

**A COMPETÊNCIA LEXICAL DE ALUNOS
UNIVERSITÁRIOS APRENDENDO A LER EM INGLÊS
COMO LÍNGUA ESTRANGEIRA***

(The Lexical Competence of
University Students to Read in EFL)

Matilde V. Ricardi SCARAMUCCI (UNICAMP)

ABSTRACT: This article presents a characterization of the lexical competence (vocabulary knowledge and use) of students learning to read in EFL in a public university in São Paulo state. Although vocabulary has been consistently cited as one of the EFL reader's main source of difficulty, there is no data in the literature which shows the extent of the difficulties. The data for this study is part of a previous research, which investigates, from the perspective of an interactive model of reading, the relationship between lexical competence and EFL reading comprehension. Quantitative as well as qualitative data was considered. For this study, the quantitative data is the product of vocabulary tests of 49 subjects while the qualitative data comprises pause protocols of three subjects, with levels of reading ability ranging from good to poor, selected upon their performance in the quantitative study. A rich concept of vocabulary knowledge was adapted and used for the development of vocabulary tests and analysis of protocols. The results on both studies show, with a few exceptions, the lexical competence of the group to be vague and imprecise in two dimensions: quantitative (number of known words or vocabulary size) and qualitative (depth or width of this knowledge). Implications for the teaching of reading in a foreign context are discussed.

RESUMO: Neste artigo apresento uma caracterização da competência lexical (conhecimento de vocabulário e uso desse conhecimento) de alunos aprendendo a ler em inglês como língua estrangeira (LE doravante) em uma universidade estadual paulista. Apesar de se afirmar muito frequentemente que o vocabulário é uma das maiores fontes de dificuldade do leitor em LE, não se tem dados

* Uma versão reduzida deste trabalho foi apresentada no XI Encontro Nacional da ANPOLL, GT de Linguística Aplicada, de 02 a 06 de junho de 1996, em João Pessoa.

na literatura que mostrem a extensão dessas dificuldades. Essa caracterização se baseia, em parte, em dados de uma pesquisa anterior mais ampla, que envolve duas fases – quantitativa e qualitativa – visando a investigar, na perspectiva de um modelo de leitura interativo, a relação entre competência lexical e compreensão em leitura em LE. Para o presente trabalho, foram considerados o produto de testes de vocabulário de 49 sujeitos (fase quantitativa) e os protocolos de leitura de três sujeitos de diferentes graus de proficiência em leitura, selecionados a partir da fase anterior (fase qualitativa). Um conceito rico de vocabulário foi adaptado e usado tanto para o desenvolvimento dos testes de vocabulário como para a análise dos protocolos. Os resultados dos dois estudos mostram, com algumas exceções, que a competência lexical do grupo investigado é vaga e imprecisa, tanto do ponto de vista do número de palavras conhecidas, como do ponto de vista da profundidade desse conhecimento. Implicações para o ensino da leitura em LE são discutidas.

Key words: Reading; English; Foreign language; Rich Concept of Vocabulary Knowledge; Lexical Competence.

Palavras-chave: Leitura; Inglês; Língua Estrangeira; Conceito Rico de Vocabulário; Competência Lexical.

0. Introdução

Meu objetivo neste trabalho é apresentar uma caracterização da competência lexical de alunos aprendendo a ler em inglês como LE em uma universidade estadual paulista. Essa caracterização se baseia em parte dos dados coletados em um estudo anterior (Scaramucci, 1995), cujo objetivo é investigar, na perspectiva de um modelo de leitura interativo, o papel do vocabulário na compreensão em leitura em inglês como LE. Caracterizações desse tipo, importantes não apenas em estudos que visam a estabelecer sua relação com a leitura – como é o caso da pesquisa que originou estes dados – mas também naqueles que têm por objetivo orientar o ensino de LE, não têm sido encontradas na literatura, apesar de se afirmar muito frequentemente que o vocabulário é uma das maiores fontes de dificuldades do leitor em LE.

Este artigo está subdividido em quatro partes: na primeira, apresento o conceito de vocabulário usado na definição de competência lexical. Na segunda, caracterizo a metodologia, os sujeitos e os instrumentos envolvidos; na terceira, analiso e discuto os resultados, enquanto na quarta e última, saliento algumas implicações para o ensino de inglês como LE.

1. Competência lexical: um conceito *rico* de vocabulário e uma dimensão de uso

Com relação à definição de competência lexical, o que uma revisão da literatura mostra são, com algumas exceções, modelos implícitos (vide Scaramucci, 1995, para uma revisão desses modelos), inferíveis a partir dos testes de vocabulário utilizados. Em geral, esses modelos são reducionistas: conhecer uma palavra é conhecer um significado, que teria formas equivalentes em duas línguas (*paired-associate*), ou ainda ser capaz de reconhecer ou lembrar uma definição. As medidas de vocabulário que se baseiam nesses modelos, portanto, consideram apenas a extensão do vocabulário do aprendiz ou o número de palavras conhecidas/desconhecidas (*size* ou *breadth*). Alguns trabalhos, entretanto, como por exemplo, Richards 1976 (vide também Read, 1987; Nation, 1984 e Wallace, 1982) têm tentado ampliar esse conceito, nele incluindo um leque de componentes que envolve aspectos lingüísticos, psicolingüísticos e sociolingüísticos. Esse conceito de vocabulário tem sido reconhecido na literatura como um conceito *rico*, pois abrange, além da dimensão "extensão", também uma dimensão qualitativa, ou de profundidade (*depth*) desse conhecimento. Para o seu proponente, os componentes são definidos a partir de uma série de pressuposições sobre o conhecimento lexical de um falante nativo em termos ideais, e poderiam servir de parâmetro para se estimar que tipo de conhecimento e produto final de aprendizagem podem ser esperados de um aprendiz em LE. Apresento, a seguir, os aspectos considerados por Richards (1976) como parte desse conceito *rico*:

- 1.1 O falante nativo de língua contínua expandindo seu vocabulário na idade adulta, embora haja comparativamente, pouco desenvolvimento da sintaxe. Nesse caso, o autor parece focalizar, levando em conta outras considerações que faz no texto, a importância de um vocabulário amplo e de uma expansão

contínua, o que interpretei como se referindo ao aspecto *extensão* (*size*);

Os demais itens abaixo referem-se ao aspecto *profundidade* (*depth*).

Conhecer uma palavra significa:

- 1.2 Saber o grau de probabilidade de encontrar essa palavra na fala ou na escrita. Esse item diz respeito a dois tipos de conhecimento: frequência e colocabilidade. Dada uma lista de palavras, o falante nativo é capaz de classificá-las em diferentes níveis de frequência (frequente, mais ou menos frequente ou não frequente), além de ser capaz de reconhecer sua colocabilidade, isto é, a probabilidade de ocorrência com outras palavras ou suas relações sintagmáticas. Por exemplo, a palavra *fruta* coloca-se com *verde*, *madura*, *gostosa*; a palavra *céu* coloca-se com *azul*;
- 1.3 Conhecer as limitações impostas ao seu uso de acordo com variações de função e de situação, incluindo variação temporal, social, geográfica., de área ou do modo de discurso. Pode-se dizer, portanto, que refere-se às características de registro;
- 1.4 Pressupor conhecimento de sua forma subjacente e das derivações e flexões que podem ser feitas a partir dela;
- 1.5 Conhecer suas propriedades gramaticais e estruturais. Este é o aspecto que mais contraria a divisão tradicional entre gramática e vocabulário;
- 1.6 Conhecer a rede de associações com outras palavras da língua ou suas relações paradigmáticas. Por exemplo, a palavra *acidente* está associada a *hospital*, *ambulância*, *médico*, *enfermeira*, dentre outras;
- 1.7 Conhecer seu valor semântico ou seu significado denotativo;
- 1.8 Conhecer os diferentes significados associados a ela ou seus significados conotativos.

