

Gêneros científicos e o processamento da leitura por alunos universitários

Aluna: Juliana da Silva Neto
Orientadora: Erica dos Santos Rodrigues

Introdução e objetivos

O trabalho a ser relatado aqui se vincula ao projeto de pesquisa “A leitura e a produção escrita numa abordagem psicolinguística: demandas cognitivas e especificidades de processamento” [1], cujos objetivos específicos são (i) investigar as características linguísticas e textuais de gêneros científicos normalmente adotados em cursos universitários; (ii) delimitar habilidades linguísticas e discursivas necessárias à leitura e à produção dos gêneros científicos examinados, e (iii) contribuir para a construção de uma escala, em termos de níveis de complexidade, dos diferentes tipos de habilidades em leitura e escrita requeridos para cada um dos gêneros científicos analisados.

A pesquisa de iniciação científica “Gêneros científicos e o processamento da leitura por alunos universitários”, que está sendo desenvolvida no âmbito desse projeto, foca-se nos seguintes objetivos específicos: (i) caracterização de habilidades necessárias à leitura de gêneros textuais científicos usados em cursos universitários e (ii) identificação dos fatores que permitem distinguir leitores com diferentes graus de proficiência em leitura.

Na seção seguinte, apresentaremos uma breve resenha da bibliografia utilizada, que explora questões associadas às habilidades requeridas na compreensão do discurso escrito. Na segunda seção, iremos reportar a metodologia adotada e os resultados da verificação de características formais presentes em dois capítulos de livros didáticos acadêmicos. Concluímos com uma breve síntese do relatório, apontando os próximos estágios da pesquisa.

1. Breve resenha da literatura

1.1. As habilidades linguísticas requeridas para a compreensão textual

A partir das leituras de textos que abordam a questão da compreensão textual, chegamos a algumas conclusões com relação a essa questão. A primeira delas foi a de que a compreensão do discurso escrito envolve um conjunto de habilidades linguísticas e cognitivas, que vão desde a decodificação do material escrito e análise estrutural das sentenças a processos que implicam a integração de informações novas e a construção do sentido global do texto [2, 3]

A segunda conclusão foi a de que os leitores proficientes e não proficientes diferenciam-se em relação a um conjunto de habilidades, em especial o processamento de inferências: leitores pouco proficientes apresentam dificuldades para estabelecer inferências relacionadas ao assunto do texto e para integrar novas informações às já disponíveis. Leitores proficientes, ao contrário, são mais capazes de fazer inferências apropriadas para preencher lacunas conceituais durante a leitura e de integrar o sentido de estruturas textuais [4].

Quanto aos fatores que permitem explicar diferenças individuais relativas à proficiência em leitura, a bibliografia consultada apresenta um conjunto de hipóteses [4, 5, 6]. Há pesquisadores que vinculam tais diferenças à capacidade de memória de trabalho: leitores com maior capacidade de memória de trabalho teriam mais facilidade de manter e processar as informações do texto, o que é crucial no estabelecimento de determinados tipos de inferências. Outros afirmam que a proficiência em leitura vincula-se à capacidade de suprimir informações irrelevantes em um dado contexto. Outra linha de investigação considera que conhecimento de mundo é fator crucial para diferenciar os leitores: os mais proficientes possuiriam maior conhecimento sobre o tópico do texto e seriam capazes de usá-lo durante a leitura. Há, ainda, uma quarta proposta, que se apóia no emprego de estratégias de leitura: leitores proficientes teriam mais conhecimento sobre estratégias de leitura, o que lhes permitiria usar eficientemente seus conhecimentos.

Outras leituras possibilitaram o contato com estudos nos quais se investiga a interação entre fatores individuais, como os destacados acima, e propriedades textuais, em especial a questão da coesão textual. Os trabalhos consultados indicam que (i) textos com alto grau de coesão facilitam a compreensão textual; (ii) a compreensão global está nitidamente relacionada ao conhecimento prévio dos leitores e (iii) o grau em que os participantes se beneficiam de textos coesos depende das habilidades de leitura dos participantes. Leitores com conhecimento sobre o tópico do texto mas que possuem dificuldades de processar pistas coesivas tendem a confiar mais em seu conhecimento de mundo e processar superficialmente o texto, o que tem sido caracterizado como “expertise reversal effect” [6].

