

5

Metodologia

Neste capítulo iremos examinar o corpus, os procedimentos e critérios de análise do trabalho que desenvolvemos.

5.1

O corpus

O corpus consiste de uma coleção intitulada Cadernos Didáticos de EJA, disponibilizada pela Secretaria de Educação Continuada, Alfabetização e Diversidade, criada pelo MEC para implementar políticas públicas voltadas para pessoas com 15 anos ou mais que não completaram o Ensino Fundamental. Foi elaborada por uma equipe de professores da Unitrabalho, uma Rede Nacional de Universidades que apóia a busca de melhores condições de vida e trabalho. Em uma parceria com o MEC e a Secretaria de Educação Continuada, essa equipe planejou e desenvolveu a Coleção Didática Cadernos de EJA.

A coleção é de 2006 e composta de 13 cadernos de textos para os alunos (Cadernos do Aluno) e 13 cadernos de atividades sugeridas para o professor (Cadernos do Professor), além de um caderno metodológico com a concepção pedagógica do material e com perguntas mais frequentes feitas por professores que acompanharam as oficinas pedagógicas, realizadas nas cinco regiões do Brasil. Esses cadernos foram concebidos tendo como público alvo alunos da 1ª a 8ª série do Ensino Fundamental, atualmente 2º a 9º ano.²

Os Cadernos do Aluno da Coleção Cadernos de EJA apresentam uma coletânea de textos de gêneros variados organizados em treze temas em torno do tema trabalho, conforme indicado a seguir:

Caderno 1: Cultura e Trabalho

Caderno 2: Diversidades e Trabalho

Caderno 3: Emprego e Trabalho

² Observe-se que a coleção foi elaborada em 2006, portanto antes da última LDB e da mudança do conceito para ano.

Caderno 4: Economia Solidária

Caderno 5: Globalização e Trabalho

Caderno 6: Juventude e Trabalho

Caderno 7: Meio Ambiente e Trabalho

Caderno 8: Mulher e Trabalho

Caderno 9: Qualidade de Vida, Consumo e Trabalho

Caderno 10: Segurança e Saúde no Trabalho

Caderno 11: Trabalho no Campo

Caderno 12: Tecnologia e Trabalho

Caderno 13: Tempo Livre e Trabalho

Dentre os gêneros presentes nos cadernos acima citados destacam-se os seguintes: artigos de opinião, artigos de divulgação didática (em páginas da Web), poesias, poesias em letras de música, leis e normas, etc. A coleção explora o uso de imagens (fotografias e ilustrações, principalmente) como forma de maximizar o sentido do texto.

Não há um ordenamento prévio dos livros do aluno em sequência de graus de dificuldade, seja em série ou ciclo. Nos Cadernos do Professor, é dito que o material pode ser usado por todo o Ensino Fundamental. A articulação entre os livros e a organização das aulas deve ser discutida em um planejamento coletivo dos professores. Assim, conforme expresso nas orientações ao professor, um professor de início de primeiro segmento poderá mostrar um texto de qualquer um dos 13 Cadernos do Aluno apenas para que seus alunos tenham noção do que é um texto (já que não são crianças) e trabalhar apenas com as palavras do título, por exemplo.

Os Cadernos do Professor apresentam atividades, com indicação, por meio de cores de margem diferente, do nível a que se destinam - amarela, para o 1º segmento; vermelha, para o 2º segmento; laranja, para os dois segmentos. Cabe ao professor decidir quais partes dos cadernos vai usar, de acordo com o grau de dificuldade dos textos e das atividades propostas, e de acordo com a série que leciona e a facilidade que seus alunos tenham para ler e compreender.