Embora esse modelo tenha sido proposto para língua materna (LM doravante), considero-o importante também para LE, na medida em que amplia o conhecimento da palavra para além de seu significado, passando a incluir suas frequências/colocações, registros, comportamento sintático, derivações/flexões, associações, significados denotativo e conotativo. Apesar da importância dessa base teórica ampliada, entretanto, devo salientar que ela ainda se refere a um conceito estático, por não fazer referência a uma dimensão ou

capacidade de uso. Embora incorpore, como parte do conhecimento lexical, vários outros níveis que, na lingüística tradicional e também no ensino de línguas são considerados separadamente ou como parte da gramática (flexões/derivações, relações sintáticas) e embora faça menção a regras de uso (variações de função e situação, incluindo variação temporal, social, geográfica da área ou modo do discurso), ainda está restrita ao aspecto semântico (denotativo/conotativo), não considerando um nível pragmático/discursivo que inclui a capacidade de uso desse conhecimento, responsável pela realização ou atualização desses níveis na construção da compreensão.

Neste trabalho, portanto, adoto esse conceito de conhecimento lexical, que incorpora vários aspectos especificados no conceito *rico*, ou seja, os níveis lexical, sintático, morfológico, semântico, ao qual adiciono um nível discursivo/pragmático, incorporando à dimensão lingüística também uma dimensão de uso, ou cognitiva, ao invés de considerá-la separadamente, levando-se em conta a interação entre elas. Essa dimensão de uso focaliza as habilidades necessárias ao acesso a significados de palavras através do reconhecimento automático/não automático, isto é, habilidades de nível mais baixo ou de decodificação, assim como procedimentos de nível mais alto, ou seja, aqueles usados na construção do sentido do texto.

2. A metodologia

O desenho desta pesquisa procurou levar em conta, dentre outros aspectos, a necessidade de se avaliar as duas dimensões acima salientadas, ou seja, uma avaliação qualitativa e quantitativa do conhecimento lexical dos sujeitos envolvidos, assim como uma dimensão de uso. Por isso, apresenta dois focos de análise ou uma divisão em duas fases:

- * Fase I, de ênfase quantitativa, focalizando o conhecimento lexical através do produto de testes de vocabulário¹; e
- * Fase II, de ênfase qualitativa, focalizando a competência lexical (conhecimento lexical/uso desse conhecimento)

¹ No estudo original (Scaramucci, 1995), nesta fase também foram usados testes de compreensão em leitura, dentre outros, além de questionários.

durante o processo de construção do sentido.

2.1. Fase I - Estudo de ênfase quantitativa com foco no produto

Os sujeitos são 49 alunos de graduação da área de Exatas (Engenharia Elétrica e Matemática Aplicada), com idades variando entre 16 e 25 anos, cursando Inglês Instrumental I (17 sujeitos) e II (32 sujeitos) em uma universidade estadual paulista, apresentando níveis de proficiência em leitura que variam de *muito bom a muito fraco*. A disciplina Inglês Instrumental é obrigatória para esses alunos durante dois semestres, com 60 horas/aula, distribuídas em 4 horas/aula semanais. A grande maioria dos sujeitos (39) já havia estudado inglês durante mais de 5 anos, tendo mais da metade (29 sujeitos) freqüentado escolas de línguas. Em uma auto-avaliação que faz parte da bateria de testes utilizada, o vocabulário foi apontado como o maior problema por 40 dos sujeitos.

A abordagem de ensino utilizada tinha como objetivos principais tornar os alunos mais conscientes das estratégias usadas na leitura em LM assim como praticá-las na leitura de textos em inglês. Consistia, portanto, em um trabalho centrado em estratégias, tais como: localização de informação específica, levantamento de hipóteses a partir do título, inferência lexical, dentre outras, através de textos de vulgarização científica de temas variados, extraídos de jornais e revistas.

Baseado no conceito de competência lexical acima explicitado, foi desenvolvida uma bateria de quatro testes de vocabulário. Esses testes não constituem, entretanto, uma medida absoluta de conhecimento de vocabulário, ou da extensão total ou do número de palavras conhecidas do vocabulário dos sujeitos, mas relativa, ou das palavras conhecidas/desconhecidas em um texto. A partir desse conhecimento espera-se fazer inferências ou generalizações com relação à extensão do conhecimento lexical do sujeito.

A Tabela 1 abaixo apresenta os testes e seus focos de avaliação:

Testes de vocabulário	Focos de avaliação
Teste I (T1)	Familiaridade
Teste II (T2)	Frequência
Teste III (T3, T4, T5, T6, T7, T8)	Conceito <i>rico</i> de vocabulário
Teste IV (T9)	Inferência em contexto

Tabela 1 – Testes de vocabulário e seus focos de avaliação

O teste de vocabulário I (T1) é uma auto-avaliação da familiaridade do sujeito com as palavras, conduzida através de uma escala de verificação (*checklist*), baseada em Dale (1965). São 30 as palavras que compõem esse teste — 20 reais (extraídas de um texto, comuns a todos os testes, selecionadas por alunos de nível de proficiência semelhante ao dos sujeitos) e 10 inventadas (comuns apenas ao teste II (T2) ou teste de frequência) — seguidas de uma escala de três níveis (*nunca vi essa palavra antes, essa palavra existe mas não sei o que quer dizer, essa palavra existe e eu conheço seu significado*). A inclusão das palavras inventadas se justifica enquanto medida de consistência e também de controle contra possíveis adivinhações. Elas foram inventadas segundo dois critérios: (1) palavras reais mudando-se uma ou duas letras e (2) combinações diferentes de radical mais afixo (*novel base-and-affix combinations*).

As palavras reais foram extraídas de um texto de 350 palavras em inglês intitulado *Galactic Birth?*, extraído da revista *Time* de 11 de agosto de 1989, reproduzido no Apêndice A. Em Scaramucci (1995) esse texto foi usado também para o teste de leitura.

O teste de vocabulário II (T2), ou de frequência, consiste na mesma lista de 30 palavras usada no teste de familiaridade (T1), seguida de uma escala de três níveis (*frequente, mais ou menos frequente, não frequente*). Sua finalidade é verificar a percepção dos sujeitos com relação à frequência das palavras que normalmente encontram. Espera-se, também, verificar a relação entre a frequência e familiaridade, uma vez que palavras mais frequentes tendem a ser mais familiares e vice-versa.

O teste de vocabulário III (T3) é composto por seis subtestes com finalidades, formatos e instruções diferentes. O objetivo geral é uma medida abrangente dos vários aspectos que compõem o conhecimento de vocabulário de acordo com o conceito *rico* acima exposto. A bateria de seis testes é a operacionalização desse conceito, tendo envolvido algumas alterações, como poderá ser observado a seguir na Tabela 2. Por exemplo, os aspectos de registro foram eliminados, assim como os aspectos conotativos. Para esses subtestes, não foram consideradas as palavras inventadas, mas apenas as 20 reais.

Para a elaboração desse teste, assim como para os demais, partiu-se de um princípio corrente na área de avaliação, que relaciona a extensão de um teste com sua confiabilidade, ou seja, testes mais longos tendem a ser mais confiáveis. Entretanto, essa extensão não poderia comprometer sua praticidade, introduzindo algum tipo de efeito fadiga. Procurou-se, portanto, um equilíbrio entre extensão, confiabilidade e praticidade. A Tabela 2, abaixo, apresenta esses subtestes e seus focos de avaliação.

Subteste	Focos de avaliação
III-A (T3)	Derivações (quando a palavra é apresentada em sua forma primitiva)
	Forma subjacente da palavra (no caso da palavra derivada)
	Flexões (singular/plural)
III-B (T4)	Significado das palavras
	Significados adicionais das palavras.
III-C (T5)	Associações
III-D (T6)	Colocações
III-E (T7 e T8)	Contextos gramaticais possíveis para cada palavra, ou seja, as relações gramaticais que cada palavra apresenta, sem uso de metalinguagem
	Classes gramaticais, usando-se metalinguagem

Tabela 2. Testes elaborados de acordo com um conceito *rico* de conhecimento de vocabulário

O subteste III-E, como se pode notar, subdivide-se em duas partes. Na primeira, ou T7, o que realmente se deseja verificar é a função ou as funções gramaticais de cada uma das palavras avaliadas dentro da oração (substantivo, adjetivo, por exemplo) sem o uso de metalinguagem, que o sujeito poderia desconhecer. Para isso foi apresentada uma lista (embora não exaustiva) de contextos possíveis e impossíveis, que foram mantidos em um nível mínimo, de forma a não influenciarem o resultado ou não permitirem inferências em nível de significado. O formato, portanto, difere dos demais testes uma vez que, neste caso, optou-se por um método de múltipla escolha. Para cada item há um número diferente de opções (*distractors*), que variam de 3 a 10. Na segunda parte do subteste III-E, ou seja, T8, foi solicitado o mesmo tipo de informação da primeira parte, ou seja, classe ou função gramatical da palavra em contexto. Entretanto, neste caso, o conhecimento da metalinguagem é necessário, uma vez que o sujeito é solicitado a explicitar a(s) classe(s) gramatical(ais) de cada palavra.

