

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
CURSO DE PÓS-GRADUAÇÃO EM LITERATURA

Deise J. T. de Freitas

A composição do estilo do contista Machado de Assis

Tese

Florianópolis

2007

Deise J. T. de Freitas

A composição do estilo do contista Machado de Assis

Tese apresentada ao Curso de Pós-Graduação em Literatura da Universidade Federal de Santa Catarina como requisito para a obtenção do título de Doutora em Literatura.

Orientador: Prof. Dr. Alckmar Luiz dos Santos

Florianópolis

2007

Este trabalho é dedicado a Sara, Leomar e Jacques.

AGRADECIMENTOS

Aos meus pais.

Ao Jacques e a todos os meus amigos e familiares que não me deixaram desistir.

A Alckmar e meus colegas do Nupill, que encararam o desafio de entender o Hyperbase e se aventuraram comigo no universo da estatística textual para autodidatas.

Ao prof. Dr. Carlos Maciel.

Ao curso de Pós-Graduação em Literatura.

Ao CNPq e à Capes, que financiaram parcialmente este trabalho.

RESUMO

FREITAS, Deise J. T. de. **A composição do estilo do contista Machado de Assis**. 2009. 204 f. Tese (Doutorado em Literatura) – Programa de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis.

Esta tese discute a percepção, ainda vigente em parte da crítica literária, de que a obra de Machado de Assis é cindida em duas partes, como se fosse possível a um autor ter dois estilos distintos. Amparada na revisão da fortuna crítica machadiana e com o método da estatística textual mediante a utilização do programa Hyperbase, compara bases de dados formadas pelo conjunto de contos de Machado, cotejando-os com os romances do autor e a base Portext. A análise exploratória dos dados permite descrever a anatomia do material que compõe o conto machadiano, enquanto as funções estatísticas viabilizam a busca de padrões e transformações no léxico e na distribuição do texto. Os resultados da análise qualitativa, ao indicarem que há poucas variações de classe gramatical e de vocabulário no material, contrapõem-se à ideia de ruptura estilística e reforçam a hipótese de que a transformação do estilo de Machado de Assis no conto é gradual e encontra-se fundamentalmente não no material linguístico, mas na composição.

Palavras-chave: Machado de Assis. Conto. Estatística Textual. Estilo.

ABSTRACT

FREITAS, Deise J. T. de. **A composição do estilo do contista Machado de Assis**. 2009. 204 f. Tese (Doutorado em Literatura) – Programa de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis.

This thesis investigates the statement, still in use by some literary critics, that the work of Machado de Assis is divided in two parts, as if could be possible to an author to have two different styles. The review of Machado's specialized critics, the textual statistics and the software Hyperbase are the main methodological tools. The research analyses data-basis witch contains all the short stories written by this author, and compares it with all his novels and the Portext basis. The exploratory analysis of the data allows to describe the material anatomy of Machado's short stories, as the statistics functions allows the search for patterns and transformations in two elements: the lexicon and the text distribution. Finding low variations of grammar classes and vocabulary, the quantitative analysis challenges the idea of stylistic rupture and reinforces the hypothesis that style transformation of Machado's short stories is gradual and can be found mainly in the composition, instead of being at the linguistic material.

Keywords: Machado de Assis. Short story. Textual Statistics. Style.

LISTA DE QUADROS

Quadro 1 –	Códigos dos contos de Machado de Assis por volume e bases integrantes do <i>corpus</i>	73
Quadro 2 –	Lista de textos e autores da base Portext e respectivos códigos e gêneros	78
Quadro 3 –	Ordem dos textos nas bases	82
Quadro 4 –	Evolução do uso de substantivos, verbos, outros e nomes próprios nos volumes de contos publicados por Machado de Assis	85
Quadro 5 –	Evolução do uso de substantivos, verbos, outros e nomes próprios nos volumes de contos agrupados de Machado de Assis	86
Quadro 6 –	Substantivos e verbos mais usados por Machado de Assis nos volumes agrupados	87
Quadro 7 –	Substantivos e verbos mais usados por Machado de Assis nos contos publicados em livro	87
Quadro 8 –	Substantivos mais frequentes nos volumes de contos publicados por Machado de Assis e nos avulsos	88
Quadro 9 –	Verbos mais frequentes nos volumes de contos publicados por Machado de Assis e nos avulsos	89
Quadro 10 –	Quarenta verbos mais usados na língua portuguesa	89
Quadro 11 –	Verbos no infinitivo utilizados em todos os volumes de contos publicados por Machado de Assis	90
Quadro 12 –	Pontuação forte em cada um dos sete volumes publicados com valores reais (observados), teóricos e desvios (números relativos)	121
Quadro 13 –	Extensão média de frase nos volumes de contos de Machado de Assis	124
Quadro 14 –	Extensão média de frase nos volumes de contos agrupados e romances de Machado de Assis, e no Portext	125

LISTA DE GRÁFICOS

Gráfico 1 –	Evolução do uso de substantivos, verbos, outros e nomes próprios nos volumes de contos publicados de Machado de Assis	85
Gráfico 2 –	Análise fatorial dos substantivos na base Flumeia	93
Gráfico 3 –	Análise em árvore dos substantivos na base Flumeia	93
Gráfico 4 –	Análise fatorial dos substantivos na base AVG1	94
Gráfico 5 –	Análise em árvore dos substantivos na base AVG1	95
Gráfico 6 –	Análise fatorial dos substantivos na base 5VOL	96
Gráfico 7 –	Análise em árvore dos substantivos na base 5VO	96
Gráfico 8 –	Análise fatorial dos substantivos na base AVG2	97
Gráfico 9 –	Análise em árvore dos substantivos na base AVG2	97
Gráfico 10 –	Análise fatorial dos substantivos na base 7VOL	98
Gráfico 11 –	Análise em árvore dos substantivos na base 7VOL	99
Gráfico 12 –	Análise fatorial dos substantivos mais usados nos contos publicados de Machado de Assis na base Portext	100
Gráfico 13 –	Análise em árvore radial dos substantivos mais usados nos contos publicados de Machado de Assis na base Portext	100
Gráfico 14 –	Riqueza lexical na base Flumeia	103
Gráfico 15 –	Riqueza lexical na base AVG1	104
Gráfico 16 –	Riqueza lexical na base 5VOL	105
Gráfico 17 –	Riqueza lexical na base AVG2	106
Gráfico 18 –	Riqueza lexical na base 7VOL (todos os volumes publicados)	107
Gráfico 19 –	Riqueza lexical na base Romances	108
Gráfico 20 –	Riqueza lexical na base Portext	109
Gráfico 21 –	Crescimento lexical na base 7VOL	110
Gráfico 22 –	Histogramas da riqueza lexical sobre hápax em Flumeia	112
Gráfico 23 –	Histogramas da riqueza lexical sobre hápax em AVG1	113
Gráfico 24 –	Histogramas da riqueza lexical sobre hápax em 5VOL	113
Gráfico 25 –	Histogramas da riqueza lexical sobre hápax em AVG2	114

Gráfico 26 –	Histogramas da riqueza lexical sobre hápax em 7VOL	122
Gráfico 27 –	Histogramas da riqueza lexical sobre hápax em todos os romances	123
Gráfico 28 –	Histogramas da riqueza lexical sobre hápax na base Portext	124
Gráfico 29 –	Desvio da pontuação forte nos sete volumes publicados	129
Gráfico 30 –	Extensão média de frase nos volumes de contos de Machado de Assis	133
Gráfico 31 –	Extensão dos textos, calculada sobre as ocorrências (N) em 7VOL	134
Gráfico 32 –	Evolução do uso do ponto em Flumeia	135
Gráfico 33 –	Evolução do uso do ponto em AVG1	136
Gráfico 34 –	Evolução do uso do ponto em 5VOL	137
Gráfico 35 –	Evolução do uso do ponto em AVG2	138
Gráfico 36 –	Evolução do uso do ponto em 7VOL	139
Gráfico 37 –	Evolução do uso do ponto nos romances	140
Gráfico 38 –	Evolução do uso da vírgula em Flumeia	141
Gráfico 39 –	Evolução do uso da vírgula em AVG1	142
Gráfico 40 –	Evolução do uso da vírgula em 5VOL	143
Gráfico 41 –	Evolução do uso da vírgula em AVG2	144
Gráfico 42 –	Evolução do uso da vírgula em 7VOL	145
Gráfico 43 –	Evolução do uso da vírgula nos romances	146
Gráfico 44 –	Análise em árvore da distância lexical na base R1	149
Gráfico 45 –	Análise em árvore da distância lexical na base R2	150
Gráfico 46 –	Análise em árvore da distância lexical em todos os romances	151
Gráfico 47 –	Análise em árvore (radial) da distância lexical na base Portext	152
Gráfico 48 –	Análise em árvore (retangular) da distância lexical na base Portext	153
Gráfico 49 –	Análise em árvore da distância lexical na base Flumeia	154
Gráfico 50 –	Análise em árvore da distância lexical na base AVG1	155
Gráfico 51 –	Análise em árvore da distância lexical na base 5VOL	156
Gráfico 52 –	Análise em árvore da distância lexical na base AVG2	157
Gráfico 53 –	Análise em árvore da distância lexical em todos os volumes	158

LISTA DE TABELAS

Tabela 1 –	Indicadores de extensão do <i>corpus</i> em todos os contos de Machado de Assis	77
Tabela 2 –	Indicadores de extensão do <i>corpus</i> nos cinco volumes de contos publicados em livro por Machado de Assis	78
Tabela 3 –	Pontuação forte nos volumes de contos de Machado de Assis	120
Tabela 4 –	Percentual de pontuação forte pelo número de ocorrências dos volumes	120
Tabela 5 –	Teste de Associação Qui-quadrado (X^2) por volume (variáveis: cronologia e pontuação forte)	122
Tabela 6 –	Pontuação forte nos volumes de contos agrupados de Machado de Assis	122
Tabela 7 –	Pontuação forte nos volumes de contos agrupados de Machado de A	123
Tabela 8 –	Pontuação forte em todos os contos de Machado de Assis, por grupos: valores reais, téóricos e desvios (números relativos)	123
Tabela 9 –	Teste de Associação Qui-quadrado (X^2) por grupo (variáveis: cronologia e pontuação forte)	123