O Caderno Metodológico, conforme já dito, indica a concepção de leitura que orientou a elaboração do material didático. Essa concepção vê “a leitura como um processo criativo, crítico e dialético, envolto pela exigência de uma série de

habilidades cognitivas” (página 20). Como os alunos possuem diversos graus de capacidade leitora, o professor atua como mediador incentivando cada um a manifestar suas impressões sobre o texto, baseados em seus conhecimentos de mundo e construindo assim um sentido compartilhado por todos:

Um objetivo importante que esse material procura atingir é servir de MOTIVO para desafiar os educandos a avançarem no seu processo de conhecimento e ação sobre a realidade em que vivem. Desafiar o aluno a explorar um texto que ele ainda não está apto a compreender sozinho leva-o a perceber melhor o significado do processo pedagógico. O aluno tenta ler um texto e vê que é difícil. Diante desse obstáculo, ele pode recuar ou, com a ajuda do(a) professor(a) e de colegas, pode compreender o texto, por meio de aproximações sucessivas e releituras. Aos poucos, por meio dessa atividade mediada pela ajuda dos outros, ele vai internalizando as capacidades necessárias para entender os conceitos e estruturas dos textos mais complexos e, finalmente, começa realizar essa leitura de modo independente.

Já na página 24 do Caderno Metodológico, há uma referência à proposta pedagógica dos Cadernos de EJA:

Portanto, na proposta pedagógica dos Cadernos de EJA, o texto assume um papel central, estabelecendo conexões entre as experiências vividas pelos educandos no mundo do trabalho, as reflexões de autores que abordaram esse tipo de experiência na sua produção escrita e os conteúdos escolares que possibilitam embasar a análise dessas vivência em conceitos científicos e fundamentos teóricos.

5.2

Procedimentos de análise

Foram analisados, para fins desta pesquisa, tanto os textos dos Cadernos do Aluno como a relação de atividades sugeridas nos Cadernos do Professor. Para isso, procedeu-se inicialmente a uma caracterização dos gêneros dos textos dos 13 Cadernos do Aluno. Ao realizar esta tarefa, baseamo-nos, em especial, no trabalho de Marcuschi (2008), que, conforme apresentado na resenha da literatura, classifica os gêneros textuais levando em consideração a questão dos domínios discursivos e do suporte.

Partindo-se do pressuposto de que o significado do texto é construído a partir da articulação entre material verbal e visual, foram analisados os recursos multimodais presentes Cadernos do Aluno. A terminologia empregada na

classificação dos elementos visuais seguiu basicamente a informação disponibilizada no próprio corpo dos textos, em que legendas sinalizavam o tipo de material (ilustração, fotografia, mapa, etc.)

A legibilidade dos textos dos Cadernos Didáticos foi também objeto de análise, sendo aferida pela comparação dos índices obtidos com a submissão dos textos à ferramenta Coh-Metrix-Port (Aluisio et al.,2008), como detalharemos na seção 4.2.1.

Com base nos Cadernos do Professor, tivemos acesso à relação das atividades sugeridas para trabalhar cada um dos textos em sala de aula, identificando o segmento a que se destinam.

Todos os resultados foram armazenados em tabelas para serem comparados. Uma tabela geral foi confeccionada com o título e autor do texto, o segmento do Ensino Fundamental (1 ou 2) para o qual é recomendado, o gênero textual a que pertence, os recursos multimodais presentes na página do texto e a atividade sugerida para o professor trabalhar o texto com os alunos. Esse procedimento foi elaborado para todos os textos dos 13 Cadernos Didáticos do Aluno e do Professor. No anexo I, apresentamos, como exemplo, a tabela geral referente a um dos Cadernos.

Foi elaborada também uma tabela com os registros das frequências em que os gêneros textuais apareceram nos textos. Buscamos verificar tanto o total de ocorrências de cada gênero por caderno didático como o total de ocorrências de cada gênero nos treze cadernos.

Uma outra tabela foi confeccionada com a lista das atividades sugeridas, em que se observou a frequência de cada tipo de atividade: escrever uma carta, preencher um cheque, elaborar um currículo e assim por diante.

A multimodalidade foi tema de outra tabela em separado onde registramos os gêneros visuais, indicando o número total de ocorrências em cada caderno didático e também no conjunto de cadernos. Foram registrados tanto os recursos multimodais que acompanham os textos, bem como aqueles que estão integrados à composição gráfica do texto, como, por exemplo, fundo de página com ilustrações, fotografias, cores diferentes, etc.

