

---

UNIVERSIDADE DE SÃO PAULO

Pró-Reitoria de Graduação  
Curso de Ciências Moleculares

---

**PROJETO DE INICIAÇÃO CIENTÍFICA**  
**Rastreando a leitura literária: um estudo dos**  
**movimentos oculares**

---

Aluno Leonardo Zeine Mendes de Souza  
e-mail: leonardo.zeine@gmail.com

Orientador Prof. Dr. Renato Vicente  
Departamento de Matemática Aplicada  
Instituto de Matemática e Estatística - USP  
Rua do Matão 1010, Cidade Universitária  
05580-090 São Paulo - SP  
e-mail: rvicente@ime.usp.br

Co-orientadores Prof. Dr. Hamilton Haddad Júnior  
Departamento de Fisiologia  
Instituto de Biociências - USP  
Rua do Matão, travessa 14, 101  
05580-090 São Paulo - SP  
e-mail: haddad@usp.br

Dr. Samira Murad  
Departamento de Letras Modernas  
Faculdade de Filosofia e Ciências Humanas - USP  
Rua do Lago, 717, Cidade Universitária  
05508-080 - São Paulo - SP  
e-mail: smurad@usp.br

Período Segundo Semestre de 2015

# 1 Introdução e Justificativa

Há muito se sabe da importância dos movimentos oculares nos processos de cognição que envolvem processamento de linguagem. Os primeiros resultados robustos que se teve nessa área, no entanto, apareceram nas décadas de 50, com Alfred L. Yarbus e 70, com Rayner. Neste ponto do estudo, já se sabia, por exemplo, que a leitura não é um movimento contínuo, e sim intercalado por fixações (repouso da visão sobre palavras específicas) e sacadas ("pulos" entre palavras, sem necessidade de processamento), e que o leitor costuma voltar àquelas palavras ou segmentos que lhe chamam atenção.

Marcel Just e Patricia Carpenter (1987) também produziram resultados relevantes usando a técnica de eye-tracking. O modelo de leitura desse estudo assume que assim que o leitor passa por uma palavra, ele tenta compreender seu sentido (Hipótese da Imediatez); e que o tempo de fixação do olho em cada palavra é proporcional ao tempo de compreensão (Hipótese olho-mente). Segundo este trabalho também, o tempo de fixação é proporcional à raridade da palavra na língua, ao tamanho da palavra e à complexidade semântica e/ou sintática daquela palavra no trecho.

Vários estudos caminharam nessa direção, no entanto pouco se sabe sobre leitura de trechos mais longos. Os trabalhos mais importantes da área analisam leitura de palavras e/ou sentenças e quase nunca, parágrafos ou textos. Portanto, sabe-se pouco sobre a estrutura esquemática chamada gramática do texto, ou gramática da história, que descreve as inferências intra-textuais, sequência de ações, personagens, ambientação e etc. Neste sentido, a ideia do presente trabalho é analisar trechos grandes (de 5 ou 6 páginas) e por isso, atacar um problema pouco explorado até o momento.

Em acréscimo, este trabalho tem também o objetivo de estudar a estrutura textual de textos literários, em especial de Marcel Proust, que é tido por grande parte de seus leitores como um escritor de textos extremamente difíceis e desconexos. A escolha do texto pode ser interessante no sentido de que ele é um extremo, o que será útil na hora de comparação entre outros textos literários, e de que a dificuldade da leitura dos textos de Proust não é óbvia, como veremos a seguir, e o entendimento dessa arquitetura textual

complexa pode sugerir aspectos mais gerais sobre o processamento textual.

Para que o trabalho se aproxime do processamento de textos literários, escolhemos como objeto de estudo o caso da leitura das primeiras páginas do primeiro volume do romance *Em Busca do Tempo Perdido* de Marcel Proust, tido por grande parte de seus leitores como um texto de leitura bastante difícil. Entender como esta dificuldade se dá é um dos objetivos deste trabalho, o que, espera-se, pode apontar, para aspectos mais gerais do processamento de textos literários.

## 2 Objetivos

Análise de movimentos oculares durante a leitura de textos literários: trecho de *Em Busca do Tempo Perdido* de Marcel Proust, como texto principal, trecho de *Os Subterrâneos do Vaticano* de André Gide, como controle, *A Incapacidade de Ser Verdadeiro* e *A Lanterna*, ambos de Carlos Drummond de Andrade e ambos utilizados como textos de aquecimento e ambientação.