No teste de vocabulário IV ou de *inferência em contexto* (vide Tabela 1), as 20 palavras dos testes anteriores são recolocadas no texto *Galactic Birth?* de onde foram extraídas, para serem inferidas.

Para a análise dos dados, nesta fase I, foi usada uma técnica estatística exploratória denominada Análise de Correspondência. Essa técnica (vide Scaramucci (1995), para uma descrição mais detalhada) é apropriada para tabelas de frequência onde o número de variáveis e associações são numerosas, como é o caso neste trabalho. "Correspondência", nesse caso, refere-se a um sistema de associações entre os elementos dos conjuntos constituídos pelas categorias da tabela (linhas e colunas).

2.2. Fase II - Estudo de ênfase qualitativa com foco no processo

A fase II, por sua vez, focaliza o segundo aspecto do tópico investigado ou a dimensão dinâmica do conhecimento de vocabulário. Essa dimensão envolve a capacidade de uso desse conhecimento na construção do sentido do texto, geralmente mediado pelo acesso automático e não automático a significados, ou pelo reconhecimento de palavras durante a decodificação. Nessa dimensão, portanto, estão envolvidos não apenas as fontes de conhecimento/procedimentos de

uso, mas também a eficiência e a velocidade ou a automaticidade com que tal reconhecimento é operacionalizado pelo sujeito.

Participaram dessa fase três sujeitos, apresentando níveis de leitura caracterizados como *intermediário avançado*, *intermediário e iniciante*, selecionados a partir de seu desempenho na fase I acima descrita. Para essa seleção, além do critério teórico, ou seja, a necessidade de se avaliar sujeitos de níveis diferentes de proficiência em leitura, adotou-se também um critério prático: a disponibilidade dos sujeitos para a coleta dos dados, uma vez que ela não foi conduzida durante as aulas, mas em sessões especiais.

O leitor intermediário avançado (LIA doravante) cursava Engenharia Elétrica e a disciplina Inglês Instrumental no nível II. Estudou inglês durante 5 anos, tendo freqüentado escolas de línguas. Na época da coleta, não fazia nenhum outro curso de inglês, além do Inglês Instrumental. Para esse leitor, assim como para os outros dois, o vocabulário é apontado como sua maior dificuldade na leitura. O leitor intermediário (LIO doravante), por sua vez, cursava Matemática Aplicada e Computacional e a disciplina Inglês Instrumental no nível I. Estudou inglês durante 2 anos, tendo também freqüentado escolas de línguas. O leitor iniciante (LIE), finalmente, cursava Matemática Aplicada e Computacional e a disciplina Inglês Instrumental no nível I, na época da coleta dos dados. Estudou inglês durante 3 anos, mas nunca freqüentou escolas de línguas ou qualquer outro curso de inglês, além do Inglês Instrumental.

Uma técnica mentalista e introspectiva, mais especificamente o protocolo de pausa (vide Cavalcanti 1983 e 1989; Scaramucci 1990 e 1995, para uma explicação detalhada desse método), foi escolhida para a elicitación dos dados, por ser aquela que, a meu ver, mais se aproxima do processo de construção da compreensão. Essa técnica consiste em solicitar aos sujeitos que leiam silenciosamente (monitorando seu processo de leitura), com o objetivo de fazerem um resumo oral, e pensem alto toda vez que observarem pausas no seu processo. O dados do processo são complementados por dados do produto, isto é, dos resumos que os sujeitos são solicitados a elaborarem no final de cada parágrafo, trazendo maior validade e confiabilidade a esses dados.

O texto-base para o protocolo (vide Apêndice B) consistiu em um artigo extraído da revista *New Scientist* de 1 de setembro de 1990, cujo tema é a ecologia na União Soviética; tendo aproximadamente 700 palavras, apresenta uma estrutura problema-solução (Hoey, 1979). Os critérios que nortearam sua seleção foram os mesmos usados para a seleção dos textos na fase I:

- * Tópico familiar e de interesse do leitor. O tema já havia sido discutido durante as aulas e despertado o interesse dos alunos; e
- * Texto bem organizado, apresentando um grau de dificuldade que possivelmente apresentaria alguns desafios para os sujeitos, principalmente para os de nível intermediário e iniciante.

3. Análise e discussão dos resultados

Os resultados dos dois estudos são discutidos separadamente a seguir.

3.1 Fase I - Estudo de ênfase quantitativa com foco no produto

A análise conduzida nos resultados dos testes de vocabulário mostra o desempenho geral do grupo nos vários testes, as associações entre eles, além de gerar um escore para cada sujeito em cada um dos testes. Esse escore é usado como pano de fundo para a análise no estudo da fase II.

Através das respostas do grupo como um todo no teste de *familiaridade* (pode-se classificar as palavras em três níveis:

- * Palavras mais familiares: *close, slowly, empty, spring, receiving e appeared;*
- * Palavras de familiaridade média: *cloud, *adjustion, picked up, pattern, shaken, *observement, throughout, massive, fairly, fortuitously e disk-shaped;*
- * Palavras menos familiares: **drealled, *glim, *tark, *trivel, loomed, spawned, *lodings, *leam, resembling, *resks, *skapping, cluster e focusing.*

As palavras precedidas de * são as palavras inventadas. Algumas delas, tais como *adjustion* e *observement*, talvez pelo fato de serem semelhantes a palavras em português, ou "cognatas", foram identificadas como de familiaridade média.

Os resultados também revelaram uma forte associação entre os aspectos *familiaridade* e *freqüência*, confirmando a relação que tem sido geralmente identificada na literatura em LM, ou seja, palavras mais freqüentes tendem a ser mais conhecidas. Palavras de alta freqüência compõem um vocabulário básico e, portanto, têm sido denominadas de palavras básicas (*core vocabulary*) ou de vocabulário procedimental (*procedural vocabulary*) (vide Robinson 1989); são palavras de uso generalizável, que não pertencem a nenhum esquema em particular, e, portanto, mais difíceis de serem inferidas, uma vez que são usadas na construção de outros significados. Entretanto, com relação a esse aspecto, algumas observações são apresentadas a seguir.

Para que um parâmetro aproximado de comparação com os resultados dos leitores em LE deste estudo pudesse ser estabelecido, utilizei-me de dados de um grupo de falantes nativos que serviram de critério para alguns dos testes de vocabulário. Embora ciente de que a proficiência em LE seja distinta daquela apresentada pelo falante nativo, considero esses dados importantes na medida em que dão consistência ao conceito *rico* de vocabulário proposto por Richards (1976), que, como o autor mesmo reconhece, é baseado em pressuposições sobre o conhecimento do falante nativo e não parece ter sido "testado" em situações reais. Além disso, dão validade à bateria de testes que são sua operacionalização. Os dados dos falantes nativos, entretanto, também foram importantes para mostrar que muitas das palavras consideradas freqüentes pelos nativos não o eram para os leitores em LE, uma vez que a escala de freqüência dos dois grupos mostrou ser diferente. Neste estudo, portanto, um dos fatores que distingue o vocabulário dos dois grupos é a freqüência das palavras conhecidas/desconhecidas. Enquanto o falante nativo conhece palavras de alta freqüência, podendo apresentar problemas com palavras de baixa freqüência, o leitor em LE apresenta problemas também no reconhecimento das palavras de alta freqüência.

Esse aspecto é mais uma vez observado e parece se confirmar através dos resultados de uma outra tarefa que compõe a bateria de

testes deste estudo, ou seja, a tarefa de grifar palavras.² Além das palavras selecionadas para o teste de vocabulário, muitos leitores (pelo menos 14) identificaram como desconhecidas outras palavras no texto (vide cópia no Apêndice A). Uma análise dessas palavras mostra que são, em sua maioria, palavras de alta frequência. Esse resultado é importante na caracterização da competência lexical do grupo, mostrando que outras palavras, além daquelas selecionadas para os testes de vocabulário, são também desconhecidas. Isso significa, em termos do grupo estudado, que se pode identificar, na realidade, dois subgrupos: um em que tem seus problemas de vocabulário limitados às palavras do teste de vocabulário, e o outro, que além das palavras do teste, ainda desconhece outras palavras no texto. Isso significa dizer que, enquanto o primeiro grupo se caracteriza pelo conhecimento de palavras de frequência alta e desconhecimento de palavras de frequência média e baixa, o segundo se caracteriza também pelo desconhecimento de palavras de frequência alta ou seja, palavras básicas (*core words*).