SUMÁRIO

INTRODUÇÃO	13
CAPÍTULO 1 – OS CRÍTICOS E A TEORIA DO <i>TWICE BORN</i>.....	17
1.1 Machado de Assis: Processo ou Ruptura?.....	17
1.2 O Projeto Estético-Literário de Machado de Assis	24
1.2.1 A nacionalidade como tema	26
1.2.2 A crítica ao Naturalismo.....	30
1.3 O Conto e o Projeto Machadiano	34
1.3.1 O conto como laboratório do romance	36
1.3.2 O conto como parte do projeto estético.....	38
CAPÍTULO 2 – O DESAFIO TEÓRICO-METODOLÓGICO DOS ESTUDOS LITERÁRIOS AUXILIADOS POR COMPUTADOR	41
2.1 Um Novo Campo.....	41
2.1.1 Características.....	41
2.1.2 Publicações e fontes	43
2.1.3 No Brasil.....	44
2.2 Histórico	45
2.2.1 A vertente de língua inglesa	45
2.2.2 A vertente francesa	47
2.3 Principais Contribuições.....	50
2.3.1 <i>Junius Letters</i>	50
2.3.2 <i>Federalist Papers</i>	51
2.3.3 Análise multidimensional ou multivariada.....	53
2.3.4 Distinção de gênero literário.....	54
2.3.5 Análise temática	55
2.4 Nasce uma Nova Crítica?	56
CAPÍTULO 3 – ESTILO E QUANTIFICAÇÃO.....	61
3.1 Questões de Estilo	61
3.2 A Ferramenta Hyperbase	67
3.2.1 Conceitos básicos	68
3.3 O <i>Corpus</i>	69
3.3.1 Características do <i>corpus</i>	72

3.3.2 Léxico, convenções e dificuldades	80
3.3.3 Ordem dos textos no Hyperbase	81
CAPÍTULO 4 – AS ALTAS FREQUÊNCIAS	83
4.1 Altas Frequências de Nomes Próprios, Verbos, Substantivos e Outros	83
4.2 Análise Fatorial e em Árvore dos Substantivos	90
CAPÍTULO 5 – RIQUEZA LEXICAL E HÁPAX	102
5.2 O Crescimento Lexical	109
5.3 Hápax	111
CAPÍTULO 6 – A PONTUAÇÃO E A EXTENSÃO DE FRASE	118
6.1 Pontuação	118
6.1.1 O ponto	127
6.1.2 A vírgula	133
CAPÍTULO 7 – DISTÂNCIA LEXICAL	140
7.1 O Romance	141
7.2 Machado, Outros Autores e Outros Gêneros	144
7.3 O Conto	146
CONSIDERAÇÕES FINAIS.....	151
REFERÊNCIAS BIBLIOGRÁFICAS	154
BIBLIOGRAFIA	158
APÊNDICE	171

INTRODUÇÃO

*Um clássico é uma obra que provoca incessantemente uma nuvem de discursos críticos sobre si, mas continuamente as repele para longe*¹.

Machado de Assis foi, de fato, um autor *twice-born*? Houve, de fato, uma ruptura estética radical entre os escritos anteriores a 1880 e os posteriores? Ou o que houve foi simplesmente o desenvolvimento da escrita de um grande autor, graças ao estudo, ao exercício, à autocrítica, acrescidos da experiência trazida pela maturidade? O grande Machado já estava presente nos primeiros escritos de alguma forma?

É em busca de respostas para essas questões que escrevemos o presente trabalho.

Harold Bloom define cânone como tudo aquilo que é “obrigatório em nossa cultura”². Machado de Assis é referência obrigatória, não só em literatura, na qual é o nosso maior escritor, como também na crítica, em que também se destacou.

Ainda hoje, passados quase cem anos da morte de Machado, o Brasil é um país dependente e periférico. Pior do que isso é ser o campeão das desigualdades sociais e econômicas na América Latina³. Embora já não haja senhores que possam montar em seus escravos como fazia o jovem Brás Cubas, há ainda no Brasil (e não só aqui) outras formas de escravidão e de humilhação impostas e sofridas pelos homens. O poder, o dinheiro e a corrupção continuam bem casados, o que mantém atuais muitas das narrativas machadianas, como *Conto de escola* ou *Teoria do medalhão*.

O fato de nosso país se manter dependente e periférico (embora essa condição tenha configuração diferente da época de Machado), padrão de desenvolvimento em que as relações de compadrio ainda valem, mesmo que de forma disfarçada, é apenas um dos aspectos que confere atualidade à obra machadiana. É claro que sua atualidade não se reduz a isso. O autor ainda hoje é dos mais produtivos (se não o mais) de nossa história literária, tendo experimentado todos os gêneros e se tornado mestre no conto, no romance e na crítica.

¹ CALVINO, Ítalo. *Por que ler os clássicos*. Tradução de Nilson Moulin. São Paulo: Companhia das Letras, 1994, p. 12.

² BLOOM, Harold. *O cânone ocidental*. Tradução de Marcos Santarrita. Rio de Janeiro: Objetiva, 1995, p. 23.

³ Segundo dados do Banco Mundial, no Brasil, os 10% mais pobres recebem 0,9% da renda do país, enquanto os 10% mais ricos ficam com 47,2% (FOLHA ONLINE, 2003).

O conto machadiano é o objeto da presente pesquisa. Trabalhos exclusivos sobre o assunto são escassos, embora a crítica não só tenha, por diversas vezes, reconhecido e reiterado a importância e a qualidade artística dos contos de Machado como também afirmado, por nomes de peso da crítica como Augusto Meyer e Lúcia Miguel-Pereira, que é nesse gênero que o autor teria tido seu melhor desempenho como escritor.

O próprio Machado sempre defendeu o conto, considerando-o “gênero difícil, a despeito da sua aparente facilidade”⁴. Sua afinidade com ele foi tão grande que o cultivou dos 19 anos até dois anos antes de morrer, somando 205 textos, dos quais somente 76 selecionou para publicação em livro.

Contrariamente à abundância de trabalhos sobre o romance, são poucos os estudos dedicados exclusivamente aos contos. Vários autores já o constataram⁵. Pela falta de estudos sistemáticos sobre o assunto, geralmente lembrado como tema das sempre populares antologias, reiteramos a opinião de Gledson de que os contos de Machado “parecem ter ficado em segundo plano” e que, “a despeito de sua popularidade, não são levados tão a sério quanto mereceriam”⁶.

Quanto ao nosso trabalho, ele toma como premissa a contestação de uma afirmativa crítica que nasceu quase que concomitantemente com a obra de Machado: a da existência de uma ruptura radical que transformou o autor em dois, o da primeira fase e o da segunda. Essa afirmação tem uma origem tão remota que Sílvio Romero – logo ele – já se contrapunha a ela em 1897.

Nosso objetivo principal é investigar se essas fases existiram de fato devido a uma ruptura, e não como processo no conto machadiano. Como nosso *corpus* é formado pelo total de contos, 205 textos, optamos pelo recurso à estatística textual auxiliada por computador. A partir de um programa de análise de grandes grupos de textos, obtivemos, a princípio, um dicionário com todas as palavras usadas pelo autor e as listas de frequências capazes de nos dar o número de vezes que determinada palavra é repetida, bem como a possibilidade de identificar onde as palavras aparecem no texto. Além dessas ferramentas exploratórias, descritivas, o programa Hyperbase⁷ dispõe de funções estatísticas capazes de viabilizar a comparação entre textos de diferentes extensões.

⁴ MACHADO DE ASSIS, Joaquim Maria. *Obras completas*. Rio de Janeiro: Nova Aguilar, 1997, v. III, p. 806.

⁵ GLEDSON, John. *Por um novo Machado de Assis: ensaios*. São Paulo: Companhia das Letras, 2006, p. 35; DIXON, Paul B. *Os contos de Machado de Assis: mais do que sonha a filosofia*. Porto Alegre: Movimento, 1992, p. 10; CUNHA, Patrícia Lessa Flores da. *Machado de Assis: um escritor na capital dos trópicos*. Porto Alegre: IEL; São Leopoldo: Unisinos, 1998, p. 17.

⁶ “O machete e o violoncelo”. GLEDSON, op. cit., p. 35.

⁷ BRUNET, Etienne. *Hyperbase* : manuel de référence (Version 5.4 par Windows), janv. 2002.

Os critérios qualitativos usados para a divisão do *corpus* para fins comparativos foram o gênero – comparar textos de mesma natureza – e a cronologia, pois, se queremos verificar a distinção de fases, é necessário respeitar a divisão cronológica feita pela crítica. Como o marcador da transição na prosa machadiana foi *Memórias póstumas de Brás Cubas*, em 1881, optamos por esse ano para dividir os contos. Embora a crítica trate *Papéis avulsos*, de 1882, como o divisor das fases no gênero, optamos pela data de publicação das *Memórias* porque ela é anterior, ou seja, se o autor já havia feito a transição com a publicação do romance, em 1881, e o livro fora escrito antes, os contos que foram produzidos após 1880 já estariam na segunda fase. Outro critério usado foi relativo à publicação *em livro*. Os contos publicados tiveram sua divisão em volumes respeitada, somando 76 textos. Os não-publicados serão tratados como Avulsos e também foram divididos por ordem cronológica: um grupo formado pelos textos escritos entre 1858 (data do primeiro conto) e 1880; e um segundo grupo, de 1881 até 1906, data do último conto.

Para viabilizar o tratamento estatístico feito por computador, foram necessárias versões digitalizadas das obras de Machado de Assis. Nossas fontes para esse fim foram o sítio do núcleo de Pesquisas em Informática Literatura e Linguística (Nupill)⁸, da Universidade Federal de Santa Catarina, e o sítio sobre o conto machadiano de Cláudio Weber Abramo⁹. As versões digitalizadas do Nupill têm como fonte as *Obras completas* de Machado, publicadas pela Nova Aguilar, que o sítio de Weber Abramo também utiliza (embora, para os Avulsos, conte com outras fontes).

Além de nosso *corpus*, utilizaremos para fins comparativos os romances de Machado de Assis, já que fazem parte da prosa de ficção do autor junto com o conto. Usaremos também a base de dados Portext¹⁰, que congrega textos de diferentes gêneros, autores e épocas da literatura brasileira.

O estudo está dividido em seis capítulos. No primeiro, problematizamos questões apresentadas pela crítica, notadamente a ideia de “dois Machados” – contra a qual defendemos a noção de projeto, que contempla o exercício, a autocrítica e o amadurecimento contínuo e gradual do estilo do autor. Para isso, amparados na fortuna crítica, apontamos o papel crucial do conto. No segundo capítulo, apresentamos os fundamentos histórico-epistemológicos do campo da estatística textual. No terceiro capítulo, detalhamos o modelo de análise, cujo ponto de partida é o conceito de estilo como um desvio em relação à norma,

⁸ <http://www.nupill.org/>.