1.2. Livro didático acadêmico como suporte de gêneros

A partir das leituras sobre gêneros e suportes textuais realizadas até o momento, observaram-se divergências quanto à classificação do livro didático como gênero ou suporte [7,8]. Marcuschi (2003) considera o livro didático como um suporte textual de diferentes gêneros textuais, os quais têm sua funcionalidade alterada ao serem incorporados nesse suporte. Bezerra (2006) adota mesma classificação para os livros acadêmicos. Nesta pesquisa, utilizamos a definição de livro adotada por esses dois autores. Consideramos, assim, que o livro didático acadêmico é um suporte que comporta uma ou mais variedades de gêneros textuais e que pode apresentar diferentes características formais de acordo com as disciplinas a que esse tipo de livro se destina. Passamos agora a relatar a metodologia empregada para a futura análise do grau de legibilidade de textos de livros didáticos acadêmicos de diferentes disciplinas.

2. Metodologia e resultados preliminares

A metodologia seguida neste projeto é orientada por uma visão processual da leitura e envolve o emprego de ferramentas computacionais que permitem avaliar o grau de legibilidade de textos, levando em consideração vários níveis de análise linguística: léxico, sintático, discursivo e conceitual.

Até o momento, foram selecionados dois textos de disciplinas introdutórias dos cursos de Letras e de Administração da Pontifícia Universidade Católica do Rio de Janeiro. O primeiro texto é “Linguagem, língua e linguística”, da autora Margarida Petter [9]. O segundo, “A administração e suas perspectivas: delineando o papel da administração”, de Idalberto Chievenato [10]. O critério de seleção para os textos foi o de serem os das primeiras disciplinas dos referidos cursos e por introduzirem conceitos básicos das respectivas áreas.

A ferramenta computacional utilizada para verificar o grau de legibilidade dos textos e, a partir daí, deduzir as habilidades necessárias para a compreensão dos mesmos foi Coh-Metrix-Port [11] (ver <http://caravelas.icmc.usp.br/manual.pdf>). Essa ferramenta, em português, foi desenvolvida a partir das métricas da ferramenta Coh-Metrix, criada na Universidade de Memphis. A versão 1.0 do Coh-Metrix-Port utiliza 34 das 60 métricas disponíveis na versão livre da ferramenta Coh-Metrix. Com base nessas métricas, realiza cálculos que avaliam características linguísticas dos textos, tais como grau de coesão, coerência e número de resoluções anafóricas. Para tais cálculos, a ferramenta utiliza vários níveis de análise linguística: léxico, sintático, discursivo e conceitual. Antes de reportar os resultados do processamento dos textos por essa ferramenta, apresentaremos algumas observações sobre as especificidades de cada um deles.

2.1. Análise dos textos selecionados

Nesta seção serão relatadas algumas particularidades observadas em cada um dos dois textos selecionados para análise. Embora os dois capítulos façam parte de livros didáticos acadêmicos que visam a introduzir conceitos básicos requeridos pelas disciplinas dos cursos, ambos apresentam diferenças significativas quanto à organização gráfica que podem comprometer seu processamento.

O texto de Linguística – “Linguagem, língua e lingüística” – é organizado graficamente em seções nas quais a autora apresenta suas ideias linearmente, a partir de um texto expositivo. Terminada a discussão de um tópico, ela introduz outro, iniciando outra seção, com a mesma estrutura e composição gráfica da(s) anterior(es). Ao final, apresentam-se mais duas partes: a bibliografia e as sugestões de leitura.

