINSON, D. (1992) The evolution of medical research writing from 1735 to 1985: The case of the 'Edinburgh Medical Journal'. *Applied Linguistics* **13**: 337-74.
- _____. (1993) A historical discourse analysis of scientific research writing from 1675 to 1975: The case of the 'Philosophical Transactions of the Royal Society of London'. Tese de PhD inédita, Department of Linguistics, University of Southern California, EUA.
- BIBER, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- _____. (1995). *Dimensions of Register Variation – A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.
- CONNOR, U. (1996). *Contrastive Rhetoric – Cross-cultural Aspects of Second Language Writing*. Cambridge: Cambridge University Press.
- GHADESSY, M. (1996). Criteria for English text types. Comunicação apresentada no 8º Euro-International Systemic Functional Workshop, Nottingham Trent University, Nottingham, Reino Unido, 24 de julho de 1996.
- _____. (no prelo). Review of Biber, D (1995) Dimensions of register variation: A cross linguistic comparison. *System*.
- HALLIDAY, M. A. K., E R. HASAN (1989). *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. 2ª ed. Oxford: Oxford University Press.
- HOEY, M. (1983). *On the Surface of Discourse*. London: George Allen & Unwin.
- _____. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- PACHECO, L. (1997). *Variação intercultural na escrita: Contrastes multidimensionais em inglês e português*. Tese de Doutorado inédita. São Paulo: PUCSP, LAEL.
- SHIMAZUMI, M. (1998). Investigating EFL writing: A Multidimensional analysis. Comunicação apresentada na 6ª Convenção Braz-TESOL, Recife, PE, 13 a 16 de julho de 1998.
- STUBBS, M. (1996). *Text and Corpus Analysis – Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell.
- SWALES, J. M. (1990). *Genre Analysis – English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- VAN DIJK, T. A. (1980). *Macrostructures. An Interdisciplinary Study of Global Structures in Discourse, Interaction and Cognition*. Hillsdale, NJ: Erlbaum.

ANEXO

Escore médios de cada registro nas seis dimensões

Registros	Dimensões					
	1	2	3	4	5	6
Biografias	-12.4	2.1	1.7	-0.7	-0.5	-0.3
Cartas pessoais	19.5	0.3	-3.6	1.5	-2.8	-1.4
Cartas profissionais	-3.9	-2.2	6.5	3.5	0.4	1.5
Conversaão face a face	35.3	-0.6	-3.9	-0.3	-3.2	0.3
Conversaão telefnica	37.2	-2.1	-5.2	0.6	-3.7	-0.9
Cultura popular	-9.3	-0.1	2.3	-0.3	0.1	-0.8
Documentos oficiais	-18.1	-2.9	7.3	-0.2	4.7	-0.9
Editoriais jornalsticos	-10	-0.8	1.9	3.1	0.3	1.5
Entrevistas	17.1	-1.1	-0.4	1	-2	3.1
Fico cientfica	-6.1	5.9	-1.4	-0.7	-2.5	-1.6
Fico de aventura	0	5.5	-3.8	-1.2	-2.5	-1.9
Fico de mistrio	-0.2	6	-3.6	-0.7	-2.8	-1.9
Fico geral	-0.8	5.9	-3.1	0.9	-2.5	-1.6
Fico romntica	4.3	7.2	-4.1	1.8	-3.1	-1.2
Humor	-7.8	0.9	-0.8	-0.3	-0.4	-1.5
Palestras espontneas	18.2	1.3	1.2	0.3	-2.6	2.6
Palestras preparadas	2.2	0.7	0.3	0.4	-1.9	3.4
Passatempos	-10.1	-2.9	0.3	1.7	1.2	-0.7
Prosa acadmica	-14.9	-2.6	4.2	-0.5	5.5	0.5
Rdio e TV	-4.3	-3.3	-9	-4.4	-1.7	-1.3
Religio	-7	-0.7	3.7	0.2	1.4	1
Reportagem jornalstica	-15.1	0.4	-0.3	-0.7	0.6	-0.9
Resenhas jornalsticas	-13.9	-1.6	4.3	-2.8	0.8	-1

(Recebido em fevereiro de 1999; Aceito em junho de 1999)