⁹ <http://www.uol2.com.br/machadodeassis/fim.html>.

¹⁰ Desenvolvida pelo Prof. Dr. Carlos Maciel, da Universidade Sophia Antipolis, de Nice (França), pesquisador do mesmo laboratório que desenvolveu o *Hyperbase*.

como afirma Pierre Guiraud, que complementa: “A estatística, com efeito, é a ciência dos desvios e a especificidade da obra é entre outras coisas um desvio em relação a uma norma, que resta, evidentemente, a definir”¹¹. Daí nossa opção pela estatística textual. No quarto capítulo, pelos aspectos relacionados às altas frequências, iniciamos a apresentação dos dados e o desenvolvimento da interpretação do conto machadiano, em diálogo com a base Portext. O Capítulo 5 é dedicado à descrição do léxico (riqueza lexical e hápax); e, no capítulo sexto, a análise é concluída, a partir dos gráficos, quadros e tabelas relativos à pontuação.

Este trabalho foi parcialmente financiado por bolsas do CNPq (de dezembro de 2002 a dezembro de 2004) e Capes (de fevereiro a junho de 2005/PDEE). Em função de imperativos administrativos, relacionados ao cumprimento de exigências impostas pela política de concessão de bolsas desses órgãos oficiais, este trabalho foi enviado aos avaliadores sem que nem mesmo a autora o julgasse concluído. A tirania dos prazos limitou a possibilidade de intervenção do orientador e restringiu a menos de duas semanas o período para a leitura dos integrantes da banca. O tempo da burocracia derrotou o tempo do pensamento; esse fato deve ser levado em conta na apreciação deste trabalho.

¹¹ GUIRAUD, Pierre. *Essais de stylistique*. Paris: Klincksieck, 1985, p. 16.

CAPÍTULO 1 OS CRÍTICOS E A TEORIA DO *TWICE BORN*¹²

1.1 Machado de Assis: Processo ou Ruptura?

Em ensaio publicado em 1959 no jornal o Estado de S.Paulo, Otto Maria Carpeaux comenta o estudo “De Machadinho a Brás Cubas”, de Augusto Meyer. Neste, Meyer foca sua análise nas tentativas de explicar a mudança estilística de Machado de Assis – evidenciada pela publicação de *Memórias póstumas de Brás Cubas*, em 1881 –, a partir da ideia de uma evolução gradual. Segundo Carpeaux, Meyer “compara essas tentativas à situação da biologia antes de se terem descoberto as mutações bruscas, admitidas pela teoria mendeliana”¹³, ou seja, para Meyer, com quem Carpeaux concorda, a ideia de um desenvolvimento gradual do estilo serviria a nossa necessidade e desejo de coerência.

Na tentativa de explicar essa “mutação brusca”, Carpeaux afirma que “o instrumento psicológico foi forjado por William James, em *The Varieties of Religious Experience: a Study in Human Nature*. É a teoria de que certas pessoas, excepcionais nesse ou naquele sentido, passam por um “segundo nascimento”¹⁴. Como na história bíblica em que é dito a Nicodemus que só os que nascerem de novo poderão ver o reino de Deus, William James

[...] limitou essa necessidade de renascer no espírito a certos eleitos, aos *twice-born*. Estes se encontram com freqüência inquietante entre os grandes espíritos religiosos, que passaram por conversão repentina: Lutero, Pascal, Kierkgaard; e, sobretudo o apóstolo São Paulo. São os eleitos de Deus. Mas aos defensores desta ou aquela ortodoxia, um ou outro daqueles *twice-born* antes se afigurará eleito do Diabo. O renascimento poderia ter como conseqüência a perda da fé; pode ser anticonversão. E seria esse o caso do *twice-born* Machado de Assis¹⁵.

Comparando com o caso de seu contemporâneo italiano Giovanni Verga, cujo público e crítica rejeitaram sua “conversão”, Carpeaux afirma que, em relação a Machado,

¹² A expressão em inglês remete a BOSI, Alfredo et al. *Machado de Assis*. São Paulo: Ática, 1982, p. 440: “Todos reconhecem nas *Memórias póstumas* o divisor de águas da obra machadiana. Otto-Maria Carpeaux chegou a falar em Machado de Assis como um desses raros escritores *twice born*, nascidos duas vezes, à maneira dos convertidos Santo Agostinho ou Pascal”. Bosi não indica com precisão em que texto Carpeaux teria afirmado isso; no único texto do crítico citado na bibliografia geral do livro não há o que justifique tal menção.

¹³ CARPEAUX, Otto Maria. *Ensaaios Reunidos (1946-1971)*. Rio de Janeiro: UniverCidade; Topbooks, 2005, p. 456.

¹⁴ CARPEAUX, op. cit., p. 457-458.

¹⁵ CARPEAUX, op. cit., p. 458.

Erros de julgamento também os houve e ainda os há [...] porque falsa piedade ou supervalorização de fatos estilísticos ou aquela teoria genético-evolucionista não querem admitir a inferioridade da primeira fase de Machado, o que significa diminuir a superioridade das obras da segunda fase.

Grande parte da crítica reproduz a ideia de que Machado é um autor dividido em duas fases: primeira e segunda, romântica e realista, juvenil e madura. Como afirma uma personagem de Machado¹⁶, “se uma coisa pode existir na opinião, sem existir na realidade, e existir na realidade, sem existir na opinião, a conclusão é que das duas existências paralelas a única necessária é a da opinião”. Será essa a situação desse dogma da crítica machadiana? Quando uma ideia está de tal modo sedimentada na crítica, acabamos por aceitá-la como verdadeira, sem valorizarmos as vozes dissonantes e, pior ainda, sem nos darmos ao trabalho de verificar até que ponto a obra corrobora de fato aquela afirmação.

São recorrentes as vezes em que o próprio Machado enfatiza o trabalho do tempo na formação de um escritor, de uma escola, de uma literatura. Desde cedo, mais precisamente em texto de 1858 – ano da escrita de seu primeiro conto, *Três tesouros perdidos* –, Machado já manifesta essa crença em relação à necessidade de emancipação da literatura brasileira:

Uma revolução literária e política fazia-se necessária. O país não podia continuar a viver debaixo daquela dupla escravidão que o podia aniquilar. [...] Mas após o *fiat* político, devia vir o *fiat* literário, a emancipação do mundo intelectual, vacilante sob a ação influente de uma literatura ultramarina. Mas como?¹⁷

Tal crença será reafirmada no clássico *Instinto de nacionalidade*, de 1873: “Esta outra independência não tem Sete de Setembro nem campo de Ipiranga; não se fará num dia, mas pausadamente, para sair mais duradoura; não será obra de uma geração nem duas; muitas trabalharão para ela até perfazê-la de todo”¹⁸.

Um dos primeiros críticos a enfatizar o desenvolvimento gradual de Machado foi Sílvio Romero. Apesar das posições equivocadas que tomou em relação à obra de Machado no que tange à “ausência de cor local”, bem como às explicações pseudocientíficas e preconceituosas que associam os traços de estilo à gagueira ou a uma suposta timidez que Romero atribui à origem, racial ou de classe, do autor, o crítico foi dos primeiros a ver na obra

¹⁶ “O segredo do Bonzo”. MACHADO DE ASSIS, op. cit., p. 325.

¹⁷ “O passado, o presente e o futuro da literatura”. MACHADO DE ASSIS, op. cit., v. III, p. 787.

¹⁸ MACHADO DE ASSIS, op. cit., v. III, p. 804.

de Machado um desenvolvimento gradual. Em seu livro sobre Machado, escrito em 1897, quando o escritor já tinha seu nome consagrado, Romero posiciona-se contra a “apregoadá antinomia entre a primeira e a segunda fase da carreira do ilustre autor, entre a sua antiga *maneira* e a que depois adotou”¹⁹:

Julgam geralmente que existe um valo quase invariável entre os dois períodos.

A nova maneira de Machado de Assis não estava em completa antinomia como seu passado, sendo apenas o desenvolvimento normal de bons germes que ele nativamente possuía [...] ²⁰.

Para Romero, Machado

tem certamente em si vivos os sinais de evolução e progresso; mas esses não se fizeram como antítese de suas primeiras revelações na arena das lides espirituais, e sim como normal continuação e desdobramento delas.

O progresso tem consistido no melhor manejo da linguagem, na maior correção do estilo, no mais apurado da observação, no mais penetrante da análise, no alargamento das idéias²¹.

Outro medalhão da crítica, contemporâneo de Romero, José Veríssimo confirma a posição de seu colega. Para ele, “ao contrário de alguns notáveis escritores nossos que começaram pelas suas melhores obras e como que nelas se esgotaram, tem Machado de Assis uma marcha ascendente. Cada obra sua é um progresso sobre a anterior”²².

Apesar de afirmar a marcha ascendente de Machado, Veríssimo acaba por se contradizer, de certa forma, ao sucumbir ao senso comum da crítica. Não só incorpora o discurso da ruptura como também filia Machado ao Romantismo. “As *Memórias Póstumas de Brás Cubas* eram o rompimento tácito, mais completo e definitivo de Machado de Assis, com o Romantismo sob o qual nascera, crescera e se fizera escritor”, afirma, para logo depois atenuar essa influência:

Aliás, conquanto necessariamente lhe sofresse a influência, nunca jamais se lhe entregara totalmente nem lhe sacrificara o que de pessoal e original havia no seu engenho, e acharia em Brás Cubas a sua cabal expressão²³.

¹⁹ ROMERO, Sílvio. *Machado de Assis*. 2. ed. Rio de Janeiro: José Olympio, 1936, p. 24-25. (Grifo do autor.) Note-se a longevidade do dogma do Machado *twice-born*. Como em todo mito, é difícil caracterizar sua origem.

²⁰ ROMERO, op. cit., p. 24-25.

²¹ ROMERO, op. cit., p. 25.

²² VERÍSSIMO, José. *História da Literatura Brasileira*. VERÍSSIMO, José. *História da Literatura Brasileira*. Rio de Janeiro: Francisco Alves, 1916, p. 214.

²³ VERÍSSIMO, op. cit., p. 217.