Comparada organização gráfica dos dois textos, foi observado que a do segundo – “A administração e suas perspectivas: delineando o papel da administração” – é menos simples, pois o autor apresenta suas ideias de forma menos linear. Ele também organiza a apresentação dos assuntos por seções, mas intercala sequências textuais diferentes da principal. No capítulo selecionado, três tipos de sequências foram encontradas:

- 1- situações-problema: exercícios que requerem o relacionamento do que foi lido anteriormente no capítulo para sua resolução. Nessa sequência, o autor utiliza situações hipotéticas como exemplos;
- 2- aplicação dos conceitos discutidos anteriormente em caso de empresas reais. Com essa sequência, parece-nos que o objetivo do autor é aprofundar o entendimento dos conceitos expostos e
- 3- explicações mais detalhadas sobre conceitos ou questões expostos.

Ao final, são apresentados, ainda, a bibliografia, o resumo do capítulo, as perguntas sobre os principais conceitos apresentados, e o glossário.

Por essas observações, chegamos à conclusão de que, apesar de esse último texto ser simples como o primeiro no que se refere à exposição de conceitos básicos, sua organização textual e composição gráfica podem dificultar a compreensão para o leitor não-familiarizado com essa forma de introdução de sequências secundárias, que interrompem a linearidade do texto. Descritas essas observações, apresentamos, a seguir, os resultados do processamento desses dois textos pela ferramenta referida acima.

2.2. Resultados do processamento pelo Coh-Metrix-Port

Nesta seção, apresentaremos os resultados de um primeiro processamento dos capítulos de Linguística e de Administração pela ferramenta Coh-Metrix Port.

Registramos todas as métricas utilizadas pelo programa, com uma breve explicação das mesmas na última coluna da tabela.

Para o texto “Linguagem, língua e linguística”, a ferramenta gerou os resultados apresentados no campo 1 da tabela a seguir. Para o texto “A administração e suas perspectivas”, dadas as diferenças comentadas na seção 2.1 quanto à composição dos capítulos das duas áreas, foi necessário realizar dois processamentos: o primeiro apenas com o texto principal, cujos resultados estão no campo 2, e o segundo, com os textos principal e os secundários, além do resumo e glossário, disponíveis ao final do capítulo – campo 3.

Os resultados a seguir serão alvo de análise nas próximas etapas da pesquisa. No momento, gostaríamos apenas de destacar os números referentes ao parâmetro “Índice de Legibilidade Flesch” (ILF). Esse índice busca avaliar a legibilidade textual com base em uma correlação entre tamanhos médios de palavras e sentenças e a facilidade de leitura. O cálculo é feito com base na seguinte fórmula:

$$ILF = 164.835 - [1.015 \times (\text{Número de palavras por sentença})] - [84.6 \times (\text{Número de sílabas do texto} / \text{Número de palavras do texto})]$$

Aplicado à língua portuguesa [12], esse índice permite identificar quatro faixas de dificuldades de leitura:

- textos classificados como muito fáceis (índice entre 75 - 100), que seriam adequados para leitores com nível de escolaridade até a quarta série do ensino fundamental;
- textos fáceis (índice entre 50 - 75), que seriam adequados a alunos com escolaridade até a oitava série do ensino fundamental;
- textos difíceis (índice entre 25 - 50), que seriam adequados para alunos cursando o ensino médio ou universitário e;
- textos muito difíceis (índice entre 0 - 25), que em geral seriam adequados apenas para áreas acadêmicas específicas.

Com base nesse parâmetro, pode-se afirmar que os capítulos de Linguística e de Administração (versão completa) apresentam índices bastante próximos, sendo possível situá-los na fronteira entre textos classificados como muito difíceis e difíceis. Logo, embora voltados para áreas acadêmicas específicas, seriam adequados a alunos universitários iniciantes e alunos de ensino médio.

A comparação entre a versão parcial (apenas texto principal – campo 2) e a versão completa (campo 3) do texto de Administração revela uma diferença importante quanto ao índice Flesch. Note-se que, quando não processamos os textos secundários – os quais podem ser vistos como material auxiliar, complementar à leitura, – a legibilidade textual sofre uma redução, fazendo com que o capítulo de Administração torne-se menos legível.