Analisando-se esses resultados, pode-se verificar que grande parte dos sujeitos, ou seja, 35, grifaram apenas palavras que faziam parte da lista de inventadas. Entretanto, 14 sujeitos grifaram, além das inventadas, também outras palavras do texto. São elas: *while, birth* (7 ocorrências), *discovered, shortly, happened* (6), *signals, outer, Rosetta Stone, announced* (5), *gestation, announcement, calibrate, Big Bang stone, thought, embryo, after, some, ago, astrophysicist, in order to, Milky Way, travels, contained* (3), *during, form, earth, born* (4), *since* (2). Como se pode notar, palavras muito comuns ou de frequência bastante alta fazem parte dessa lista, assim como palavras consideradas cognatas, tais como *announced, announcement, discovered, signals, calibrate, embryo, form*.

Esses resultados são importantes por mostrar que o critério usado para selecionar as palavras problemáticas no teste de vocabulário

²Essa tarefa faz parte de um teste de leitura usado no estudo original. Devo lembrar que o teste de leitura se baseia no mesmo texto do qual foram retiradas as palavras para os testes de vocabulário. Antes de responderem ao teste, os sujeitos deveriam grifar todas as palavras desconhecidas.

superestimou o vocabulário dos sujeitos, ou seja, deixou de considerar palavras desconhecidas. Uma análise das palavras escolhidas pelo grupo que serviu de critério para a seleção das palavras mostra que um número maior de palavras, na realidade, tinham sido identificadas como problemáticas. Entretanto, como a inclusão de todas no teste de vocabulário o tornaria muito longo, foi necessário a adoção de um ponto de corte, selecionando assim palavras que tinham sido escolhidas por pelo menos 10% dos leitores. Portanto, o que se pode concluir a partir desses resultados é que o conhecimento de vocabulário do grupo estudado contém, na realidade, mais lacunas do que se imaginava.

Uma outra associação, a mais forte delas, foi observada entre os aspectos que compõem o conhecimento *rico* de vocabulário (vide Tabela 2) e o teste de *familiaridade* (escala de verificação ou *check-list*) (vide Tabela 1), mostrando que aqueles que identificaram a palavra como familiar, em geral conheciam seu significado, além de saber também associar, colocar, flexionar, o que poderia ser indicativo da adequação de uma escala de verificação enquanto medida de conhecimento de vocabulário.

Outras associações, desta vez entre os vários aspectos que compõem os conceito *rico* de vocabulário, também foram observadas, podendo ser consideradas uma medida de consistência da bateria desenvolvida para avaliar o vocabulário e, principalmente, a adequação do conceito de vocabulário operacionalizado através dos testes. Apesar dessa associação, alguns dos testes se mostraram, para o grupo como um todo, mais difíceis que outros, como se houvesse um contínuo de dificuldade entre eles. Nesse sentido, o aspecto *colocações* mostrou-se como o mais difícil, seguido do aspecto *associações*, enquanto o de *classes gramaticais* como o mais fácil. Esse resultado é curioso, uma vez que esse teste exigia a classificação das palavras em suas classes gramaticais, e, portanto, o uso de metalinguagem. O teste de *contextos gramaticais*, onde os contextos possíveis e os impossíveis foram apresentados para escolha, foi considerado mais difícil pelos sujeitos, embora não fosse necessário nenhum conhecimento metalingüístico. Apesar de apresentarem níveis diferentes de dificuldade, os dois testes apresentam uma grande associação. O teste de *significado* foi o segundo mais fácil, seguido pelo de *derivação/flexão*. As palavras não foram definidas, mas sim

"traduzidas", uma vez que o tipo de resposta mais comum foi um sinônimo em português.

Nos testes considerados "difíceis" houve um número muito grande de lacunas e respostas em branco. À primeira vista, essas lacunas poderiam ser indicativas de deficiências do próprio instrumento de avaliação. Entretanto, nesse caso, mais uma vez, os dados dos falantes nativos foram importantes para eliminar essa hipótese. Embora os resultados dos falantes nativos mostrem diferenças entre si com relação a algumas palavras, ou seja, que o conhecimento de algumas palavras é mais completo do que de outras, pode-se afirmar, pelo menos para as palavras avaliadas, que um conceito *rico* se mostrou bastante adequado. Nos subtestes que compõem esse conhecimento, por exemplo, *flexão/derivação*, *significado*, *colocações* (vide Tabela 2), dentre outros, os falantes nativos raramente deixam itens em branco e as respostas apresentam uma profusão de detalhes. As muitas lacunas apresentadas pelos falantes não nativos em seus testes não impedem que se considere adequado um conceito *rico* de vocabulário também para uma situação de LE, uma vez que fortes associações entre os várias subtestes de *conhecimento de vocabulário* foram identificadas. Essas lacunas, nesse caso, podem ser consideradas, com mais segurança, como insuficiências no conhecimento desses sujeitos e não limitações do instrumento.

Com relação ao aspecto *significado*, os acertos que se observam referem-se apenas à parte I, ou seja, que solicita seja dado um significado para cada palavra. A parte II, que solicitava *significados adicionais* para cada uma das palavras, não foi respondida por nenhum sujeito, ficando totalmente em branco. Isso parece mostrar que, para o grupo como um todo, os significados das palavras se restringem a apenas um. Esse aspecto também é observado no estudo com ênfase qualitativa, discutido mais adiante.

Uma possível explicação para a relativa facilidade de alguns testes e dificuldade de outros acima salientada poderia estar relacionada ao tipo de abordagem de ensino de LE a que esses sujeitos têm sido submetidos. No caso da *derivação/flexão*, a facilidade poderia estar associada à abordagem instrumental, que enfatiza prefixos e sufixos. Com relação à dificuldade com *significados*

adicionais, associação e colocação também as abordagens de ensino de língua poderiam ser responsáveis, uma vez que, em geral, focalizam as palavras isoladamente, sem a consideração de suas redes semânticas e possibilidades de colocações e o significado como o único aspecto relevante, sugerindo ao aluno que cada palavra tem apenas um significado correspondente em LM.

Um outro aspecto observado através dos resultados do teste *inferência em contexto* refere-se a um padrão de inferência ligado a conhecimento de vocabulário: os sujeitos que apresentaram um escore alto de *conhecimento de vocabulário* (obtido a partir do conjunto de todos os subtestes) obtiveram também um escore alto no teste de *inferência em contexto*, uma vez que conseguiram inferir um número maior de palavras. O processo de inferência de palavras é considerado mais detalhadamente no estudo de ênfase qualitativa.

Concluindo, poderia afirmar que a grande importância do estudo de ênfase quantitativa com foco no produto foi a de mostrar que a competência lexical do grupo em geral não é apenas insuficiente sob o ponto de vista qualitativo ou de profundidade, envolvendo os vários aspectos do conceito *rico* de vocabulário, mas também no número de palavras e natureza das palavras desconhecidas.

3.2. Fase II - Estudo de ênfase qualitativa com foco no processo

Os dados obtidos nesta fase complementam e confirmam aqueles obtidos na fase anterior. Enquanto aquela mostrou uma dimensão quantitativa, horizontal, baseada no universo dos 49 sujeitos, vistos enquanto um grupo, esta introduz uma dimensão qualitativa, vertical, focalizada no universo de três sujeitos, a partir de uma tarefa de compreensão e protocolo de pausa.

Embora o foco da análise dos protocolos seja a dimensão de uso da competência lexical dos sujeitos, ou seja, a maneira como o conhecimento identificado no primeiro estudo é efetivamente usado pelos leitores na construção do sentido do texto, o que será abordado mais adiante, pode-se observar, a partir desses dados de processo, também aspectos que se relacionam ao conhecimento de palavras, ampliando, portanto, a caracterização acima apresentada no estudo da fase I. Os dois subgrupos identificados no estudo anterior podem ser

ilustrados de forma bastante clara pelos dados dos protocolos dos sujeitos desta fase. O leitor *intermediário avançado* (LIA) pode ser visto como um exemplo do primeiro grupo, enquanto os leitores *intermediário* (LIO) e *iniciante* (LIE) como exemplos do segundo grupo.