O texto-controle, que é contemporâneo a Proust, é, intuitivamente, de mais fácil compreensão, mesmo que tenha léxico e tamanho próximos ao texto foco. É interessante ressaltar que esta intuição vem, por um lado, da aceitação histórica de suas obras - Proust teve dificuldades por muito tempo com a crítica e para publicação - e, por um lado mais estrutural, do uso frequente, nos textos de Proust, de orações longas, subordinadas, parênteses, travessões, que alongam os períodos, e digressões; características estas não encontradas no texto de Gide e que têm como resultado a dificuldade por parte dos leitores de entender o porquê da narrativa e de fazer previsões. A análise do trecho de *Os Subterrâneos do Vaticano* trará características comuns a textos literários, porém sem as especificidades da estrutura complexa de Proust.

Os textos de aquecimento são curtos, um e dois parágrafos, e de fácil compreensão. A ambientação deve trazer ao leitor a expectativa de ler textos literários durante o experimento, porém sem a previsão de tema dos textos foco e controle, que não tratam, intencionalmente, de assuntos similares.

Em primeiro momento, deve-se caracterizar a leitura de textos literários,

tendo Proust como um extremo do gênero. Em um segundo momento, a construção de um modelo de previsão de performance na leitura será o foco: dadas informações sobre os hábitos de leitura do leitor, deve-se inferir o tempo que ele levará durante o experimento e o nível de compreensão do texto. É como uma medida de aptidão para a leitura de Proust.

É importante ressaltar que a diferença no tempo de leitura médio dos textos do Proust e do Gide, o que já foi observado em laboratório, não se dá imediatamente nem pela dificuldade ou raridade lexical nem pelo tamanho das palavras; cujo aumento causa maior tempo de leitura, como já se sabe: na Figura 1, pode-se ver que Proust utilizava menos palavras com 7 ou mais letras que Gide e de mesma forma, é observado na figura 2.A e 2.B que Proust utilizava, proporcionalmente, mais palavras não-raras do que Gide, enquanto os dois textos se assemelham nas proporções de palavras raras.

### **3 Métodos**

Para a captação de dados de movimento ocular, será utilizado o eye-tracker Tobii modelo TX300.

### **4 Análise de dados**

Como reflexo da prioridade dos estudos de eye-tracker por textos menores, os softwares usados nessas técnicas não possuem a ferramenta de cruzar as coordenadas das palavras no arquivo .pdf com as coordenadas de cada fixação. O que se tem pronto é a marcação manual de áreas de interesse e com isso, o tempo de fixação dentro dessa área. Como se trata de um texto longo, estamos escrevendo um código que extrai as coordenadas das palavras do arquivo .pdf, as redimensiona para a escala do software da Tobii e as cruza com as coordenadas e tempos de fixação. Após o término do programa, o aluno o disponibilizará para outros grupos na internet.

Dado esse primeiro passo, deve-se procurar fixações de interesse. Algumas perguntas importantes são: palavras de duplo sentido, que como já se sabe

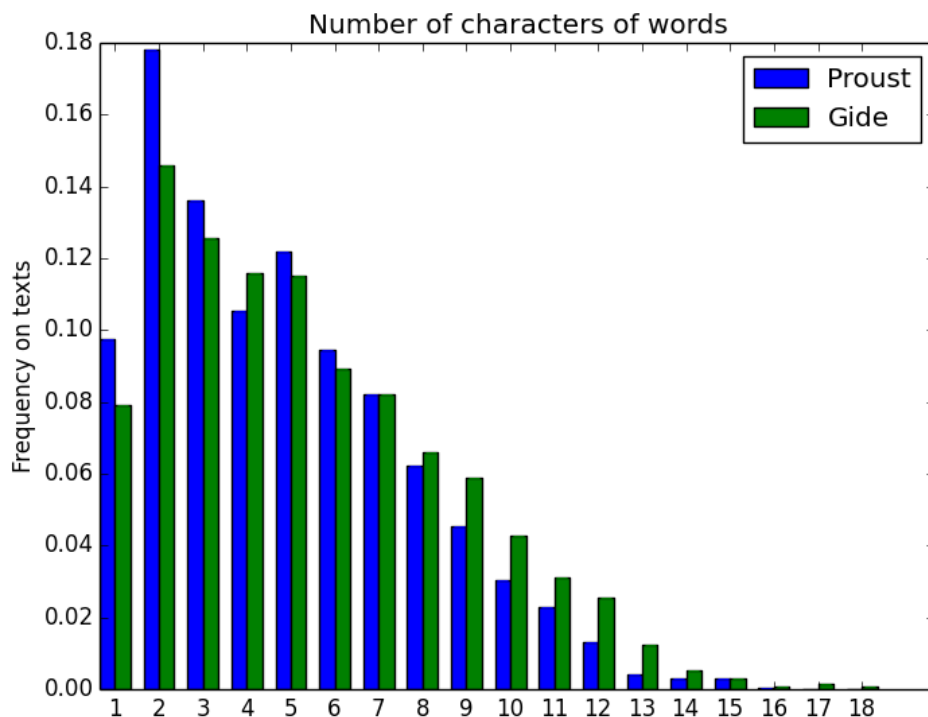


Figura 4.1: Relação entre frequência de aparição no texto e número de letras das palavras