É importante observar que o índice Flesch mede apenas a legibilidade com base em fatores como tamanho de palavras e de sentenças e, nesse sentido, não pode ser tomado isoladamente como um parâmetro de avaliação da compreensibilidade textual. No entanto, é certo que tais variáveis certamente precisam ser consideradas na análise dos fatores que contribuem para a maior ou menor dificuldade/ facilidade de compreensão de um texto, visto que, como apontado na resenha da literatura, um dos elementos que distinguem bons e maus leitores é a capacidade de retenção e de recuperação de informação durante a leitura.

Nas próximas etapas da pesquisa, iremos explorar em detalhes os resultados relativos às outras métricas, procurando integrar essa análise dos fatores textuais a questões de gênero/suporte.

Contagens Básicas				
	Campo 1	Campo 2	Campo 3	
Índice Flesch	23.06743175 01571	16.7102	21.276870740975 7	Índice Flesch
Número de Palavras	5894.0	3556.0	7009.0	Número de palavras do texto.
Número de Sentenças	223.0	175.0	383.0	Número de sentenças de um texto.
Número de Parágrafos	87.0	31.0	115.0	Número de parágrafos de um texto. Parágrafos são apenas onde há quebra de linha (não identações).
Palavras por Sentenças	26.43049327 35426	20.32	18.300261096605 7	Número de palavras dividido pelo número de sentenças.
Sentenças por Paragrafos	2.563218390 8046	5.645161290 32258	3.3304347826087	Número de sentenças dividido pelo número de parágrafos.
Sílabas por Palavras de Conteúdo	3.159428737 97727	3.458595641 64649	3.3876749324822	Número médio de sílabas por palavras de conteúdo (substantivos, verbos, adjetivos e advérbios).
Incidência de Verbos	145.0627757 04106	124.0157480 31496	134.25595662719 4	Incidência de verbos em um texto.
Incidência de Substantivos	288.5985748 21853	293.5883014 62317	300.47082322727 9	Incidência de substantivos em um texto.
Incidência de Adjetivos	113.3355955 20869	129.3588301 46232	111.85618490512 2	Incidência de adjetivos em um texto.
Incidência de Advérbios	35.12046148 62572	33.74578177 72778	34.527036667142 2	Incidência de advérbios em um texto.
Incidência de Pronomes	52.42619613 16593	56.80539932 50844	53.074618347838 5	Incidência de pronomes em um texto.
Incidência de Palavras de Conteúdo	582.1174075 33084	580.7086614 17323	581.11000142673 7	Incidência de Palavras de Conteúdo (substantivos, adjetivos, advérbios e verbos).
Incidência de Palavras Funcionais	379.5385137 42789	390.3262092 23847	384.64830931659 3	Incidência de Palavras Funcionais (artigos, preposições, pronomes, conjunções e interjeições).
Operadores Lógicos				
Incidência de Operadores Lógicos	32.57550050 89922	63.83577052 86839	61.207019546297 6	Incidência de operadores lógicos em um texto. Consideramos como operadores lógicos: e, ou, se, negações e um número de condições.
Incidência de E	23.24397692 56871	48.93138357 70529	45.084890854615 5	Incidência do operador lógico <i>e</i> em um texto.
Incidência de OU	6.786562606 04004	12.09223847 01912	10.557854187473 2	Incidência do operador lógico <i>ou</i> em um texto.
Incidência de SE	0.339328130 302002	1.124859392 57593	1.141389641889	Incidência do operador lógico <i>se</i> em um texto.
Incidência de Negações	1.526976586 35901	1.124859392 57593	1.4267370523612 5	Incidência de Negações. Consideramos como negações: <i>não, nem, nenhum, nenhuma, nada, nunca e jamais.</i>
Frequências				
Frequências	197632.5491 11046	186000.1384 98789	197161.65062607 4	Média de todas as frequências das palavras de conteúdo encontradas no texto. O valor da frequência das palavras é retirado da lista de frequências do corpus Banco do Português.
Mínimo Frequências	2417.825870 64677	96615.69047 61905	79295.864553314 1	Identifica-se a menor frequência dentre todas as palavras de conteúdo em cada sentença. Depois, calcula-se uma média de todas as frequências mínimas. A palavra com a menor frequência é a mais rara da sentença.