Os resultados dessa análise serão reportados no capítulo 6. A seguir, detalhamos os procedimentos de submissão dos textos à ferramenta Coh-Metrix-Port

5.2.1

Coh-Metrix-Port – procedimentos de submissão dos textos e parâmetros de legibilidade

5.2.1.1

Procedimentos de Submissão

A Ferramenta Coh-Metrix-Port foi desenvolvida na USP/São Carlos a partir do Coh-Metrix da Universidade de Memphis. Essa ferramenta faz parte do Projeto PorSimples, nome pelo qual ficou conhecido o projeto que envolve a Simplificação Textual do Português para Inclusão e Acessibilidade Digital (Aluisio et al., 2008). A proposta é desenvolver tecnologias para facilitar o acesso à informação dos analfabetos funcionais e de outras pessoas com deficiências cognitivas, como afasia e dislexia.

A figura a seguir refere-se à tela inicial do Coh-Metrix-Port, onde o usuário pode se cadastrar ou, se já for cadastrado, pode acessar o sistema para submeter um texto à ferramenta ou ver os textos que já foram analisados e que ficam armazenados para consulta.



Figura 2: Página inicial do Coh-Metrix-Port

Qualquer pessoa pode se cadastrar no sistema acessando o link da Página Inicial que vai remeter à página seguinte:

Figura 3: Página de cadastramento do Coh-Metrix-Port

Quando o usuário já está cadastrado, o texto que será submetido à ferramenta precisa ser preparado para o processo. Todas as ilustrações, fotografias e outros recursos multimodais precisam ser apagados e o texto precisa ser convertido para Word, se estiver em PDF, por exemplo.

O usuário cadastrado deve preencher os campos requeridos com os seguintes metadados: título do texto, autor, fonte e o gênero. Em seguida, colar o texto e clicar em *analisar*. As figuras a seguir ilustram estes procedimentos:

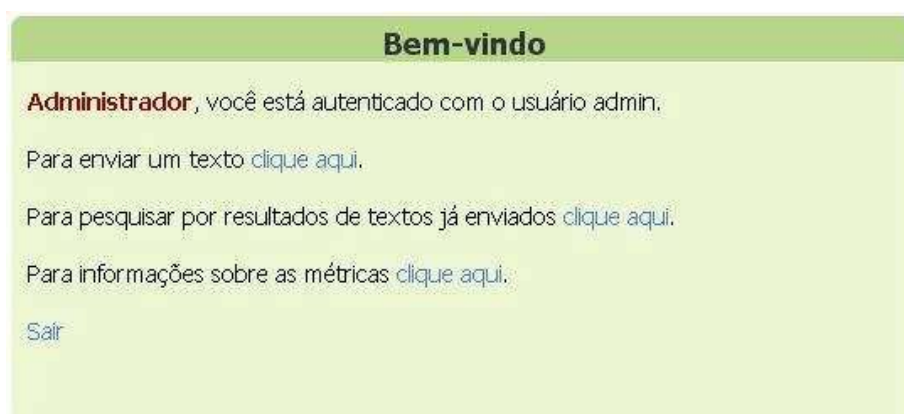


Figura 4: Página que registra o cadastramento do novo usuário e suas opções

Página inicial > [Enviar novo texto](#)
Enviar novo texto
 Daniel de Almeida | Sa

Título *
 os lobos uivam para a lua cheia?

Autor
 Débora Boccacino

Fonte
 CHC

Data de Publicação
 2008

Gênero
 Científico

Texto *
 Por meio dos uivos - e também de rosnadas e pequenas mordidas! -, os lobos que comandam a matilha também instruem os demais, como forma de manter a ordem e a hierarquia. Um macho também pode uivar para atrair a atenção de uma fêmea. Ou vice-versa.

Já os cães, quando estão sosinhos, costumam uivar para chamar a atenção do dono ou quando têm alguma necessidade: fome, sede, para sinalizar alguma dor... Nos gatos, o miado que parece um uivo aparece quando um macho encontra com outro macho. Isso porque, quando dois gatos se encontram, sempre surge uma certa agressividade, seja porque eles vão brigar por uma fêmea ou por um território. Nesse caso, o miado contínuo semelhante ao uivo serve para tentar intimidar o oponente, marcar o território, mas evitando o enfrentamento, a briga propriamente dita. Afinal, se dá para afugentar o adversário só na garganta, para que gastar energia trocando unhas?