O estudo de processo mostrou que todas as pausas (com exceção de uma) identificadas nos processos de construção de sentido dos leitores deste estudo, assim como outros problemas, não marcados por pausas, foram "detonados" por dificuldades que, na superfície, podem ser considerados lexicais, atribuídas ao desconhecimento de palavras, mas que uma análise mais detalhada mostra insuficiências ou dificuldades também em outros níveis (fonológico, morfológico, sintático, semântico e pragmático-discursivo), que, para serem selecionados, exigem do leitor recursos que vão além do significado da palavra. Isso significa dizer que a falta de palavras não introduz apenas lacunas de significados mas de blocos de informações, uma vez que as palavras trazem consigo instruções de como devem se relacionar entre si. Na perspectiva das soluções, poder-se-ia dizer que a efetiva atualização de significados potenciais em uma situação de uso (nível discursivo) parece exigir adaptações e relações em níveis que não se restringem ao lexical, ou apenas ao significado de palavras isoladas. Esses resultados, mais uma vez, parecem mostrar a adequação de um conceito *rico* de vocabulário.

A Tabela 3 a seguir mostra o número de pausas/problemas e as soluções de cada um dos três leitores.

Leitores	LIA	LIO	LIE
Pausas	23	81	96
Soluções	14 (61%)	11 (13%)	15 (16%)

Tabela 3 – Pausas/problemas e soluções dos sujeitos

Um dos traços salientes da competência vaga e imprecisa apresentada pelos LIO e LIE, e principalmente LIE, é a dificuldade com verbos, principalmente aqueles que não aparecem no texto em sua forma infinitiva. O LIE tenta, constantemente, acessar esses significados, mas suas tentativas são, na maioria das vezes, mal-sucedidas. O desconhecimento dos verbos torna difícil a segmentação,

necessária para a leitura em blocos maiores de significados, contribuindo para a leitura palavra por palavra. O LIO e LIE mostram desconhecer verbos tais como *reach, lead, ban, charge, maintain, damage, become*, dentre outros. Para o LIE, a lista ainda inclui *introduce, get, give, monitor, contain, begin*. Verbos frasais, tais como *clean up, set up, move away, step up, hand over, take steps, shut down*, são particularmente difíceis para esses dois sujeitos (vide protocolo mais adiante). As tentativas mal-sucedidas de solução mostram que a combinação das duas partes dos verbos não foram suficientes para o reconhecimento de seus significados. Em todos esses exemplos, a presença de outras palavras desconhecidas na mesma oração os impediram de rejeitar os sentidos errôneos.

Além disso, uma análise das palavras desconhecidas mostra que o LIA desconhece palavras que poderiam ser classificadas como de média e baixa frequência (*ills, timber, profligate*), dentre outras, pouco numerosas no texto-base, enquanto o LIO e o LIE, além dessas, demonstram dificuldade também no reconhecimento de palavras de alta frequência, básicas, tais como *bring, talk, reach, clean without, every, unit, times, should*, além de muitas outras, como *set, take*, que entrariam na formação de verbos frasais, confirmando o que se observou no estudo I. Enquanto LIO e o LIE desconhecem substantivos, verbos e adjetivos, o LIO e o LIE também desconhecem, além disso, palavras de função, como preposições, advérbios, quantificadores, dentre outros. A falta de palavras básicas, ou de um vocabulário que pode ser denominado *procedimental (procedural)* dificulta enormemente a tarefa de inferência e de construção do sentido desses leitores, uma vez que palavras *básicas* são difíceis de serem inferidas, pois em geral não se relacionam a um assunto ou conteúdo específico.

É interessante notar, pelos comentários em seu protocolo, que os leitores, principalmente o LIO, têm uma noção clara das palavras desconhecidas/conhecidas. Uma *escala de familiaridade* poderia ser *deduzida* a partir de alguns trechos do protocolo do LIO a seguir:

"*reach* eu não sei o que que é..."; "*unscheduled, Swansea* também não sei"; "Bom eu sei o que é ... *clean*. *Clean* se não me engano é *apagar* Agora *clean up* é *pra cima apagar pra cima* eu não entendi"; "Eu não entendi muito bem o que significa *joint venture*

tá?"; Bom eu deparei com uma palavra que eu não conheço é *announced* (...) não sei o que que é"; "Bom no próximo parágrafo que eu tô lendo aqui eu não entendi qui o que é a palavra /•ut/ *shut* .. é *shut down* certo? (... ..)"; "Uma palavra aqui também eu bom ... também eu (...) Bom, eu já vi essa palavra mas eu não lembro o significado dela, que é /BoutX/ *both* certo? E tem a palavra seguinte que é também /aBRoudX/ *abroad* que eu também não sei o que que é, nunca vi na vida (... ..)"; "Tem outra palavra aqui também que eu não entendi nunca vi na vida: *entrepreneurial* certo? Não sei o que que é (...); "Eu não sei o que quer dizer *bring* também esqueci Eu conheço essa palavra mas esqueci o significado".

Esses diversos graus de familiaridade com as palavras sugerem que o conhecimento de significados de palavras pode ser representado por uma escala de familiaridade, confirmando, mais uma vez, através dos dados também deste estudo de processo, a validade de uma escala ou lista de verificação (*checklist*) como aquela usada na fase I, enquanto instrumento para avaliação do vocabulário.

Um foco na dimensão de uso propriamente dita, objetivo deste segundo estudo, mostra os efeitos dos diferentes níveis de competência, acima descritos, no desempenho dos sujeitos na construção do sentido do texto.

As várias confusões com formas de palavras "parecidas" encontradas nos protocolos dos LIO e LIE, sugerem que, para esses leitores, muitas palavras são conhecidas apenas superficialmente. O longo protocolo do LIE abaixo mostra suas idas e vindas ao texto, assim como suas incertezas e tentativas de adivinhação das palavras *run* (talvez chuva?), *change* (talvez troca?), dentre outras:

The /eKolxgkka/ ecological /KRzXs/ crisis in the Soviet Union é .. ecologia .. ecologia .. /KRzXs/ *crisis* .. não sei não sei o que que é /KRzXs/ *crisis* não sei .. na União Soviética .. *is* .. /SoM/ *some* .. um ponto .. *change* não sei /KRzXs/ também não na União Soviética .. vou voltá né esse texto 'tá difícil *The /eKolxgkka/ ecological* .. a ecologia na União Soviética .. a um ponto .. como que .. foi muito .. a vida do povo na União Soviética .. a um ponto .. vão falar de ecologia .. a União Soviética é um ponto que tem muito a ver com a vida .. do povo soviético certo? *In danger danger* .. deve sê o perigo deve sê o

perigo o perigo o perigo é .. está em perigo O povo está em perigo o povo soviético está em perigo Então um ponto .. é muito .. o soviético povo está em perigo .. o povo soviético está em perigo e .. a troca não .. homem .. próprio não sei .. uma troca não .. 'pera aí economia correndo? /HCN/ run /HCN/ run chuva chuva can pode /SoLvX/ solve solve não sei .. é isso solver solver a /KRxzXs/ crisis .. crisis é chave .. você tem que descobrir o que é crisis .. crisis é palavra chave .. aqui volta lá /eKolXgXkal/ ecological /KRxzXs/ crisis .. crisis ah crisis .. crisis de crise ahn crise .. crise política crise ecológica crise econômica Ah agora sim ecological crisis .. ecologia a crise na ecologia na União Soviética Ah se descobri o que é crise agora .. a um ponto .. onde .. ela .. eles .. muito .. o povo .. perigo Então o povo soviético em perigo .. e cada troca de seus .. economia chuva .. pode solver a crise Então econômica o modo econômico pode diminuir a crise né pode solver a crise .. Ah eu estou aqui crise ..