Hiperônimos				
Hiperônimos de verbos	0.466494845 360825	0.475247524 752475	0.4689331770222 74	Hiperônimos de verbos.
Pronomes, Tipos e Token				
Incidência de Pronomes Pessoais	2.205632846 96301	268.8413948 25647	2.9961478099586 2	Incidência de pronomes pessoais em um texto. Consideramos como pronomes pessoais: <i>eu, tu, ele/ela, nós, nós, eles/elas, você e vocês.</i>
Pronomes por Sintagmas	0.033976795 9375627	0.682008368 200837	0.0273439558721 476	Média do número de pronomes que aparecem em um texto pelo número de sintagmas.
Type/Token	0.458758379 481201	6.309523809 52381	0.4335870365823 72	Número de palavras únicas dividido pelo número de tokens dessas palavras. Cada palavra única é um tipo. Cada instância desta palavra é um token.
Constituintes				
Incidência de Sintagmas	261.7916525 27995	268.8413948 25647	276.92966186331 9	Incidência de sintagmas nominais por 1000 palavras.
Modificadores por Sintagmas	0.679196370 706416	0.682008368 200837	0.6130860381246 78	Média do número de modificadores por sintagmas nominais, <i>adjetivos, advérbios e artigos</i> , que participam de um sintagma.
Palavras antes de verbos principais	5.084577114 42786	6.309523809 52381	4.9452449567723 3	Média de palavras antes de verbos principais na cláusula principal da sentença.
Conectivos				
Incidência de Conectivos	67.86562606 04004	73.11586051 74353	76.473106006563	Incidência de todos os conectivos que aparecem em um texto.
Conectivos Aditivos Positivos	29.69121140 14252	33.46456692 91339	35.383078898559	Incidência de conectivos classificados como aditivos positivos.
Conectivos Aditivos Negativos	3.562945368 17102	2.812148481 43982	2.8534741047225	Incidência de conectivos classificados como aditivos negativos.
Conectivos Temporais Positivos	7.295554801 49304	5.061867266 59168	6.5629904408617 5	Incidência de conectivos classificados como temporais positivos.
Conectivos Temporais Negativos	0.0	0.0	0.0	Incidência de conectivos classificados como temporais negativos.
Conectivos Causais Positivos	35.45978961 65592	34.58942632 17098	34.669710372378 4	Incidência de conectivos classificados como causais positivos.
Conectivos Causais Negativos	1.526976586 35901	0.562429696 287964	0.5706948209445	Incidência de conectivos classificados como causais negativos.
Conectivos Lógicos Positivos	33.59348489 98982	27.27784026 99663	29.533456983877 9	Incidência de conectivos classificados como lógicos positivos.
Conectivos Lógicos Negativos	3.562945368 17102	1.124859392 57593	1.4267370523612 5	Incidência de conectivos classificados como lógicos negativos.
Ambiguidades				
Verbos	5.467836257 30994	5.453514739 22902	5.4729011689691 8	Ambiguidade de Verbos.
Substantivos	2.143230869 84958	2.242303872 88977	2.3440294581799 1	Ambiguidade de Substantivos.
Adjetivos	1.974550898 20359	2.091304347 82609	2.0688775510204 1	Ambiguidade de Adjetivos.
Advérbios	0.0	0.0	0.0	Ambiguidade de Advérbios.
Correferência				
Sobreposição de argumentos adjacentes	0.845	0.383233532 934132	0.4104046242774 57	Sobreposição de argumentos em sentenças adjacentes.
Sobreposição de argumentos	0.576169154 228856	0.215212432 278301	0.1798903899651 85	Sobreposição de argumentos em todos os pares de sentenças.

Sobreposição de radicais de palavras adjacentes	0.965	0.568862275 449102	0.5809248554913 29	Sobreposição de argumentos em sentenças adjacentes.
Sobreposição de radicais de palavras	0.820796019 900498	0.346378671 228971	0.3128050507237 93	Sobreposição de radicais de palavras em todos os pares de sentenças.
Sobreposição de palavras de conteúdo	0.805	0.407185628 742515	0.4479768786127 17	Sobreposição de palavras de conteúdo em sentenças adjacentes.
Anáforas				
Referência anafórica adjacente	0.134328358 208955	0.267857142 857143	0.1412103746397 69	Referência anafórica em sentenças adjacentes.
Referência anafórica	0.159203980 099502	0.303571428 571429	0.1671469740634 01	Referência anafórica em até cinco sentenças anteriores.