Para Veríssimo, as primeiras obras de prosa de ficção de Machado (*Contos fluminenses, Histórias da meia-noite, Ressurreição, A mão e a luva, Helena e Iaiá Garcia*)

traziam ressaibos românticos, embora atenuados pelo congênito pessimismo e nativa ironia do autor. Ora, o Romantismo não comportava nem a ironia nem o pessimismo, na forma desenganada, risonha e resignada de Machado de Assis. [...] Desde, portanto, os anos de 70, renunciando ao escasso Romantismo que nele havia, criava-se Machado de Assis uma maneira nova, muito sua, muito particular e muito distinta e por igual estreme daquela escola e das novas modas literárias.²⁴

A complexidade e a originalidade da obra de Machado, de fato, desnortearam a crítica, principalmente a de sua época, quando predominava o viés determinista. A riqueza do projeto estético-literário do escritor chocava-se com o reducionismo imposto pela crítica. Independentemente da inflexão desta, é certo que o biografismo e o determinismo que a assombravam empobreceram a análise da obra machadiana. Raymundo Faoro cria uma imagem bastante interessante dessa situação:

Um pobre homem do século XIX, emparedado nos preconceitos e na perspectiva do seu tempo, desvenda o quadro social do Segundo Reinado, com luz feita de aproximações, tímida e sombreada. Ele caminha tateando o terreno, passo a passo, avançando e recuando, enganando os contemporâneos, testemunhas de igual momento. Houve quem, convivendo na mesma hora com a gente evocada no papel impresso, lhe negou a autenticidade do retratista. Ao lado do crítico embaído, um escritor, mais próximo da realidade, reconheceu-o genuíno homem de seu meio, lutando, pensando e escrevendo como homem do seu tempo. Dois juízos, o áspero e o benévolo, mordidos de equívocos semelhantes, preocupados em reduzir o escritor ao homem, e o homem à época, à concreta situação social e histórica, e daí armar o critério de validade poética e literária da obra²⁵.

Apesar das divergências entre Romero e Veríssimo, além de terem concepções semelhantes quanto ao amadurecimento da escrita de Machado, ambos apontam uma mesma característica já presente nos escritos da juventude que ficará mais forte nos da maturidade: o que o primeiro chama de “observação da realidade”, o segundo chama “análise do ambiente”.

Outro crítico canônico corrobora a ideia do amadurecimento gradual. Afrânio Coutinho afirma explicitamente que descarta a ideia de uma mutação repentina na forma de escrever de Machado. Para ele,

²⁴ VERÍSSIMO, op. cit., p. 217-218.

²⁵ FAORO, Raymundo. *Machado de Assis: a pirâmide e o trapézio*. 4. ed. ver. São Paulo: Globo, 2001, p. 525.

não há ruptura brusca entre as duas fases. É mais justo afirmar que uma pressupõe a outra, e por ela foi preparada. Há, antes, continuidade. E, se existe diferença, não há oposição, mas sim desabrochamento, amadurecimento. Isto sim: maturação. O desenvolvimento de Machado de Assis é um longo processo de maturação, ao longo do qual vai acumulando experiência e fixando vivências, que gerarão o seu credo espiritual e estético e a sua concepção técnica. Nada disso resultou de uma modificação súbita nem por geração espontânea no espírito do escritor, mas de transformação lenta em zonas profundas e obscuras, na intimidade das fontes vitais²⁶.

Mais uma vez, temos a ênfase na “análise de costumes” e no “gosto psicológico”. Além disso, Coutinho destaca o humor como outra característica presente nos livros iniciais. Mas um humor de natureza diferente do que veríamos na maturidade, ainda “faceto, quase alegre”, o que o autor atribui ao fato de a primeira etapa ainda “estar embebida do sentimentalismo romântico”. Afrânio reforça tanto a posição de Romero quanto a de Veríssimo e o faz com quase as mesmas palavras ao afirmar que nos romances da primeira fase encontram-se “em germe” os recursos técnicos e estilísticos que seriam posteriormente desenvolvidos e apurados pelo autor e na afirmação de um progresso “constante e ascensional”²⁷. Em relação a Veríssimo, embora este não o diga de modo tão explícito, Coutinho tem mais um ponto em comum, que é o de considerar o desenvolvimento da escrita de Machado mais do que um amadurecimento pretensamente natural, devido ao exercício da literatura ao longo do tempo, mas como algo intencional, produto de um processo “gradativo e consciente”. Isso nos leva a outro ponto de discussão deste capítulo, a saber, o projeto estético-literário de Machado de Assis.

Antes, porém, de entrar em novo tema, vejamos ainda alguns autores que nos dão argumento para considerar o desenvolvimento gradual da escrita de Machado, em contraste com a tese da ruptura. Um exemplo vem de Jean-Michel Massa. Ao tratar das *Aquarelas* (textos publicados por Machado na revista *O Espelho*), o autor afirma que ali já se encontravam “muitos dos caracteres que descreveu mais tarde em seus romances”²⁸.

Outro adepto da ideia de continuidade em relação à propagada ruptura é Massaud Moisés. Ele parte do princípio de que as obras de um autor, por mais divergentes que sejam, identificam-se pelos mesmos denominadores comuns, o que chamamos de padrões. Moisés acredita que, apesar do designativo das fases, isso não implica necessariamente que haja características identificadoras, em razão da doutrina estética perfilhada. Segundo ele, “não há

²⁶ COUTINHO, Afrânio. *Machado de Assis na literatura brasileira*. 2. ed. Rio de Janeiro: São José, 1966, p. 16.

²⁷ COUTINHO, op. cit., p. 17.

²⁸ MASSA, Jean Michel. A juventude de Machado de Assis. In: BOSI, Alfredo et al. *Machado de Assis*. São Paulo: Ática, 1982, p. 397.

separação fundamental entre elas. As diferenças são, em geral, de gradação, de ênfase ou de perspectiva”²⁹. O crítico considera que algumas das características mais relevantes da obra de ficção de Machado já estavam presentes em seu primeiro romance – sem, no entanto explicitá-las.

Machado, em carta de agradecimento a José Veríssimo, por uma crítica feita a um “velho livro e seu velho autor”, expressa o que pensa sobre a divisão de sua obra em duas etapas: “O que você chama a minha segunda maneira naturalmente me é mais aceita e cabal que a anterior, mas é doce achar quem se lembre desta, quem a penetre e desculpe, e até chegue a catar nela algumas raízes de meus arbustos de hoje”³⁰. O livro era *Iaiá Garcia*, o ano, 1898.

Essa opinião deixa claro que o próprio autor via suas mudanças estilísticas como um processo, e não como uma ruptura que apartaria artificialmente o primeiro Machado do segundo. Vários autores já se pronunciaram a respeito do autodidatismo do escritor, que foi crescendo em qualidade gradativamente ao longo de sua carreira, para chegar à maturidade artística aos 40 anos. Sua produção forma um “todo coerentemente organizado”, no qual “à medida que seus textos se sucedem cronologicamente, certas estruturas primárias e primeiras se desarticulam e rearticulam sob forma de estruturas diferentes, mais complexas e mais sofisticadas”³¹.

Um dos maiores nomes da crítica machadiana, Roberto Schwarz, embora adote a divisão em fases da obra do Machado, aponta uma série de elementos comuns aos primeiros romances e aos da maturidade. Uma das análises mais pertinentes – que, no mínimo, derruba a denominação do período das primeiras obras como “fase romântica” – é a afirmação de que Machado, ao conceber “já de entrada” o Romantismo como “ideologia de segundo grau”, “escapava às implicações desse percurso”, mas “via-se a braços com um problema novo e capital, para o qual só mais tarde acharia solução: qual a curva própria à vida de suas figuras? Qual a forma para o seu enredo?”³²

Ainda sobre os primeiros romances, um elemento que será recorrente na obra machadiana já está presente: “também eles trazem na composição a marca da dependência nacional”, afirma Schwarz. Tais textos são, porém, considerados pelo crítico como

²⁹ MOISÉS, Massaud. A ficção machadiana: ressurreição e permanência. In: *Machado de Assis: ficção e utopia*. São Paulo: Cultrix, 2001, p. 35-40 (publicado originalmente no “Suplemento Literário” de *O Estado de S.Paulo*, em 16/05/1959).

³⁰ MACHADO DE ASSIS, op. cit., p. 1044.

³¹ SANTIAGO, Silviano. Retórica da verossimilhança. In: *Uma Literatura nos Trópicos: ensaios sobre dependência cultural*. 2. ed. Rio de Janeiro: Rocco, 2000, p. 27.

³² SCHWARZ, Roberto. *Ao vencedor as batatas*: forma literária e processo social nos inícios do romance brasileiro. São Paulo: Duas Cidades; Ed. 34, 2000, p. 99-100.

“deliberada e desagradavelmente conformistas”³³. Nessa afirmação, o autor aponta tanto para um denominador comum, o tema da dependência, quanto para um ponto de afastamento, o caráter conformista, entre os primeiros romances e os da fase madura, ou seja, aponta para uma mudança “ *muito mais de enfoque do que propriamente de rumo*, atualiza-se e justifica-se na própria ideia de *processo* que subjaz à composição da sua obra”, como bem afirma Patrícia Lessa Flores da Cunha em livro sobre a contística de Machado³⁴.

Em sua análise de *A mão e a luva*, Schwarz mostra como o autor ainda está em busca de uma forma que contemple as idiossincrasias que pretende retratar: “impostura honesta, simulação sincera e mais outros paradoxos, o movimento repete-se e consiste em suspender o sistema das oposições românticas, depois de o ter trazido à baila”³⁵. Esses paradoxos seriam traduzidos por Machado de um modo em que:

a terminologia do cinismo e da virtude são levadas a coincidir, e o comportamento condenado é exatamente o que convém. Em suma, o movimento da cooptação entrosa cálculos e sentimentos numa mesma aspiração, e modifica os termos do problema, guardando-lhes, no entanto a nomenclatura conflituada. Os lados ignóbeis desta ‘harmonia’ estarão entre as matérias favoritas do segundo Machado. Já o primeiro tem nela um ideal. Assim, o desmentido que a realidade inflige às apreciações românticas veio a ser um elemento formal, algo como um timbre de prosa³⁶.

Sobre o fundamento da singular brasilidade de Machado, “sem pitoresco” – que, conforme Schwarz, estaria na maneira como o autor se apropria e retrata as contradições ligadas ao convívio das ideias liberais com a cultura paternalista brasileira, ancorada em práticas como as relações de favor –, é certo que “só na segunda fase” essa lógica estaria “desenvolvida sem entraves”, afirma ele. *A mão a luva*, conclui, “elabora-lhe alguns elementos e beneficia deles, subordinados, porém, à inconseqüência rigorosa da literatura leve”³⁷.