Mas se, até aqui, não conhecemos nenhum animal que uive especialmente porque a Lua apareceu no céu, de onde será

Analisar

© 2009 PorSimples | Desenvolvido por: Carolina Evaristo Scarton e Daniel Machado de Almeida Design por: Felipe Vianna Perez | Baseado no Coh-Metrix

Figura 5: Página onde o usuário cola o texto a ser submetido

5.2.1.2

Métricas do Coh-Metrix-Port

As métricas do Coh-Metrix foram selecionadas para serem adaptadas e implementadas para o português pelos pesquisadores da USP/São Carlos. O índice Flesch, que avalia a inteligibilidade de um texto e será detalhado na seção 5.2.1.3, já havia sido adaptado ao português (Martins et al., 1996), mas é um índice superficial, que, por si só, tem pouca valia. Como Aluísio (2009, p.) expõe, “a língua portuguesa já dispõe de várias ferramentas e recursos de Processamento de Língua Natural (PLN) que poderiam ser utilizados para a criação de uma ferramenta que analisasse vários níveis da língua e fosse calibrada com textos de vários gêneros”.

O Coh-Metrix 2.0 de Memphis possui 60 índices divididos em 6 classes que são:

1. *Identificação Geral e Informação de Referência*, que corresponde às informações que referenciam o texto, como título, gênero entre outras;
2. *Índices de Inteligibilidade*, que contém os índices de inteligibilidade calculados com as fórmulas Flesch Reading Ease e Flesch Kincaid Grade Level;
3. *Palavras Gerais e Informação do Texto*, que possui 4 subclasses: Contagens Básicas, Frequências, Concretude, Hiperônimos;
4. *Índices Sintáticos*, que possui 5 subclasses: Constituintes, Pronomes, Tipos e Tokens, Conectivos, Operadores Lógicos e Similaridade sintática de sentenças;
5. *Índices Referenciais e Semânticos*, que possui 3 subclasses: Anáfora, Co-referência e Latent Semantic Analysis (LSA) (Deerwester et al., 1990);
6. *Dimensões do Modelo de Situações*, que possui 4 subclasses: Dimensão Causal, Dimensão Intencional, Dimensão Temporal e Dimensão Espacial.

Foram escolhidas do Coh-Metrix para o Coh-Metrix-Port as classes 3. *Palavras Gerais e Informações do Texto* e suas subclasses, com exceção de Concretude; 4. *Índices Sintáticos* e suas subclasses, com exceção de Similaridade sintática de sentenças e 5. *Índices Referenciais e Semânticos* e suas subclasses, com exceção de Latent Semantic Analysis. Também foram mantidos os metadados (Classe 1) e o índice Flesch (Classe 2).

Para esclarecimento dessas classes, apresentamos, a seguir, o detalhamento de suas subclasses.

Classe 3: *Palavras Gerais e Informação do Texto*

Subclasse - *Contagens Básicas*.

Essa subclasse busca verificar aspectos ligados ao total de ocorrências de itens vocabulares, considerando-se a classe gramatical a que pertencem; número de sentenças e de parágrafos, além de cruzar informações desses elementos. São, portanto, registrados o **Número de Palavras** que o texto possui; **Número de Sentenças**, contadas pelo ponto final de cada parágrafo; **Palavras por Sentença**, calculadas pelo número de palavras dividido pelo número de sentenças; **Número de parágrafos** de um texto, considerando somente a quebra de linha (não

identificações); **Sentenças por Parágrafos**, número de sentenças dividido pelo número de parágrafos; **Sílabas por Palavras de Conteúdo**, número médio de sílabas por palavras de conteúdo (substantivos, verbos, adjetivos e advérbios). Deve-se esclarecer que a ferramenta chama de sentença o que tradicionalmente chamamos de período, visto que, como dito, são identificadas a partir do sinal de ponto final.

Em relação às classes gramaticais, a ferramenta disponibiliza a **Incidência de Verbos**, número de verbos em um texto; **Incidência de Substantivos**, número de substantivos em um texto; **Incidência de Adjetivos**, número de adjetivos em um texto; **Incidência de Advérbios**, número de advérbios em um texto; **Incidência de Palavras Funcionais**, número de palavras funcionais em um texto, ou seja, número de artigos, preposições, pronomes, conjunções e interjeições em um texto.