3 - Considerações finais

Foi apresentado no texto o desenvolvimento das primeiras etapas dessa pesquisa, que consistiram em levantamento e análise da bibliografia – brevemente apresentada acima –, seleção de textos para análise e processamento dos mesmos na ferramenta Coh-Metrix-Port.

As próximas etapas do trabalho consistem em (i) aumentar a composição de *corpus* de textos representativos de gêneros científicos mais recorrentes na bibliografia de cursos de graduação de diferentes áreas; (ii) analisar os textos selecionados por ferramentas computacionais, como Coh-Metrix-Port; (iii) analisar criticamente os critérios empregados na definição de níveis de proficiência em testes de avaliação de habilidades em leitura (ex. provas do *Programme for International Student Assessment - PISA*); (iv) construir uma matriz que permita relacionar propriedades linguístico-textuais dos gêneros examinados e habilidades em leitura necessárias ao processamento desses textos.

Referências

1-RODRIGUES, E. dos S. **A leitura e a produção escrita numa abordagem psicolinguística: demandas cognitivas e especificidades de processamento.** Projeto de Pesquisa, Depto. de Letras, PUC-Rio.

2- MATLIN, Margaret W. Linguagem I: Introdução à linguagem e à compreensão da linguagem. In: _____. **Psicologia cognitiva.** Rio de Janeiro: LTC, 2004, cap. 8, p.186 – 213.

3- GERNSBACHER, M. Language comprehension. In: L. Nadel (Ed.). **Encyclopedia of Cognitive Sciences.** London, UK: Nature Publishing Group, 2002.

4- MCNAMARA, Danielle S.; O'REILLY, T. Theories of comprehension skill: knowledge and strategies versus capacity and suppression. In: A. M. F. Columbus (Ed.). **Advances in Psychology Research,** v.62, Hauppauge, NY: Nova Science Publishers, 2010, p. 113-136.

5- MCNAMARA, Danielle S. The importance of teaching reading strategies. **Perspectives on language and literacy,** p. 34-40, Spring 2009.

6- OZURU, Yasuhiro; DEMPSEY, Kyle; MCNAMARA, Danielle S. Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. **Learning and instruction,** v.19, n. 3, p.228-242, jun. 2009.

7- MARCUSCHI, Luiz Antônio. A questão do suporte dos gêneros textuais. A questão do suporte dos gêneros textuais. **DLCV: Língua, linguística e literatura.** João Pessoa, v. I, n. 1, p. 9-40, out. 2003.

8- BEZERRA, Benedito Gomes. **Gêneros introdutórios em livros acadêmicos.** Tese de Doutorado – Centro de Artes e Comunicação da Universidade Federal de Pernambuco, Recife, 2006.

9- PETTER, Margarida. Linguagem, língua e lingüística. In: FIORIN, José Luiz (org.). **Introdução à linguística I: objetos teóricos.** 5ª ed. São Paulo: Contexto, 2006. p. 11-25.

10- CHIAVENATO, Idalberto. A administração e suas perspectivas: delineando o papel da administração. In: _____. **Introdução à teoria geral da administração.** 7ª ed. Rio de Janeiro: Elsevier, 2004. pp. 9-25.

11- ALMEIDA Daniel Machado de; ALUÍSIO Sandra Maria. **Manual de uso do Coh-Metrix-Port 1.0.** Technical Report NILC-TR-09-05, 13 p. Agosto 2009, São Carlos-SP.

12- MARTINS, T.B.F.; GHIRALDELO, C.M.; NUNES, M.G.V.; OLIVEIRA JR O.N.. **Readability Formulas Applied to Textbooks in Brazilian Portuguese.** Notas do ICMSC-USP ICMC. 1996.

Disponível em: < <http://www.nilc.icmc.usp.br/nilc/download/Reltec28.pdf>>.