Alguns problemas identificados no estudo da fase I, tais como associação da palavra a apenas um significado, também são encontrados nos dados de protocolo. O protocolo do LIO a seguir ilustra esse aspecto:

...Bom acho que agora eu descobri /GRextestX/ greatest Deve ser alguma coisa muito (...) o maior problema (...) Bom eu deparei com uma palavra que eu não conheço é /aNauNsedX/ announced (...) não sei o que que é (...) Bom tem uma palavra aqui chamada assim /baunDerXs/ boundaries certo? Bom eu sei que /baunDerXs/ boundary deve ser contorno certo? /baunDerXs/ boundaries eu não sei o que que é Pode ser um falso cognato certo? (...) Bom o que ele diz aqui no nono parágrafo [referindo-se ao oitavo] é o seguinte que a república russa é .. contém a maioria dos fontes naturais da União Soviética e devido a isso é .. tem os maiores problemas do meio ambiente ambientais né? Mas o governo da república certo? Não sei o que é que é /LedX/ led Sei que /LedX/ led é a luzinha que acende no drive só isso Tá tem é .. /alRedX/ already Sei que /RedX/ /RedX/ ready é pronto tá? pronto certo? /alRedX/ already não sei o que que é Começou com pequenos com passos pra remediar esse problema certo? Quer dizer são os problemas ambientais tá? Ele recentemente é .. anunciou sua intenção de fazer um controle de todas as fontes naturais certo? com /baunDerXs/ boundaries que eu não sei o que que é certo?

Bastante interessante é a tentativa de inferência de *boundaries* pelo LIO. Embora explicitamente conheça a palavra *boundary* como *contorno*, não consegue, através do contexto da oração e do parágrafo, fazer uma adaptação para *fronteira* ou *limite*, mais adequado para a construção do significado da sentença. Essa falta de flexibilidade, que poderia ser explicada pela associação da palavra a apenas um significado, também é observada em outra pausa neste parágrafo, mais especificamente, naquela causada pela palavra *led*, tendo sido também freqüente em outros parágrafos. É curioso notar que, nessa situação, o significado explicitado pelo leitor não foi o do verbo *lead*, mais básico ou *core*, mas um termo mais especializado, ou seja, o substantivo *led*, que é associado à "luzinha do *drive* do computador". O conhecimento de um significado mais remoto ou mais técnico, sem o conhecimento do mais comum ou generalizável não parece incomum em situações de LE, dado o tipo de exposição que esses leitores têm à língua. A exposição prematura e formal à leitura, principalmente de textos especializados ou técnicos, a necessidade de uso do computador (deve-se lembrar que os programas eram, até pouco tempo, todos em inglês), distanciados de situações de vivência da língua, poderiam explicar esse tipo de inversão. Juntamente com essa situação, observa-se uma visão estreita de competência lexical, em que a palavra é vista como portadora de um único significado, não incentivando o leitor a buscar outras possibilidades para a confirmação de suas hipóteses.

O protocolo do LIA abaixo, referente ao mesmo parágrafo do protocolo do LIE, mostra claramente a diferença nos desempenhos dos dois leitores.

Bom vou começá lê o texto ((LB)) .. *The ecological /KRxzXs/ crisis in the Soviet Union is reaching the point where the very life of the Soviet people is in danger .. and only a change in the way the economy is run can solve the /KRxzXs/ crisis. According to /maxka/ Mikhail /maxka/ Mikhail Krotov an economist at the Leningrad Mechanical Institute the ecological crisis is a direct result of the Soviet Union's economic crisis ..* Bom, esse parágrafo basicamente fala que a crise ecológica na União Soviética tá atingindo pontos já .. é .. críticos né que coloca .. a a a vida das pessoas que lá moram em perigo e apenas é .. e somente uma uma mudança em .. nos rumos da economia é que pode pode assim salvá pode solucioná esse problema da crise ecológica de acordo com um economista de um de um

Instituto de Leningrado Essa crise ecológica é um di .. é um resultado direto da crise econômica né existente na União Soviética.

O protocolo a seguir, também do LIA, mostra como esse leitor constrói o sentido, mesmo encontrando palavras desconhecidas. Convém salientar que, nesse caso, a densidade dessas palavras é bastante pequena com relação às conhecidas e a inferência torna-se possível. Apesar de ter inferido erroneamente a palavra *affluent* (*afluente*) não deixou de construir um sentido para o trecho:

((LB)) *Environmental problems in the country have reached the stage where people's health is declining* Krotov said in an unscheduled talk in Swansea *The people of Leningrad for example drink water drawn from Lake Ladoga which is polluted by the effluent from nine pulp and paper plants two dairy complexes and a series of chemical factories on its shores .. Hum .. drink water drawn .. from Lake Ladoga .. polluted by the effluent* *afluente* /PuLpX/pulp .. *nine pulp and paper plants .. não sei o que é* /PuLP/ *pulp plants* fábricas indústrias indústrias de papel? .. *dairy complexes* laticínios deve ser alguma coisa relacionada a isso laticínios *and a series of chemical factories on its shores* *Shores shores* baía costa costa costa Bom o segundo parágrafo .. problemas com relação ao meio ambiente na União Soviética atingiram um estágio .. crítico né como eu havia falado no primeiro parágrafo A saúde das pessoas 'tá em declínio Um exemplo é citado aí: as pessoas né de Leningrado bebem água né que .. é poluída de .. bebem água *drawn from the lake* .. bebem a água do lago .. que é poluída pelo seu *afluente* (...) é é poluída por um *afluente* quer dizer por um rio que deságua nesse lago que é poluído por várias indústrias de papel laticínios uma série de fábricas ou indústrias químicas que estão instaladas na costa né desse rio que afluí no lago.

Todas as evidências parecem mostrar, portanto, que a competência lexical mais desenvolvida — tanto qualitativamente, ou com relação à sua profundidade, como quantitativamente, ou relativa ao número de palavras conhecidas e à natureza dessas palavras (principalmente palavras básicas ou de alta frequência), como aquela apresentada pelo LIA deste estudo — permite uma maior sensibilidade às restrições semânticas, sintáticas, fonológicas e morfológicas do texto, possibilitando a construção de um contexto de forma automática de grande parte das palavras, e também a construção de um contexto

de forma controlada ou semi-controlada, quando necessário. A importância de palavras básicas está, principalmente, no fato de servirem como conhecimento *procedimental*, permitindo o acesso às palavras de baixa frequência. Esses dois aspectos de uma competência lexical mais desenvolvida, portanto, oferecem ao leitor mais proficiente deste estudo a possibilidade de usar sua capacidade de processamento não apenas para a decodificação, mas principalmente para a construção do sentido ou processamentos de nível mais alto.

Leitores com uma competência lexical menos desenvolvida, como aquela apresentada pelos LIO e LIE, por outro lado, apresentam sérias dificuldades de processamento automático, tendo que construir um contexto de forma controlada para a inferência de cada palavra. Entretanto, há grandes dificuldades também na construção desse contexto, uma vez que sua competência lexical insuficiente não oferece outros recursos, além dos lexicais, ou seja, aqueles que levam em conta apenas a forma ou som das palavras. Como é muito grande o número de palavras desconhecidas, torna-se necessário, com muita frequência, o estabelecimento de contexto de forma controlada para a inferência dessas palavras, sobrecarregando sua capacidade de processamento. As limitações da atenção humana e da capacidade da memória, assim como as demandas cognitivas exigidas do leitor com uma competência lexical menos desenvolvida, portanto, poderiam explicar porque bons leitores em LM não conseguem, frequentemente, aplicar suas habilidades de leitura a textos em L2 ou LE, ou porque uma compensação através dos recursos da LM nem sempre é possível.

4. Conclusões e implicações

Neste trabalho, o objetivo foi apresentar uma caracterização da competência lexical de leitores brasileiros aprendendo a ler em inglês. Embora se tenha afirmado frequentemente que uma das maiores dificuldades, senão a maior, do leitor em LE seja seu vocabulário limitado, não se têm caracterizações mais precisas desse vocabulário. Os dados para esse estudo foram baseados em um conceito *rico* de vocabulário, operacionalizado através de testes no primeiro estudo e como guia para a análise nos protocolos de pausa no segundo estudo.