Subclasse – *Frequências*

A segunda subclasse pertencente à classe 3 é *Frequências*, que é a média de todas as frequências das palavras de conteúdo encontradas no texto. O valor da frequência das palavras é retirado da lista de frequência do corpus Banco de Português (BP) compilado por Tony Berber Sardinha da PUC-SP. Segundo Scarton, Almeida e Aluísio (2009), outros bancos de dados foram cogitados para alimentar a ferramenta – como o *corpus* NILC e o *corpus* de referência do LácioWeb, mas optou-se pelo BP por ser, na opinião dos autores, o maior e mais balanceado *corpus* existente para o português brasileiro. O BP possui 700 milhões de palavras (tokens).

Subclasse – *Hiperônimos de Verbos*

A terceira subclasse pertencente à classe 3 é *Hiperônimos de Verbos*, que é obtido somando-se o número de hiperônimos e dividindo o total pelo número de verbos. A hiperonímia é uma relação definida na Wordnet.Br (Dias-da-Silva et. al., 2002; Dias-da-Silva, 2003; Dias-da-Silva, 2005; Dias-da-Silva et. al., 2008 e

Scarton e Aluísio, 2009) como “supertipo de”. O verbo “sonhar”, por exemplo, tem como hiperônimos “fantasiar” e “mentalizar”.

Classe 4: Índices Sintáticos

Subclasse – Constituintes

A primeira subclasse da classe 4, *Constituintes*, abrange: **Incidência de Sintagmas**, que é a incidência de sintagmas nominais por 1000 palavras; **Modificadores de Sintagmas**, que é a média do número de modificadores por sintagmas nominais, sendo considerados modificadores os adjetivos, advérbios e artigos que participam de um sintagma; **Palavras antes de verbos principais**, que é a média de palavras antes de verbos principais na cláusula principal da sentença.

Subclasse - Pronomes, Tipos e Tokens

A segunda subclasse da classe 4, *Pronomes, Tipos e Tokens*, inclui: **Incidência de Pronomes Pessoais**, considerados os pronomes: eu, tu, ele/ela, nós, eles/elas, você e vocês e sua incidência em um texto; **Pronomes por Sintagmas**, que é a média do número de pronomes que aparecem em um texto pelo número de sintagmas; **Type/Token**, que é o número de palavras diferentes entre si dividido pelo número de ocorrências dessas palavras. Cada palavra, consideradas apenas as de conteúdo (substantivos, verbos, advérbios e adjetivos), é um tipo (type) e cada ocorrência dessa palavra em um texto é um token.

Subclasse- Conectivos

A terceira subclasse da classe 4, *Conectivos*, inclui: **Incidência de Conectivos**, que é o número de todos os conectivos que aparecem em um texto.

Os conectivos são classificados por tipo em aditivos, temporais, causais e lógicos. Também são diferenciados, em termos de polaridade, em positivos e negativos. Os positivos ampliam eventos, ou seja, a situação apresentada primeiro é retomada na situação conjugada; e os negativos param a ampliação de eventos,

não há continuidade da situação inicial na relação conjugada. Existe uma lista que pode ser consultada na página da ferramenta.

Subclasse - Operadores Lógicos

A quarta subclasse da classe 4, *Operadores Lógicos*, inclui **Incidência de e**, **Incidência de ou**, **Incidência de se**, que é o número de *e*, *ou* e *se* presentes em um texto, com exceção do *se* quando é pronome; e **Incidência de negações** (consideradas as negações: não, nem, nenhum, nenhuma, nada, nunca e jamais).

Classe 5: Índices Referenciais e Semânticos

Subclasse - Anáfora

A **Referência Anafórica** está relacionada à proporção de referências anafóricas que se referem a um constituinte presente em até cinco sentenças anteriores.

Subclasse - Co-referência

Este índice está relacionado à proporção de referências anafóricas entre sentenças adjacentes.