possuem maior tempo médio de fixação, apresentam diferenças de tempo entre os textos do Gide e do Proust? quais são as especificidades da leitura de literatura? há pontos de ruptura de encadeamento lógico comuns aos leitores? se sim, esses pontos representam alguma função sintática/morfológica interessante? pode-se apontar pontos de perda de atenção e caracterizá-los?

Nesse sentido, algumas fixações já são caracterizadas como responsáveis por integração da informação lida, os chamados pontos de wrap-up. Seguindo a hipótese olho-mente, sabe-se que nesses pontos de wrap-up, quanto maior o tempo de fixação, maior a quantidade de informação a ser integrada. Corroborando para isso, Kintsch e Keenan (1973) propõem o conceito de complexidade proposicional, que afirma que o tempo de fixação total por frase é tão maior quanto for a quantidade de conceitos básicos colocados pela

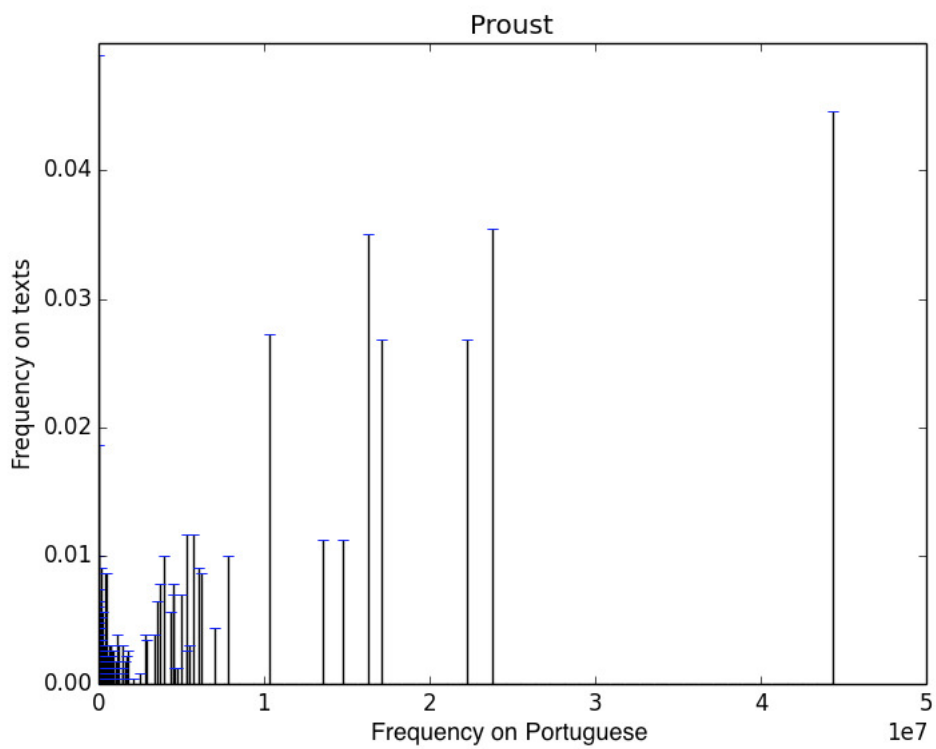
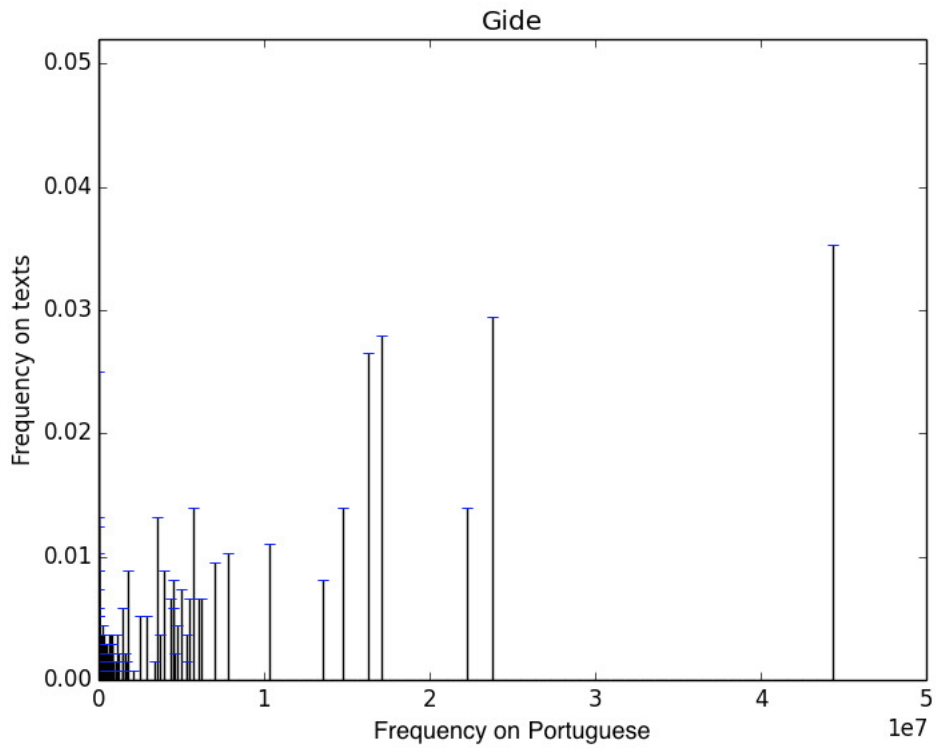