Sem dúvida alguma, os resultados desta investigação, tanto os da

fase I como os da fase II, parecem mostrar que a competência lexical dos leitores deste estudo, com algumas exceções, é vaga e imprecisa, apresentando lacunas tanto do ponto de vista quantitativo, ou seja, com relação ao número de palavras conhecidas, como qualitativo, isto é, com relação à profundidade desse conhecimento. O que mais chama a atenção, entretanto, nessa competência, é que o leitor em LE, em geral, desconhece não apenas palavras de baixa frequência, como o falante nativo, mas, principalmente, palavras de alta frequência, ou seja, aquelas que fazem parte de um vocabulário *básico* ou *procedimental*, impondo uma espécie de nível limiar³, trazendo dificuldades para a inferência de outras palavras e, conseqüentemente, para a construção do sentido do texto.

Uma contribuição para o ensino de LE, ou mais especificamente, para o ensino da leitura, que se pode depreender deste trabalho, é a necessidade de se criarem condições e oportunidades, em sala de aula, para a aquisição desse nível limiar. Embora os resultados não permitam a sua caracterização em termos absolutos ou quantitativos, indicando *quais* e *quantas* seriam as palavras suficientes para sua composição, oferecem algumas sugestões com relação às características dele. Esse limiar parece incluir duas dimensões, uma quantitativa ou de número de palavras desconhecidas e, principalmente, uma dimensão qualitativa, isto é, de profundidade desse conhecimento, envolvendo os vários aspectos de uma competência lexical definida em termos de um conhecimento *rico* de vocabulário, conforme a proposta deste estudo. As palavras, nesse limiar, seriam, principalmente, palavras básicas de alta frequência, de uso generalizável. Elas constituiriam um vocabulário procedimental (*procedural*), através do qual seria possível, através de inferências, se chegar a outras palavras de baixa frequência. A grande dificuldade de inferência de palavras de alta frequência é que elas não estão relacionadas a um tema ou tópico, como já afirmei anteriormente. Algumas sugestões nesse sentido podem ser encontradas em Robinson (1980) Hutchinson & Waters (1981) e Carter (1982, 1986, 1987a). Esse eria, portanto, um vocabulário generalizável e estratégico, que

³ Essa noção, nesse caso específico, tem sua origem no estudos sobre a Hipótese do Curto Circuito (*Short-Circuit Hypothesis*), referindo-se a um teto de conhecimento linguístico (*language ceiling*) ou nível mínimo de conhecimento linguístico necessário para a compreensão.

poderia ser usado tanto na leitura de textos gerais como acadêmicos, facilitando a inferência de outras palavras e a construção do sentido do texto e também permitindo ao leitor aprender outras palavras através delas, ampliando, assim, seu vocabulário de forma mais independente, nos moldes do que ocorre em situações de LM.

Uma outra contribuição deste trabalho diz respeito à seleção de textos usados tanto na avaliação da compreensão em leitura como no ensino. Embora a legibilidade, em um modelo de leitura interativo, não possa ser determinada, em princípio, apenas pelos elementos do texto, mas como resultado da interação das restrições do texto e conhecimentos do leitor, poderia indicar algumas áreas de dificuldades em potencial. Enquanto a seleção de textos em geral é baseada no número de palavras de baixa frequência, segundo um princípio de dificuldade definida por falantes nativos, este estudo parece mostrar que, para alunos em uma situação de LE, também as palavras de frequência média e alta são importantes enquanto fontes de dificuldade em potencial.

A necessidade de se repensar a ênfase a ser dada à competência lexical no ensino de leitura em inglês em um contexto de LE também pode ser considerada uma implicação deste estudo. Apesar da frequência com que os leitores em LE salientam sua importância e apontam as suas dificuldades, essa competência não é geralmente reconhecida e muito menos enfatizada em abordagens que se baseiam em modelos de leitura em LM, como, por exemplo, a abordagem instrumental. Não querendo negar a importância de se incentivar o aluno a conviver com a noção de vagueza e sentidos aproximados de palavras ou ignorar significados de palavras, restringindo consultas a dicionários e interrupções durante a leitura — procedimentos esses, sem dúvida alguma, importantes sob o ponto de vista da leitura — não se pode deixar de salientar que eles acabam por subestimar a importância do conhecimento de vocabulário, não permitindo um insumo adequado e significativo para o seu desenvolvimento. Não se pode esquecer que o único contato que os alunos têm como a língua em uma situação de LE é, geralmente, durante as poucas horas de aula de leitura.

Quando saliento a necessidade de um ensino de vocabulário mais sistemático não estou me referindo à volta ao ensino de língua

tradicional que, em geral, caracteriza-se pela fragmentação, descontextualização ou a visão de vocabulário através de listas de palavras e glossários. Além de evidências esparsas e muitas vezes conflitantes, não se encontram, na literatura sobre vocabulário, indicações mais precisas de como seria um ensino de vocabulário que conduza à compreensão. Embora o presente estudo não ofereça contribuições diretas com relação à abordagem a ser usada no ensino de vocabulário nas aulas de leitura, traz algumas indicações que poderão ser exploradas e avaliadas em um contexto de LE. Essa proposta, que consistiria numa ampliação da base lingüística através de um foco em palavras, é consistente com a afirmação de Nagy & Herman (1987:33) quando afirmam que

O ensino do vocabulário que realmente faz uma diferença na compreensão geralmente tem algumas das seguintes características: exposições múltiplas às palavras ensinadas, exposição às palavras em contextos significativos, informação rica e variada sobre cada palavra, o estabelecimento de ligações entre as palavras ensinadas e a própria experiência e o conhecimento prévio do aprendiz (...).

O ensino que gostaria de sugerir apresenta as características acima mencionadas, indispensáveis para a fluência de acesso às palavras ou a um reconhecimento automático que parecem caracterizar bons leitores. Esse ensino extrapola a distinção receptivo/produtivo³, uma vez que parece ser necessário muito mais conhecimento para a leitura do que essa dicotomia quer reconhecer. Ele não apenas incentiva o leitor a fazer uso dos recursos de que já dispõe, encorajando sua independência e participação ativa através de uma conscientização sobre a leitura e sobre sua experiência com o uso

⁴A dicotomia receptivo/produtivo, substituindo a dicotomia ativo/passivo, tem sido usada na literatura de ensino/aprendizagem de vocabulário geralmente para representar dois tipos diferentes de conhecimento ou dois estágios de desenvolvimento desse conhecimento: conhecer uma palavra para uso produtivo pressupõe conhecê-la melhor, para usá-la na produção e não apenas reconhecê-la ou entendê-la (uso receptivo). Entretanto, tem sido criticada por não ser suficientemente clara a extensão e o significado da lacuna entre receptivo e produtivo.

da LM, mas também procura ampliá-los, promovendo a aprendizagem de outras palavras de uma maneira mais abrangente.

Esse tipo de conhecimento lexical mais profundo, voltado para um reconhecimento automático das palavras, não elimina a necessidade de inferência ou de compensações através de outras fontes, como o contexto, mas leva a inferência a outros níveis. O que quero dizer é que leitores com conhecimento lingüístico vago e impreciso, como o dos sujeitos de nível iniciante e intermediário deste estudo, tentam construir um contexto, que é usado com uma função muito mais ampla: não apenas para decidir quais dos significados de uma palavra conhecida, por exemplo, é a mais apropriada à situação em questão, mas para levantar hipóteses e inferir o significado da palavra, o que, quando o número de palavras desconhecidas é muito grande, torna-se ineficiente. Uma base sólida de palavras de alta freqüência parece ter sido fundamental para que o leitor intermediário avançado conseguisse inferir outras palavras menos freqüentes. Embora não esteja propondo um controle das palavras a serem ensinadas ou a simplificação de textos, o que não é consistente com uma visão de leitura enquanto interação, considero importante, principalmente num primeiro estágio, salientar a construção de uma base lingüística ou mais especificamente lexical, nos moldes acima descritos, que se iniciaria com palavras mais freqüentes, básicas ou centrais (*core words*).

Os resultados deste estudo também parecem sugerir que desse grupo fizessem parte palavras que compõem conjuntos fechados ou palavras de função, como pronomes, preposições, conjunções, dentre outras. Uma ênfase especial aos verbos e principalmente a alguns verbos frasais (*two-word verbs*), também parece muito importante para facilitar a segmentação dos períodos mais longos. Sem dúvida alguma, pesquisas futuras seriam necessárias para confirmar a importância desses aspectos e sugerir outros.