Como um autor consciente dos problemas da literatura de seu país e época, Machado cultivou a reflexão sobre o assunto aliando-a a sua produção literária, que se tornou mais um elemento para análise. Num movimento de interação contínua, uma atividade alimentou a outra, gerando um projeto em constante revisão e, conseqüentemente, transformação.

É disso que trataremos a seguir.

³³ SCHWARZ, op. cit., p. 83.

³⁴ CUNHA, op. cit., p. 55. (Grifos da autora).

³⁵ SCHWARZ, op. cit., p. 97.

³⁶ SCHWARZ, op. cit., p. 98.

³⁷ SCHWARZ, op. cit., p. 101.

1.2 O Projeto Estético-Literário de Machado de Assis

*No estado atual das cousas, a literatura não pode ser perfeitamente um culto, um dogma intelectual, e o literato não pode aspirar a uma existência independente, mas sim tornar-se um homem social, participando dos movimentos da sociedade em que vive e de que depende*³⁸.

No texto *O passado, o presente e o futuro da literatura*, publicado em 1858, Machado já se preocupava com a necessidade de uma nova literatura brasileira: “É mais fácil regenerar uma nação, que uma literatura. Para esta não há gritos de Ipiranga; as modificações operam-se vagorosamente; e não se chega em um só momento a um resultado”³⁹. À época, o autor ainda não havia publicado nem o primeiro livro de contos, nem seu primeiro romance, mas o crítico já apontava o sentido da obra do contista e do romancista.

No mesmo texto, o autor critica a escravização da poesia ao modelo europeu e proclama a necessidade da busca de uma emancipação intelectual do Brasil: a necessidade de, depois de alcançada a independência política, uma revolução literária. Machado afirma a inexistência do romance e do drama, considerando a falta deste último “uma vergonha” que faria do Brasil o país mais “parvo e pobretão das nações cultas”. Quando escreveu esse texto, Machado tinha apenas 19 anos. Corria o ano de 1858, período da produção de seu primeiro conto, *Três tesouros perdidos*.

Fica claro que Machado já tinha uma reflexão sobre a literatura desde muito cedo: antes de começar a escrever a prosa de ficção que o consagrou como artista, já tinha em mente o que desejava e o que rejeitava para si como modelo. Não desprezava as referências estrangeiras, mas enxergava a necessidade da criação de modelos próprios, mesmo que isso exigisse um processo longo e lento, que levaria gerações para se concretizar.

De acordo com o que observava, Machado conduziu sua produção. De poesia e teatro inexpressivos, transparecia no crítico não só a vocação para a prosa, mas a preocupação com o papel do escritor e com um fazer literário consciente. A inquietação do autor traduziu-se em ação: pôs-se a experimentar todos os gêneros literários. Os primeiros contos apareceram na década de 1860, no *Jornal das Famílias*, publicados em livro em 1870, nos *Contos*

³⁸ MACHADO DE ASSIS, op. cit.

³⁹ MACHADO DE ASSIS, op. cit., p. 785.

fluminenses. O primeiro romance foi publicado logo em seguida, em 1872, quando Machado tinha 33 anos.

Um conceito-chave para discutirmos o projeto de Machado é o de tradição, associado à ideia de literatura como sistema, defendida por Antonio Candido. Segundo ele, esse é um modo de ver a literatura como algo que transcenda a ordem cronológica e a simples listagem de autores e obras. Candido compartilha com Otto Maria Carpeaux o bom senso dos grandes intelectuais ao considerar a complexidade do fenômeno literário, tentando conciliar apreciação estética com preocupações culturais, ideológicas e históricas. O autor só considera literatura aquele conjunto de obras ligadas por denominadores comuns que se referem a características internas – como a língua ou a temática – e elementos de natureza social e psíquica que “fazem da literatura um aspecto orgânico da civilização”⁴⁰.

Os principais elementos apontados por Candido são a existência de autores “mais ou menos conscientes de seu papel”, um conjunto de leitores e um meio que os une, a obra. Para Candido, é sobre esse tripé que a literatura se origina como um “tipo de comunicação inter-humana” estruturado como sistema simbólico capaz de transcender a experiência individual e servir àquela comunicação referida e à “interpretação das diferentes esferas da realidade”⁴¹. A partir da integração de escritores de uma época a esse sistema, forma-se uma tradição (literária) porque sua inserção no sistema implica uma aceitação de determinados valores que estabelecem um paradigma para os que vierem depois. Só a partir dessa tradição, acredita Candido, é possível à literatura atingir o *status* de “fenômeno de civilização”⁴².

Em oposição a isso, temos o que o autor chama de “manifestações literárias”. Essas seriam constituídas das obras que aparecem isoladamente, que não são representativas de um sistema. O conceito não exclui a possibilidade de aparecimento de grandes autores ou de obras de boa qualidade literária. O fator determinante aqui é, na verdade, de fundo histórico.

As reflexões constantes sobre a situação da literatura e seus rumos revelam em Machado um escritor consciente. Candido o considera “o primeiro escritor que teve noção exata do processo literário brasileiro” e afirma que, no *Instinto de nacionalidade*, Machado pretendeu mostrar que “a absorção nos temas locais foi um momento a ser superado, e que a verdadeira literatura depende, não do registro de aspectos exteriores e modismos sociais”, mas

⁴⁰ CANDIDO, Antonio. *Formação da literatura brasileira*. 7. ed. Rio de Janeiro: Itatiaia, 1993, v. I, p. 23.

⁴¹ CANDIDO, op. cit., p. 23.

⁴² CANDIDO, op. cit., p. 24.

de algo que “assegure a sua universalidade”⁴³. Graças a sua tomada de posição, Machado pôs-se a trilhar um caminho singular.

1.2.1 A nacionalidade como tema

O tema da nacionalidade está no centro do debate literário da época. Em busca de autoafirmação, tanto os escritores quanto os críticos valorizavam tudo o que estivesse relacionado ao caráter nacional, à cor local, como se esse fosse o único caminho possível para fugir da reprodução dos modelos europeus, principalmente os portugueses, dada a nossa condição colonial.

É contra essa concepção única que Machado toma posição no já citado *Instinto de nacionalidade*. O instinto do título, que o autor diz reconhecer na literatura de sua época, não é o alvo de sua crítica. Com a elegância habitual, embora considere o recurso ao retrato local um equívoco, Machado não tira o mérito dos jovens autores, pois atribui a eles o desejo de criar uma literatura mais independente. Na realidade, a questão que o incomoda é outra. O que o autor condena é a expectativa da crítica de que esses autores tivessem como compromisso forjar ou traduzir uma identidade nacional por meio da evocação à “cor local” em suas obras:

Devo acrescentar que neste ponto manifesta-se às vezes uma opinião, que tenho por errônea: é a que só reconhece espírito nacional nas obras que tratam de assunto local. [...] Não há dúvida que uma literatura, sobretudo uma literatura nascente, deve principalmente alimentar-se dos assuntos que lhe oferece a sua região; mas não estabeleçamos doutrinas tão absolutas que a empobrecem⁴⁴.

Para Machado, “o que se deve exigir do escritor antes de tudo, é certo sentimento íntimo, que o torne homem do seu tempo e do seu país, ainda quando trate de assuntos remotos no tempo e no espaço”⁴⁵. Aqui já temos mais um aspecto do projeto estético-literário do autor sendo desenhado. A preocupação política e o empenho em estar atualizado com o mundo das ideias, independentemente de viver na periferia, mostram um Machado bastante deslocado em relação aos seus pares.

⁴³ CANDIDO, Antonio. *Iniciação à literatura brasileira*. São Paulo: Humanitas FFLCH/USP, 1999, p. 55.

⁴⁴ MACHADO DE ASSIS, op. cit., v. III, p. 803-804.

⁴⁵ MACHADO DE ASSIS, op. cit., v. III, p. 804.

Ao tratar do romance ainda de influência romântica, Machado reforça sua preocupação em ampliar os horizontes para além do local: “Esta casta de obras conserva-se aqui no puro domínio de imaginação, desinteressada dos problemas do dia e do século, alheia às crises sociais e filosóficas”⁴⁶.

A consciência de que estava na periferia da história e da cultura (e não só do capitalismo, como afirma Schwarz) despertou uma sede de liberdade que se traduziu na libertação de seu estilo da influência do Romantismo e da submissão ao gosto popular, para retratar uma realidade sem final feliz. O contexto histórico exigia uma diferenciação: aquela independência com que o autor sonhava para a literatura brasileira estava, infelizmente, atrelada às outras esferas de dependência. As ideias importadas da Europa não combinavam com nossa situação de país colonizado e escravista, como nota Roberto Schwarz em um ensaio publicado há mais de 30 anos⁴⁷.

A necessidade e o desejo de buscar uma identidade para a literatura brasileira e para si próprio como escritor fizeram Machado diversificar sua formação intelectual. Como afirma Lúcia Miguel Pereira, “Machado de Assis foi uma exceção no Brasil do século XIX e ainda o seria no Brasil do século XX”, pois “os livros que amava não eram os que nutriam os seus contemporâneos”. Shakespeare, Sterne, Dickens, Pascal e Victor Hugo são alguns exemplos de seus autores favoritos: “as leituras de que se embebeu o exaltam, pois essa escolha o alçava muito acima de seu meio”⁴⁸. Antes dela, José Veríssimo já observara que Machado, “estranho a toda a petulância da juventude, estuda, observa, medita, lê, relê os clássicos da língua e as obras-primas das principais literaturas”⁴⁹. A paixão pelos clássicos europeus delineou sua formação intelectual e se transformou em objeto de pesquisa de vários trabalhos sobre a biblioteca do autor, como os de Jean Michel Massa – publicado em 1961 e recentemente reeditado – e de José Luís Jobim.

No entanto, pelos críticos de seu tempo, essa formação não era bem-vista. Aliada às críticas, já mencionadas, que fizera ao nacionalismo romântico, a formação intelectual de Machado inspirava desconfianças e acabou por criar uma falsa polêmica que perdurou por muitos anos: era a literatura de Machado representativa da identidade nacional? Era brasileira em sentido estrito?

⁴⁶ MACHADO DE ASSIS, op. cit., v. III, p. 805.

⁴⁷ “As idéias fora de lugar”, em SCHWARZ, op. cit., p. 9-32.