Figura 4.2: Figura 2: Frequência de uso no texto e frequência de uso na língua portuguesa, ou, medida de raridade das palavras, para o texto do Proust e do Gide

frase. Portanto, é de se esperar que os pontos de integração de informação nos textos do Proust sejam identificáveis pelo maior tempo de fixação; caso contrário, muito provavelmente, a compreensão do texto será pobre, ou num outro extremo, o leitor é experiente na escrita proustiana.

Encontrados pontos de interesse tanto sobre textos literários quanto sobre a escrita de Proust, deve-se agora construir um modelo matemático que, dado características prévias do leitor, seja capaz de prever a facilidade com que esse leitor lerá o texto. A ideia é ligar as habilidades do leitor com pontos específicos do texto e consolidar a importância que esses pontos têm durante a leitura do texto literário. Uma vez formado um padrão de aptidão para a leitura do texto, uma proposta do porquê da complexidade do texto vai ser formada, bem como um modelo de diferença entre a leitura de textos não-literários e literários. Para isso, durante a coleta de dados inicial e análise de resultados dos tempos de leitura dos textos, o aluno fará uma revisão bibliográfica da área de modelagem de leitura, em especial dos modelos bayesianos.

## 5 Disciplinas

- Segundo semestre de 2015
  - Iniciação à Pesquisa
  - Mecânica Estatística - IF
  - Física-Matemática I - IF
  - Percepção e Cognição - IP
  
- Primeiro semestre de 2016
  - Iniciação à Pesquisa
  - Neurociência Cognitiva: Percepção e Ação - IB Pós
  - Neurociência Cognitiva: Atenção e Memória - IB Pós
  - Probabilidade I - IME

- Etologia - IP
- Segundo semestre de 2016
  - Iniciação à Pesquisa
  - Eletromagnetismo I - IF
  - Modelos Computacionais em Ciência Cognitiva - Poli Pós
  - Etologia do Ser-humano - IP
- Primeiro semestre de 2017
  - Iniciação à Pesquisa
  - Mecânica Quântica - IF
  - Dinâmica Estocástica - IF

## Referências

- [1] E. R. Kandel, J. H. Schwartz e T. M. Jessell *Essentials of Neural Science and Behaviour*, Prentice Hall International 373-374 (1995)
- [2] Keith Rayner, *Eye Movements in Reading and Information Processing: 20 Years of Research*, Psychological Bulletin, Vol. 124, No. 3, 372-422 (1998).
- [3] David Barber, *Bayesian Reasoning and Machine Learning*, (2012)
- [4] Patricia Carpenter, Marcel A. Just, *A theory of reading: from eye fixations to comprehension* (1980).
- [5] Norris D. *The Bayesian reader: explaining word recognition as an optimal Bayesian decision process*, Psychological Review, (2006)
- [6] Erik D. Reichle, Alexander Pollatsek, Keith Rayner *E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading*, Cognitive Systems Research 7 (2006) 4–22.



- [7] Robert A. Jacobs, John K. Kruschke *Bayesian learning theory applied to human cognition*, WIREs Cognitive Science Volume 2, January/February 2011.
- [8] Kathleen M. Galotti *Cognitive Psychology: in and out of laboratory*, 5 edição (2014).