Entretanto, é relevante ressaltar que, quando menciono a necessidade da construção de uma base lingüística, não estou sugerindo que ela seria conduzida de forma isolada. A ênfase nessa questão não elimina a necessidade de se considerarem outras frentes, como já mencionei anteriormente. Além desses aspectos lingüísticos, os resultados deste trabalho também parecem indicar a necessidade de

ênfatisar a importância de recursos textuais, tais como a organização do texto, sua superestrutura, aparentemente não utilizados de forma explícita pelos sujeitos leitores. Todos esses aspectos, entretanto, não eliminam a necessidade de se desenvolver no aluno uma postura mais ativa frente aos problemas que enfrenta no processo de leitura, instrumentalizando-o para identificar de forma mais eficiente a origem de seus problemas de compreensão — sejam eles lingüísticos ou extra-lingüísticos.

(Recebido em 17/10/96. Aprovado em 14/11/96)

Referências Bibliográficas

- CARTER, R. (1982) A note on core vocabulary. *Nottingham Linguistic Circular* 2, 2: 39-51.
- _____ (1986) Core vocabulary and discourse in the curriculum: a question of subject. *RELC Journal* 17,1: 52-67.
- _____ (1987) Is there a core vocabulary? Some implications for teaching. *Applied Linguistics II*, 2, 8.
- CAVALCANTI, M. do C. (1983) The pragmatics of FL reader-text interaction: key lexical itens as source of potential reading problems. Lancaster: Universidade de Lancaster. Tese de doutorado.
- _____ (1989) *Interação leitor-texto: aspectos de interpretação pragmática*. Campinas: Editora da UNICAMP.
- DALE, E. (1965) Vocabulary measurement: techniques and major findings. *Elementary English* 42: 895-901.
- HOEY, M. (1976) Signalling in discourse. Discourse Analysis Monograph nº 6, *English Language Annals* 9: 117-29.
- HUTCHINSON, T. & A. WATERS (1981) Performance and competence in ESP. *Applied Linguistics* 2: 39-49.
- NAGY, W.E. & P.A. HERMAN. (1987) Breadth and depth of vocabulary knowledge: implications for acquisition and instruction. In: M.G. McKewon and M.E. Curtis (eds.) *The nature of vocabulary acquisition*. Hillsdale: Lawrence Erlbaum.
- NATION, I.S.P. (1984) *Vocabulary list: words, affixes and stem*. New Zealand: English Language Institute, Wellington: Victoria University of Wellington.

- OLSHAVSKY, J.E. (1976-7) Reading as problem solving: an investigation of strategies. In: *Reading Research Quarterly* 12: 654-74.
- READ, J. (1987) Towards a deeper assessment of vocabulary knowledge. Comunicação apresentada no 8th World Congress of Applied Linguistics, Sydney, Australia.
- RICHARDS, J. (1976) The role of vocabulary teaching. *TESOL Quarterly* 10, 1.
- ROBINSON, P.J. (1980) *ESP (English for specific purposes)*. Oxford: Pergamon Press.
- _____ (1989) Procedural knowledge and language learning. *Journal of Pragmatics* 13: 523-46.
- SCARAMUCCI, M.V.R. (1990) O resumo e a compreensão em leitura em língua estrangeira. *Trabalhos em Lingüística Aplicada* 15. Campinas: Universidade Estadual de Campinas, 65-86.
- _____ (1995) O papel do léxico na compreensão em leitura em LE: foco no produto e no processo. Tese de doutorado inédita, UNICAMP, Campinas.
- WALLACE, M. (1982) *Teaching vocabulary*. London: Heinemann.

Apêndice A

Texto-base para testes de vocabulário
(palavras avaliadas em negrito)

GALACTIC BIRTH?

A surprise for scientist

Astronomers have long believed that galaxies, **clusters*** that usually contain billions of stars, were all formed shortly after the Big Bang, the cataclysmic explosion some 15 billion years ago that **spawned** the universe. But that conviction was **shaken** last week when scientists announced that they had found evidence of a cosmic version of gestation: a galaxy preparing for birth. Said James Gunn, a Princeton University astrophysicist, after the announcement: "This is the Rosetta stone of galaxy formation." The apparent galactic embryo – actually a **massive, disk-shaped cloud** of hydrogen gas – was

discovered **fortuitously** last **spring** by Cornell University astronomer Martha Haynes and her colleague Ricardo Giovanelli, when they were monitoring signals in outer space with the 1,000-ft. radio telescope at Arecibo, Puerto Rico. While **focusing** the telescope on what they thought was **empty** space in order to calibrate it, the astronomers **picked up** a signal **pattern resembling** that emitted by galaxies. The invisible cloud -- estimated to be ten times as large as the Milky Way -- **loomed fairly close**, astronomically speaking: 65 million light-years from earth. Since a light-year is the distance light travels in a year, the scientists were **receiving** signals from the cloud as it **appeared** 65 million years ago. Because it apparently contained no stars, the scientists concluded that they were observing a galaxy about to be born. Said Giovanelli: "This cloud indicates that galaxies can form **slowly throughout** the history of the universe and are not something that happened during some magical period in the distant past". *Time Magazine*, 11/9/1989

Apêndice B

Texto-base para tarefa de introspecção

A CAPITAL CURE FOR THE SOVIET UNION'S ECOLOGICAL ILLS

Stephanie Pain

The ecological crisis in the Soviet Union is reaching the point where "the very life of the Soviet people is in danger", and only a change in the way the economy is run can solve the crisis. According to Mikhail Krotov, an economist at the Leningrad Mechanical Institute, the ecological crisis is a direct result of the Soviet Union's economic crisis.

Environmental problems in the country have "reached the stage where people's health is declining", Krotov said in an unscheduled talk in Swansea. The people of Leningrad, for example, drink water drawn from Lake Ladoga, which is polluted by the effluent from nine pulp and paper plants, two dairy complexes and a series of chemical factories on its shores.

Krotov insisted that industry is not to blame. "Narrow-minded people blame [the crisis] on technological progress itself, and this has

led to many industrial activities being banned for ecological reasons." The real problem, he explained, is the "lack of market mechanisms for the rational use of natural resources".

Under the current economic system, where the state has a monopoly on all natural resources, timber, oil and gas and so on are virtually free. This has encouraged profligate consumption by industry. Krotov estimates that for every unit of national income, the Soviet Union uses two to three times as many resources as other countries. It should be possible to bring consumption into line with other countries, without reducing productivity.

The best way to cure industry of its wastefulness is to charge the going rate for natural resources, maintains Krotov. He also proposes the addition of an "ecological supplement" to the price, the surcharge reflecting the scarcity of the resource or the damage done to the environment. Under the old system, research into the science and technology needed to clean up industry also stagnated. "It was just not profitable for factory managers to introduce clean, safe technology," said Krotov.

Krotov's cure for the Soviet Union's environmental ills is to privatise at least half of the state factories to introduce competition among producers. This would involve existing companies setting up new enterprises and getting involved in joint ventures with foreign firms. But the Soviet Union must move away from heavy industry and step up production of consumer goods, he said. Most importantly, the country must join the international monetary system so that the rouble becomes convertible.

A second vital component of Krotov's scheme is to give local governments control over the resources in their region, by handing over what is now state property. This would give local governments a direct interest in preventing pollution and protecting their local environment. "People with the greatest stake in the environment are the local people" he says. "Ecological standards and rules can only be applied effectively when the local authorities are the main agent for monitoring and enforcement."

The Russian republic contains most of the Soviet Union's natural resources — and has the greatest environmental problems. But the

government of the republic, led by Boris Yeltsin, has already begun to take steps to remedy its own problems. It recently announced its intention to take control of all the natural resources within its boundaries. Yeltsin has introduced a 400-day programme to revitalise the Russian economy, with backing from President Gorbachov. And from 1 January, an innovative "polluter pays" policy comes into force.

The new law means that about 100 of the 500 major industrial concerns in the Leningrad region will have to shut down because they cannot meet the new standards set by the Russian authorities. An optimistic Krotov believes that the extra money coming into the republic from the export of its resources both abroad and to other parts of the Soviet Union will help to reduce unemployment by providing funds to set up new, cleaner industries. "Economic reform in the Russian republic will make it possible to harmonise entrepreneurial activity and the rational use of resources with the protection of the environment," said Krotov.

While these reforms will bring benefits to the Russian federation, they will cause problems in other parts of the country and in the countries of Eastern Europe, which, until now, have had almost unlimited access to raw materials for virtually nothing. *New Scientist* 1/9/1990