⁴⁸ PEREIRA, Lúcia Miguel. Pesquisas psicológicas: Machado de Assis. In: *História da Literatura Brasileira: prosa de ficção: de 1870 a 1920*. Belo Horizonte: Itatiaia; São Paulo: Edusp, 1988, p. 63.

⁴⁹ VERÍSSIMO, op. cit., p. 213-214.

Veríssimo logo percebeu que a composição do texto machadiano era diferenciada e que o Brasil não estava representado nela de forma óbvia como nas obras românticas. Para ele, Machado era “o mais intimamente nacional de nossos romancistas, se não procurarmos o nacionalismo somente nas exterioridades pitorescas da vida ou nos traços mais notórios do indivíduo ou do meio”. E reafirma o caráter universal que transcende as amarras da “cor local”: “como o que sobretudo lhe interessa é a alma das cousas e dos homens, é ela que procura exprimir”⁵⁰.

É Candido que vai formular um trabalho mais profundo sobre a paisagem brasileira na obra machadiana. Para se ter uma ideia do quanto essa polêmica se estendeu, Candido registra que Alceu Amoroso Lima, “exprimindo, aliás, a opinião geral, podia dizer que Euclides da Cunha era ‘mais brasileiro’ do que Machado de Assis”⁵¹. Só em 1940, um artigo de Roger Bastide⁵² vai trazer argumentos estéticos que permitem ver onde está a brasilidade de Machado, ajudando a tirar o nacionalismo do centro das discussões sobre literatura no Brasil.

Antonio Candido partiu do ensaio de Bastide para escrever um de seus textos mais importantes sobre a obra machadiana. Em “Machado de Assis de outro modo”, Candido afirma:

Foi contra esta tradição gasta que e já duvidosa que Roger Bastide se manifestou, e costumava dizer que, pelo contrário, a haver opção, Machado seria o mais brasileiro dos dois, porque na sua obra o Brasil estava presente no miolo, não na aparência⁵³.

Para Candido, o texto de Bastide marca uma transição e uma ruptura, pois teria sido o primeiro ensaio a tratar a obra machadiana de modo “realmente contemporâneo”,

Pois não se refere à biografia, nem à psicologia, nem à sociedade, nem à correção da língua, mas à própria natureza do discurso, propondo explicitamente o conceito de latência e encarando a realidade exterior como matéria de construção literária. De fato Bastide mostra como o texto comporta uma carga de mundo que atua graças à organização efetuada pela composição literária, não à simples referência temática ou conceitual⁵⁴.

Bastide observa como a paisagem brasileira está entranhada na narrativa machadiana, graças a uma estratégia de composição que classifica como “presença na ausência”. Para o

⁵⁰ VERÍSSIMO, op. cit., p. 216.

⁵¹ CANDIDO, Antonio. *Formação da literatura brasileira*. 7. ed. Rio de Janeiro: Itatiaia, 1993, v. I, p. 106.

⁵² “Machado de Assis, paisagista”, publicado na Revista do Brasil, 3ª Fase, v. III, n. 29, 1940, citado por CANDIDO, Antonio. *Formação da literatura brasileira*...

⁵³ CANDIDO, Antonio. *Formação da literatura brasileira*..., p. 106.

⁵⁴ CANDIDO, Antonio. *Formação da literatura brasileira*..., p. 105.

autor, Machado “tomou posição deliberadamente contrária à visão exótica do Brasil, encarando o paisagismo dos românticos como perspectiva de fora para dentro, à maneira dos estrangeiros, que se interessam sobretudo pelo pitoresco”. Ou, melhor explicado por Candido, “Bastide procurou mostrar que em Machado de Assis a paisagem do Brasil está presente de maneira mais poderosa, porque não é enquadramento descrito, mas substância implícita da linguagem e da composição, inclusive como suporte das metáforas”⁵⁵.

Pelo conceito de transposição, oriundo da pintura, que consistiria em “revestir os indivíduos das cores e nuances da natureza que os cerca”⁵⁶, Bastide mostra como Machado imprime ao texto a atmosfera marítima, sedutora e intrigante do Rio de Janeiro através dos verdes olhos de ressaca de Capitu.

Esse artigo chamou a atenção para a composição machadiana e para o requintado casamento entre temática e estratégias narrativas. Ele abriu espaço para outras investigações complexas. Superada a polêmica sobre a brasilidade do autor, ficou mais clara a forma de abordagem adotada por ele para tratar a questão da identidade nacional em sua ficção.

Ainda sobre o tema da identidade nacional, John Gledson enfatiza a importância dos contos, especialmente *Papéis avulsos*, na expressão das ideias de Machado sobre a história brasileira:

Creio que ocupa uma posição central particularmente no que diz respeito à incorporação dessas idéias na ficção: acredito também que esta posição central tem que ver com a difícil questão da identidade nacional. [...] Não foi então por acaso que Machado recorreu ao conto, tão adequado à dramatização de crises de identidade, para uma primeira solução do ‘problema’ – isto é, a sua primeira tentativa de encarnar a nação num único personagem⁵⁷.

“O Espelho” é o exemplo mais forte desse procedimento. A ideia de duas almas – uma exterior e uma interior; uma com farda, uma sem – personifica a contradição entre ideologias liberais, influenciadas pelas transformações políticas na Europa, e a tradição patrimonialista local. Não é à toa que a crise de identidade de Jacobina se aprofunda no momento em que os escravos fogem da fazenda. Ele tenta redescobrir sua identidade num espelho que teria sido comprado, por seus antepassados, das fidalgas que integravam a corte de D. João VI quando

⁵⁵ CANDIDO, Antonio. *Formação da literatura brasileira...*, p. 109.

⁵⁶ BASTIDE apud CANDIDO, Antonio. *Formação da literatura brasileira...*, p. 106.

⁵⁷ GLEDSON, John. A História do Brasil em Papéis Avulsos de Machado de Assis. In: CHALOUB, Sidney; PEREIRA, Leonardo Affonso de M. (Org.). *A história contada: capítulos de história social da literatura no Brasil*. Rio de Janeiro: Nova Fronteira, 1998. p. 15-34, p. 27.

da fuga para o Brasil, em 1808. “Não sei o que havia disso de verdade; era tradição”⁵⁸. No momento em que a nova ordem se coloca, a personagem não mais consegue ver a si própria no velho espelho sem a farda, símbolo da antiga ordem e da estabilidade que Jacobina não tem mais.

1.2.2 A crítica ao Naturalismo

Se, quanto à temática, a identidade nacional dominava o cenário da crítica, quanto à forma, o caráter descritivo de algumas obras também já está desgastado, como nota Machado:

Há boas páginas, como digo, e creio até que um grande amor a este recurso da descrição, excelente, sem dúvida, mas (como dizem os mestres) de mediano efeito, se não avultam no escritor outras qualidades essenciais. Pelo que respeita à análise de paixões e caracteres são muito menos comuns os exemplos que podem satisfazer à crítica; alguns há, porém, de merecimento incontestável. Esta é, na verdade, uma das partes mais difíceis do romance, e ao mesmo tempo das mais superiores. Naturalmente exige da parte do escritor dotes não vulgares de observação, que, ainda em literaturas mais adiantadas, não andam a rodo nem são a partilha do maior número⁵⁹.

Podemos, a partir dessas considerações do autor, antever as transformações que gradualmente se impõem à sua obra, pois ainda estamos em 1873: o primeiro romance, *Ressurreição*, fora publicado há pouco e estamos no ano da publicação do segundo volume de contos, *Histórias da meia-noite*. A atividade crítica ajuda o autor a delinear de forma cada vez mais precisa o projeto estético-literário que pretende realizar.

Em texto de 1879, o tema das escolas literárias está na ordem do dia. Ao falar da nova geração de escritores, Machado questiona se isso significaria também o surgimento de uma nova estética. Estamos em plena fase de transição entre a decadência do Romantismo e a ascensão do que Machado chama de Realismo e a que chamaremos Naturalismo. Não renegar a fonte em que bebeu e não se deslumbrar com o novo é mais um princípio que o autor deixa claro como parte de seu credo de escritor: “nem tudo tinham os antigos, nem tudo têm os modernos; com os haveres de uns e outros é que se enriquece o pecúlio comum”, proclama⁶⁰.

⁵⁸ MACHADO DE ASSIS, op. cit., v. II, p. 347.

⁵⁹ MACHADO DE ASSIS, op. cit., v. III, p. 805.

⁶⁰ MACHADO DE ASSIS, op. cit., v. III, p. 810.

Mas os excessos do Naturalismo provocam Machado a se posicionar novamente. É na análise de *O primo Basílio* que o autor radicaliza sua crítica à forma explícita da nova escola. A boa recepção de *O crime do Padre Amaro* Machado atribui à força da novidade:

Era realismo implacável, conseqüente, lógico, levado à puerilidade e à obscuridade [...]. Não se conhecia no nosso idioma aquela reprodução fotográfica e servil das coisas mínimas e ignóbeis [...]. Se, por vezes, o Sr. Eça de Queirós esquecia por minutos as preocupações da escola; e, ainda nos quadros que lhe destoavam, achou mais de um rasgo feliz, mais de uma expressão verdadeira; a maioria, porém, atirou-se ao inventário⁶¹.

Quanto ao estilo de *O primo Basílio*, Machado considera-o simplesmente pura imitação de Zola e lamenta a instauração de uma estética “que não esquece nada, e não oculta nada [...]. Porque a nova poética é isto, e só chegará à perfeição no dia em que nos disser o número exato dos fios de que se compõe um lenço de cambraia ou um esfregão de cozinha”⁶². É interessante como a posição de Machado quanto à descrição e ao Naturalismo antecipa as ideias expressas por Lukács em seu ensaio *Narrar ou descrever?*, de 1936. Até o uso do termo “inventário” para caracterizar o estilo naturalista coincide em ambas as traduções.