5.2.1.3

Métricas selecionadas para análise da legibilidade dos CDs

Dentre as métricas descritas na seção anterior, utilizamos as seguintes para fins de avaliação e comparação do texto em termos de sua legibilidade: **Número de palavras por sentenças**, **Número de verbos por número de sentenças**, **Sentenças por parágrafos**, **Sílabas por palavras de conteúdo**, **Incidência de negações**, **Frequências**, **Palavras antes de verbos principais**, **Incidência de conectivos**, **Aditivos positivos**, **Aditivos negativos**, **Temporais positivos**, **Temporais negativos**, **Causais positivos**, **Causais negativos**, **Lógicos positivos**,

Lógicos negativos, Sobreposição de argumentos adjacentes, Referência anafórica adjacente.

Incorporamos o **Índice Flesch** em nossa análise, pois, embora seja uma fórmula que avalia superficialmente a inteligibilidade do texto, visto que se baseia em aspectos mais materiais do texto – como número de palavras em sentenças (entenda-se, aqui, períodos), e número de letras ou sílabas por palavras – trata-se, como apontam Scarton, Almeida e Aluísio (2009), de um índice que indica, grosso modo, a complexidade da tarefa de leitura. O índice é obtido com as métricas *Flesch Reading Ease* e *Flesch-Kincaid Grade Level*. A primeira “é a única métrica de inteligibilidade já adaptada para o português (Martins et al.,1996) e incorpora o conceito de séries escolares da segunda”. A fórmula é: $ILF=164.835-[1.015 \times (\text{Número de palavras por sentença})] - [84.6 \times (\text{Número de sílabas do texto}/\text{Número de palavras do texto})]$, levando em consideração tamanho de sentenças e de palavras e relacionando esse índice com níveis de escolaridade.

Textos cujo índice varia entre 100-75 são analisados como **muito fáceis**, adequados às primeiras séries do Ensino Fundamental; textos cujo índice varia de 75-50 são tidos como **fáceis**, adequados às séries finais do Ensino Fundamental. Os que tem Índice Flesch entre 50-25 são analisados como **difíceis** e seriam voltados para ensino médio. Por fim, os que ficam entre 25-0 são classificados como **muito difíceis** e direcionados para o ensino universitário.

Acrescente-se aqui que, em função de restrições de processamento associadas a limite de armazenamento de informação pela memória de trabalho, os fatores considerados na fórmula do Índice Flesch podem, em princípio, vir a comprometer a compreensão.

Mantivemos também o índice **Número de palavras por sentenças**. Como já esclarecido, a princípio, sentenças longas aumentam a complexidade de um texto, dificultando o processamento da leitura.

O índice **Números de Verbos por Número de Sentenças** foi por nós calculado com base em dois índices – **Número de Sentenças** e **Incidência de Verbos**. A fórmula usada para calcular a Incidência de verbos, definida na ferramenta Coh-Metrix-Port, $\frac{\text{verbo}}{\text{palavra}} \times 1000$ (número de verbos/(número de palavras/1000) nos permitiu chegar ao número de verbos de cada texto. Em seguida, dividimos o número de sentenças pelo número de verbos. A idéia era identificar se o que a

ferramenta define como sentença (que na verdade é período) corresponderia a períodos simples ou compostos. Se o índice for 2,2, por exemplo, pode significar que a maior parte dos textos daquele Caderno Didático possui períodos compostos de duas orações (cada oração correspondendo a um verbo). Trata-se de um cálculo aproximado, visto que, até onde pudemos identificar, a ferramenta não parece distinguir verbos de locuções verbais. De todo modo, considerando-se que o problema se observa na análise de todos os textos, podemos usar esse parâmetro para comparar os textos. Entendemos que, em princípio, períodos compostos sejam, em termos de processamento, mais complexos do que períodos simples. Logicamente, sabemos que outros fatores, como a natureza sintática do vínculo oracional (coordenação ou subordinação) é também fator determinante dessa complexidade.

O índice **Sentenças por Parágrafos** que foi selecionado pode indicar se os parágrafos são longos ou curtos, o que potencialmente pode ser relacionado ao grau de complexidade informacional do parágrafo.

O índice **Sílabas por Palavra de Conteúdo** foi também mantido, pois pode ser relevante para casos onde há dificuldade com a decodificação do sinal gráfico, como no caso dos analfabetos funcionais. Leitores iniciantes, no ato da leitura, adotam predominantemente um processamento ascendente (*bottom-up*), bastante preso ao próprio *input* visual e, para chegar ao sentido das palavras, precisa se apoiar na linearidade das cadeias de letras (cf. Kleiman, 2001, p.36.).