O texto de Lukács trata das diferenças de estilo entre o modo de contar de Zola e o de Tolstoi. O autor usou um critério temático para escolher os textos a serem analisados: tanto em *Naná* quanto em *Ana Karenina*, há trechos em que acontece uma corrida de cavalos. Sobre o modo de contar essa passagem em ambas as obras, é que Lukács constrói sua análise:

Em Zola, a corrida é descrita do ponto de vista do espectador; em Tolstoi, é narrada do ponto de vista do participante [...]
 A descrição de Zola é uma pequena monografia sobre a moderna corrida de trote, que vem acompanhada em todas as suas fases, desde a preparação dos cavalos até a passagem pela linha de chegada com a mesma insistência [...]
 A corrida de cavalos de *Ana Karenina* é o ponto crucial de um grande drama. [...] Todas as relações entre os principais personagens do romance entram numa fase decididamente nova, após a corrida. Esta, por conseguinte, não é um ‘quadro’ e sim uma série de cenas altamente dramáticas, que assinala uma profunda mudança no entrecho. [...]
 E será que é o caráter completo de uma descrição que torna alguma coisa artisticamente necessária? Ou não será, antes, a relação necessária dos personagens com as coisas e com os acontecimentos – nos quais se realiza o destino deles, e através dos quais eles atuam e se debatem?⁶³

⁶¹ MACHADO DE ASSIS, op. cit., v. III, p. 904.

⁶² MACHADO DE ASSIS, op. cit., v. III, p. 904.

⁶³ LUKÁCS, Georg. Ensaios de literatura. Rio de Janeiro: Civilização Brasileira, 1968, p. 48-50.

A pergunta de Lukács já revela sua posição quanto ao tipo de descrição usada por Zola, à qual atribui o “caráter de inventário”.

O contraste entre participar e o observar não é casual, pois deriva da posição de princípio assumida pelo escritor em face da vida, em face dos grandes problemas da sociedade, e não do mero emprego de um diverso método de representar determinado conteúdo ou parte de conteúdo [...].

Todo novo estilo surge como uma necessidade histórico-social da vida e é um produto necessário da evolução social [...]. A alternativa *participar ou observar* corresponde, então, a duas posições socialmente necessárias, assumidas pelos escritores em dois sucessivos períodos do capitalismo. A alternativa *narrar ou descrever* corresponde aos dois métodos fundamentais de representação próprios destes períodos⁶⁴.

A migração da observação para a participação que a transição do descrever ao narrar implica, de acordo com Lukács, revela-se gradualmente na obra de Machado mediante o fortalecimento da figura do narrador. À medida que este ganha espaço, o texto também conquista autonomia e concisão. A ambiguidade, como um espaço reservado à inteligência e à imaginação do leitor, aparece como uma das respostas aos excessos das descrições naturalistas. Voltaremos a esse assunto mais adiante, quando abordarmos o conto.

Para Luís Augusto Fischer⁶⁵, foi a consciência da crise de representação que levou Machado a buscar uma estética própria, singular, uma vez que a forma romântica já esgotada não dava conta da complexidade da realidade e o Naturalismo caía no extremo oposto, sacrificando a “realidade estética”, como Machado já apontava no *Instinto de nacionalidade*.

Podemos, então, principalmente a partir do discurso crítico de Machado, concluir que, desde antes de começar sua produção em prosa de ficção, o autor já cultivava uma reflexão sobre a literatura brasileira e sobre o papel dos novos autores. Com a atividade crítica aliada à obra de ficção, Machado foi configurando com mais nitidez os pressupostos que embasariam seu projeto estético-literário. Por um lado, este apontava para filiar-se a uma tradição literária – sem território e época definidos – e, por outro, superá-la, pela subversão formal ou temática de suas regras, sem, no entanto, deixar de tratar das idiosincrasias do Brasil. Um dos autores a destacar com mais clareza a presença desse projeto na produção machadiana é Enylton de Sá Rego. Em seu estudo sobre a influência da sátira menipeia⁶⁶ – ou tradição luciânica, como prefere –, o autor demonstra como as várias características da obra machadiana (da fase

⁶⁴ LUKÁCS, op. cit., p. 54 e 57. (Grifos do autor).

⁶⁵ FISCHER, Luís Augusto. Contos de Machado: da ética à estética. In: SECCHIN, Antonio Carlos et al. *Machado de Assis: uma revisão*. Rio de Janeiro: In-Fólio, 1998. p. 147-165.

⁶⁶ SÁ REGO, Enylton José de. *O calundu e a panacéia: Machado de Assis, a sátira menipeia e a tradição luciânica*. Rio de Janeiro: Forense Universitária, 1989.

madura) estavam presentes em seu discurso crítico desde o início. Sá Rego reitera a qualificação de Machado como um escritor consciente, feita por Candido.

Quais seriam, então, as linhas mestras, os princípios norteadores do projeto estético-literário machadiano? Listaremos a seguir algumas delas, a partir do que discutimos até este ponto:

- a) unir gêneros literários “altos” e “baixos”; usar sistematicamente a paródia e a liberdade de imaginação (não se limitando às exigências da verossimilhança); estatuto ambíguo e não-moralizante, com a coexistência do trágico e do cômico (sem a preponderância de um sobre o outro); e adotar ponto de vista distanciado “como um espectador desapaixonado que analisa não só o mundo a que se refere, como também a sua própria obra literária, a sua própria visão de mundo”, todas características apontadas por Sá Rego como parte da adoção da tradição luciânica:⁶⁷
- b) buscar a independência e a emancipação literária, sem a obrigação de se fixar na cor local;
- c) estar atualizado sobre “os problemas do dia e do século, as crises filosóficas e sociais”;
- d) não conjugar “o ideal poético e o ideal político”, fazendo de ambos “um só intuito”⁶⁸;
- e) conhecer os clássicos: não desprezar os movimentos literários passados, nem aderir totalmente aos novos; manter a postura crítica; não desprezar as contribuições que ambos podem dar ao “pecúlio comum”⁶⁹, seguindo princípio dialético entre tradição e inovação⁷⁰;
- f) produzir e incentivar a produção dos gêneros ainda inexistentes na literatura brasileira de então; e
- g) evitar dizer tudo: deixar espaço para a imaginação do leitor.

Para atender a esses princípios, Machado teve que afastar, cada vez mais, seu modo de escrever do ideal apregoado pelos críticos brasileiros de então. Para seu principal crítico, Sílvio Romero – incapaz de entender a forma de escrever do autor como uma opção –, sempre na expectativa de um estilo regular que satisfizesse seus ideais, só restava acreditar que Machado não dispunha “profusamente do vocabulário e da frase” e atribuir isso a sua “índole

⁶⁷ SÁ REGO, op. cit., p. 45.

⁶⁸ MACHADO DE ASSIS, op. cit., p. 813.

⁶⁹ MACHADO DE ASSIS, op. cit., p. 80.

⁷⁰ SÁ REGO, op. cit., p. 113.

psicológica indecisa”, ou, pior, a uma “lacuna nos órgãos da palavra”: “vê-se que ele apalpa e tropeça [...]. Realmente, Machado de Assis repisa, repete, torce, retorce tanto suas ideias e as palavras que as vestem, que nos deixa a impressão de um perpétuo tartamudear”⁷¹.

O caráter fragmentário adotado por Machado também faz parte de seu projeto e antecipa uma marca de modernidade que estaria presente no estilo de grandes autores do século XX. Conhecendo bem a crítica de sua época, Machado insinua sua técnica, nas palavras de Brás Cubas, continuando o jogo de enganar seus contemporâneos:

Tu tens pressa de envelhecer, e o livro anda devagar; tu amas a narração direita e nutrida, o estilo regular e fluente, e este livro e o meu estilo são como os ébrios, guinam à direita e à esquerda, andam e param, resmungam, urram, gargalham, ameaçam o céu, escorregam e caem...

A posição do conto dentro do projeto literário machadiano é o que veremos a seguir.

1.3 O Conto e o Projeto Machadiano

*É gênero difícil, a despeito da sua aparente facilidade, e creio que essa mesma aparência lhe faz mal, afastando-se dele os escritores, e não lhe dando, penso eu, o público toda a atenção de que ele é muitas vezes credor*⁷².

Quem observa os títulos dos volumes de contos publicados por Machado de Assis tem a falsa impressão de que o autor considerava conto um gênero menor: *Papéis avulsos*, *Histórias sem data*, *Páginas recolhidas*, *Várias histórias*. A simplicidade remete à ideia de dispersão, de uma produção pouco sistemática. Nada mais enganador, bem ao gosto de Machado. Tão perfeccionista era que demorou mais de dez anos de exercício do conto para publicar o primeiro volume e, dos mais de vinte que já havia escrito, selecionou apenas sete.

Esse rigor, que já caracterizava o estilo do jovem autor, perpassaria toda sua produção no gênero. Dos 205 contos produzidos, Machado optou por publicar apenas 76 em livro, geralmente precedidos por uma advertência ou prefácio que insistia em afirmar a despreensão daquelas páginas, sem, no entanto, jamais atribuí-la ao gênero. E ainda brincava:

⁷¹ ROMERO, op. cit., p. 55.

⁷² Sobre o conto. MACHADO DE ASSIS, op. cit., p. 806.

O tamanho não é o que faz mal a este gênero de histórias, é naturalmente a qualidade; mas há sempre uma qualidade nos contos, que os torna superiores aos grandes romances, se uns e outros são medíocres: é serem curtos⁷³.

A obra de Machado de Assis, escritor brasileiro mais estudado pelos acadêmicos brasileiros, é objeto de inúmeros ensaios, artigos, dissertações, teses e livros. Porém, a maior parte da fortuna crítica privilegia o estudo de seus romances. O conto, gênero em que foi pioneiro e do qual Machado ainda é nosso maior representante, foi pouco estudado. Alguns autores observam com igual estranheza a escassez de trabalhos na área, como Paul Dixon:

Os contos de Machado de Assis têm sido muito elogiados, mas pouco estudados [...] a análise dos relatos não passa de artigos avulsos, e algumas introduções a antologias. Até agora, nenhum livro de crítica literária se dedicou preferencialmente aos contos [...]. Não é fácil entender a falta de um livro analítico sobre os contos⁷⁴.

A opinião é corroborada por Patrícia Lessa Flores da Cunha, autora de obra inteiramente dedicada aos contos de Machado:

[...] a análise de seu romance tem sido feita à exaustão, oferecendo notáveis e reconhecidos enfoques à compreensão de sua obra. O mesmo geralmente não sucede com o estudo do conto, cuja interpretação tem quase sempre funcionado como elemento subsidiário, nem por isso menos valioso, ao alcance proposto àquele conjunto de escritura do autor – o que, de certa forma, estimula a natureza da indagação⁷⁵.