Acrescente-se, ainda que as palavras de conteúdo desempenham um papel fundamental na leitura para a compreensão. Resultados de trabalho clássico na área de movimento ocular indicam que os leitores fixam 80% do movimento dos olhos nesse tipo de palavra, o mesmo não sendo observado para palavras funcionais (Carpenter & Just, 1983, *apud* Perfetti,1999).

O índice **Palavras Antes de Verbos Principais** foi selecionado porque, segundo a documentação do Coh-Metrix, é um bom índice para avaliar a carga da memória de trabalho relacionada à manutenção de informação relativa ao sujeito quando do processamento do verbo. Resultados experimentais apontam para a relevância de informação gramatical e semântica (relativa a papéis temáticos) codificada no verbo (cf. Frazier & Rayner, 1982; Britt, 1994; MacDonald, 1994; *apud* Perfetti,1999) . O elemento inicial da sentença para que possa ser

corretamente analisado e interpretado depende, portanto, da informação codificada no verbo. É necessário manter esse elemento inicial para que seu papel temático possa ser identificado adequadamente. Nesse sentido, quanto maior for a sequência de palavras antes de verbos principais, maior o custo em termos do processamento da leitura.

O índice **Incidência de Negações** (consideradas as negações: não, nem, nenhum, nenhuma, nada, nunca e jamais) também foi utilizado, pois a incidência maior de negações parece dificultar a compreensão. Resultados de experimentos envolvendo identificação de figuras a partir de sentenças afirmativas e negativas indicaram que estas são mais difíceis em termos de processamento do que aquelas. (Carpenter & Just, 1975; Clark & Chase, 1972; Trabasso et al., 1971 *apud* Tettamanti et al., 2008). Conforme os autores esclarecem nesse artigo, “a compreensão de sentenças negativas demanda maior tempo de processamento e estas apresentam maior percentagem de erros, conforme resultados de testes que envolvem pareamento de frases e figuras”. Além disso, estudos sobre acessibilidade de informação mencionada no escopo da negação, envolvendo tarefas de reconhecimento de palavras ou tarefas de *priming*” mostraram que os tempos gastos para responder eram significativamente mais longos para os itens com negações (MacDonald and Just, 1989; Kaup, 2001; Kaup and Zwaan, 2003 *apud* Tettamanti et al., 2008).

Em uma pesquisa realizada pela Profa. Sandra Maria Aluísio e Carolina Evaristo Scarton, elaboradoras do Projeto PorSimples, sobre análise de inteligibilidade em textos reescritos para crianças, o número de negações é menor em textos reescritos (Aluísio & Scarton, 2009, p. 53), o que também pode ser tomado como indicativo de que as negações constituem um custo maior para a leitura.

O índice **Frequências** indica a média de todas as frequências das palavras de conteúdo. O valor da frequência das palavras é retirado da lista de frequências do *corpus* Banco do Português (Sardinha, 2003). A presença de palavras pouco frequentes em um texto pode, evidentemente, também contribuir para sua complexidade. Pesquisas na área de reconhecimento de palavras apontam para o reconhecimento e acesso mais rápido e mais eficiente de palavras frequentes na língua (Moreno & Van Orden, 2000). Como afirma Perfetti (2005), para a eficácia

da compreensão, faz-se necessário que a identificação de palavras seja implementada com sucesso de modo a que o significado das mesmas seja recuperado e integrado à representação mental do texto em construção.

Incidência de Conectivos foi outro índice considerado. Ele inclui Aditivos positivos e negativos, Temporais aditivos e negativos, Causais aditivos e positivos e Lógicos positivos e negativos. Alguns exemplos podem ajudar a esclarecer esta relação:

1.a. *Estamos caminhando para nosso objetivo assim como vocês também estão.* (conectivo aditivo positivo)

1.b. *Estamos caminhando para nosso objetivo ainda que vocês não estejam.* (conectivo aditivo negativo)

2.a. *A população japonesa dá exemplo ao mundo porque permanece disciplinada.* (conectivo causal positivo)

2.b. *A população japonesa está sofrendo grandes privações apesar da ajuda mundial.* (conectivo causal negativo)

Em princípio, pode-se considerar que um texto com presença de conectivos facilita o processamento, porque as relações semânticas entre as sentenças são indicadas por esses elementos, o que não ocorre quando o vínculo de ligação entre orações não é expresso. Neste caso, o leitor precisa, muitas vezes, realizar processos inferenciais para determinar de que tipo são as relações.

Perfetti (2005) afirma que leitores pouco proficientes concentram-se mais na leitura de palavras isoladas do que em buscar coerência no texto. Citando um experimento de Cain e Oakhill (1999), ele conclui que os leitores habilidosos são melhores nas atividades de inferência de conexões do texto; daí a importância da presença de conectivos para leitores pouco proficientes, que precisam desses elementos para estabelecer as relações semânticas interfrásicas.

Degand e Sanders (2002) afirmam, com base no trabalho de Gaddy, van den Broek & Sung (2001) que há realmente um suporte empírico à opinião de que os conectivos e outros marcadores de coerência desempenham uma função facilitadora durante o processo de leitura. Reportam, ainda, que os conectivos propiciariam um processamento mais rápido de segmentos de texto imediatamente subsequentes (cf. Bestgen & Vonk 1995; Britton, Glynn, Mayer & Penland 1982;

Deaton & Gernsbacher 2000; Haberlandt 1982; Sanders 1992; Sanders & Noordman 2000).

Para fins da pesquisa, consideramos que a comparação dos conectivos de tipo negativo poderia ser relevante, visto que a operação envolvida no processamento de conectivos negativos – aqueles que param a ampliação de eventos – pode implicar quebra de expectativa e maior dificuldade de compreensão.

Outros índices considerados, além dos já mencionados, foram os seguintes: **Sobreposição de Argumentos Adjacentes** e **Referência Anafórica Adjacente**.

No caso **da Sobreposição de Argumentos Adjacentes**, diferentemente dos demais parâmetros, números mais elevados podem ser tomados como facilitadores da leitura, pois, em princípio, quando há sobreposição de argumentos, a informação é retomada facilitando a manutenção da referência e o processo de progressão temática. O exemplo dado no manual do Coh-Matrix-Port é o seguinte:

“(1)Dentro do lago, existem peixes, como a traíra e o dourado, além da palometa, um tipo de piranha. (2) Ela é uma espécie carnívora que se alimenta de peixes. (3) No verão, elas ficam mais próximas das margens da barragem, atraídas pela movimentação das pessoas e por restos de comida que alguns turistas deixam na água quando lavam os pratos.”

De acordo com o Manual, “as sentenças (1) e (2) compartilham um substantivo (peixes) e, portanto, este par incrementa 1 no valor de correferência. Como também há dois pares de sentenças adjacentes ((1) com (2) e (2) com (3)), o valor final da métrica é $\frac{1}{2}=0,5$.”

A **Referência Anafórica Adjacente** diz respeito à proporção de referências anafóricas entre sentenças adjacentes. O exemplo dado no manual do Coh-Matrix é o seguinte:

*“Dentro do lago, existem peixes, como **a traíra** e o dourado, além **da palometa**, um tipo de **piranha**. **Ela** é uma espécie carnívora que se alimenta de peixes.”*

Segundo os autores, os “candidatos” à retomada anafórica pelo pronome “ela” são *traíra*, *palometa* e *piranha*. Assim, como há três “candidatos” e uma

sentença adjacente, o valor final da métrica seria $3/1 = 3$. Logo, quanto mais alto for o valor final dessa métrica mais difícil será, em princípio, o processamento da anáfora, pois mais possibilidades existirão.

Conforme esclarece Perfetti (2001), citando Frederiksen (1981), a integração de informação referencial ao longo do texto tem maior custo de processamento para leitores menos habilidosos. Foi verificado que, quando há dois possíveis referentes para um pronome, esses leitores levam mais tempo para atribuir o pronome ao referente pretendido, ao menos quando o referente não corresponde ao primeiro nome da frase precedente.

Neste capítulo relatamos os procedimentos de análise e os critérios adotados para a aferição da legibilidade dos textos contidos nos Cadernos Didáticos de EJA. No próximo capítulo, faremos uma apresentação dos resultados obtidos.