Certamente essa escassez não tem relação com o mérito, já que, no gênero, a crítica é quase unânime em compará-lo aos maiores contistas da história⁷⁶. Intelectuais como Antonio Candido e Lúcia Miguel-Pereira destacaram o conto machadiano como grande expressão de nossa literatura, além de seu importante papel como parte do processo de crescimento e amadurecimento do autor. Luís Augusto Fischer enfatiza o fato de que a contística machadiana

é uma obra vasta, que sozinha justificaria a perenidade de qualquer autor: se mais não houvesse feito, seria já um clássico da língua portuguesa e da literatura ocidental. O mesmo ninguém diria, creio, de sua poesia, nem de sua crônica, nem de sua crítica, nem de seu teatro, nem de sua atividade de

⁷³ MACHADO DE ASSIS, op. cit., p. 476 (advertência a *Várias histórias*).

⁷⁴ DIXON, op. cit., p. 10.

⁷⁵ CUNHA, op. cit., p. 17.

⁷⁶ “Ninguém nega a qualidade de Machado como contista, um dos melhores da história da literatura brasileira, digno de comparação, em muitos momentos, aos maiores contistas de sua época - Maupassant, Tchekhov ou Henry James”. GLEDSON, op. cit., p. 35.

tradutor – só de seu romance, o que seria motivo suficiente, talvez, para tomar o autor como sendo essencialmente um prosador, um autor de narrativas⁷⁷.

Essa argumentação reafirma a posição de Lúcia Miguel Pereira sobre a importância do gênero para o artista Machado de Assis:

nos romances, mesmo nos melhores, as delongas, as intromissões do autor dão à narrativa um aspecto indeciso e ziguezagueante, que tem por vezes grande encanto, mas é em outras um tanto maçante. No conto, não. Obrigada a encolher-se, a trama ganha em coesão, em resistência. Tecnicamente, literariamente, algumas de suas histórias são verdadeiras obras-primas⁷⁸.

1.3.1 O conto como laboratório do romance

Além de ser proporcionalmente o gênero em que a produção machadiana atinge maior regularidade em termos de qualidade literária, o conto teve um papel importante em sua formação de escritor de prosa de ficção. É ele que “lhe dará a oportunidade de explorar outros ângulos e categorias importantes nesta sua renovação da arte literária”⁷⁹. O conto foi uma espécie de laboratório escolhido por Machado (como o foi por outros grandes escritores) para a experimentação e o exercício de padrões estilísticos que, mais tarde, seriam incorporados a sua identidade literária. Os experimentos bem-sucedidos dos contos acabaram por compor a narrativa ficcional de maior fôlego. (O inverso poderia ter ocorrido, embora seja menos comum um autor experimentar uma técnica ou um recurso qualquer no romance para depois aplicá-lo nas narrativas curtas.) Mesmo depois do sucesso alcançado como romancista, Machado continuou a escrever e a publicar contos com muito mais assiduidade do que o fez com as narrativas longas.

Em termos de evolução intelectual do seu autor, *Papéis avulsos* (1882) é sem dúvida a mais importante das coleções de contos de Machado de Assis. Há, obviamente, uma relação crucial com *Memórias póstumas de Brás Cubas*, publicado em livro no ano anterior e, de fato, seria fácil estabelecer um

⁷⁷ FISCHER, op. cit., p. 149.

⁷⁸ PEREIRA, Lúcia Miguel. *Machado de Assis: estudo crítico e biográfico*. 6. ed. Belo Horizonte: Itatiaia; São Paulo: Edusp, 1988, p. 225-226.

⁷⁹ BRAYNER, Sonia. *Labirinto do espaço romanesco: tradição e renovação da literatura brasileira (1880-1920)*. Rio de Janeiro: Civilização Brasileira; Brasília: INL, 1979, p. 65.

paralelismo semelhante entre os últimos contos (como *Missa do Galo*) e a ironia complexa e sutileza psicológica de *Dom Casmurro*⁸⁰.

A evolução do contista se evidencia nos textos mais curtos, em que a escolha da palavra precisa denota a segurança do escritor que já sabe aonde quer chegar e a melhor maneira de fazê-lo. Nas longas descrições dos primeiros contos dos volumes *Histórias da meia-noite* e *Contos fluminenses*, verificamos um autor ainda indeciso quanto à história a ser contada e a que postura adotar, se mais participativa ou mais neutra quanto ao narrador. Foi a partir de *Papéis avulsos* (1882) que se encontraram algumas soluções, resultando num “processo fascinante (que) é parte de uma dialética complexa de experimentação e descoberta”⁸¹, como diz Gledson.

Candido trata tanto do processo de desenvolvimento do autor por meio do gênero quanto de sua variedade:

Nele se manifesta o amor da ficção pela ficção, a perícia em tecer histórias, que se aproxima da gratuidade determinativa do jogo. Deste autor habilidoso e divertido brota o Machado de Assis focalizado aqui, – numa passagem insensível, que vai levando da quase-melancolia de *Noite de almirante* à dubiedade de *D. Paula*, daí à indecisão perturbadora de *Dona Benedita*, que sobe à surpresa contundente de *A senhora do Galvão*, já no portal de um mundo estranho, – mostrando as transições quase imperceptíveis que unificam a diversidade do escritor⁸².

Lúcia Miguel Pereira destaca que o autor “custou muito a firmar-se como contista; entre 1860 e 1870, quando já é destro em crônicas, no conto ainda é fraco e indeciso”. Mas depois de *Papéis avulsos*, revelou-se um mestre no gênero. A autora compara o romance à vida, e o conto à anedota, e busca na personalidade do Machado “contido, medido e comedido” uma das razões para seu sucesso na narrativa curta. Para ela, “a própria natureza do gênero exige uma certa limitação, uma tendência a ver de perto, à moda dos míopes. O episódio, para ter realce, requer os vidros de aumento da análise minuciosa, que no romance perturbam a visão do conjunto”.

O contista Machado aborda com a mesma desenvoltura as vaidades, os sonhos, as frustrações, os crimes e as anomalias de seus personagens. Para Pereira, sob o ponto de vista temático, o autor

⁸⁰ GLEDSON, op. cit., p. 15.

⁸¹ GLEDSON, op. cit., p. 28.

⁸² CANDIDO, Antonio. *Vários escritos*. São Paulo: Duas Cidades, 1970, p. 32.

mostrou como as condições especiais da sociedade que aqui se formou no Império repercutiram sobre os elementos constitutivos da personalidade. Nas suas matronas e damas elegantes, nos seus homens ambiciosos, libertinos de corpo ou de espírito, nos seus agregados e parasitas de casas ricas [...] ⁸³.

Essas observações, por si, já nos remetem aos perfis *Uma senhora*, *A cartomante*, *O enfermeiro*; às teorias filosóficas, *O espelho*, *A igreja do diabo*, *O alienista*; ou, ainda, aos mistérios de *Missa do Galo* e *Uns braços*.

1.3.2 O conto como parte do projeto estético

O conto machadiano é um lugar privilegiado para estudarmos seu processo de amadurecimento, como bem aponta Cunha sobre o primeiro trabalho do gênero, escrito aos 19 anos:

É interessante já se poder observar, nessa pequena narrativa, a presença embrionária dos motivos perenes que sempre norteariam a escritura dos seus contos, mesmo na linguagem mais simples e singela do escritor que então apenas se formava: a presença da dúvida, a descoberta da traição, levando à constatação inevitável da dubiedade que ronda os procedimentos humanos ⁸⁴.

A autora considera que essa recorrência de temas e posturas fica bastante evidenciada no exame dos contos, mas que “é possível, senão preferível, explicá-la mais em termos de uma evolução natural e necessária do que propriamente como uma ruptura de modo de pensar”. Há, assim, uma ligação necessária, e não casual, dessa produção a “uma opção por um programa que visava ao estabelecimento dos fundamentos de uma literatura criativa, criadora e autônoma”:

[...] existiria, com efeito, um notável sentido de coerência e afirmação em toda a obra machadiana, da qual o conto não faz exceção; ao contrário, sendo um segmento particularmente coeso e bem delimitado, evidenciaria, na sua persistente e fecunda expressão, todos os matizes desse reiterado empenho ⁸⁵.

Esse é um dos motivos que nos levam a ver o conto como o melhor material para observarmos as mudanças estilísticas decorrentes do amadurecimento de Machado como

⁸³ PEREIRA, op. cit., p. 75.

⁸⁴ CUNHA, op. cit., p. 53-54.

⁸⁵ CUNHA, op. cit., p. 54.

prosador. Além da escassez de trabalhos totalmente dedicados ao conto, outro ponto a ser discutido é relativo à natureza destes. Fischer faz uma crítica bastante pertinente, afirmando “o caráter mais interpretativo do que analítico” da fortuna sobre o conto: “Quanto às constantes estruturais dos contos, porém, a tradição é sensivelmente mais pobre do que quanto às constantes temáticas”. O autor toma essa discrepância entre “a margem das interpretações e das análises”, sendo a primeira “superpovoada” e a segunda “rarefeita”, para investigar as suas causas:

Por que razões a fortuna crítica do conto machadiano seria pródiga em estudos voltados à interpretação de aspectos filosóficos, ideológicos, sociológicos, numa palavra conteudísticos dos contos, e simultaneamente avara nas considerações das estruturas, dos procedimentos, numa palavra da forma dos contos?⁸⁶

Uma das respostas possíveis, para Fischer, seria “a afeição da tradição letrada brasileira pela consideração do sentido moral, da dimensão edificante da literatura em relação à vida”. Isso se justifica em função “menos de idiosincrasia e mais de convergência entre ficção e realidade, entre literatura e história, entre arte e sociedade”, constata o autor, referindo-se à longa busca por afirmação de uma identidade nacional (não só na época de Machado, mas também posteriormente). O fato de a própria literatura ter estado, por longo período, “a serviço de uma causa extraliterária” faz com que autor considere que “nada mais razoável do que também a crítica ter-se posto à mesma tarefa”⁸⁷.

A afirmação de Fischer nos leva não só a constatar o problema da escassez de estudos mais analíticos, mas também às consequências disso. Quantas “verdades” mantiveram-se graças a essa omissão? Por que um dos pilares da crítica machadiana, como a divisão da obra de Machado em fases estanques – afirmação cuja origem já se perdeu no tempo, pois Romero, em fins do século XIX, já a discutia –, se mantém firme se isso hoje parece tão ilógico?

Um dos medalhões da crítica que ainda insiste nisso é Alfredo Bosi. Como bem coloca Fischer, além de repisar a existência de dois Machados, Bosi “estabelece distinção de mérito entre os primeiros contos e aqueles que foram aparecendo a partir dos meados de 1870”⁸⁸.

Assim, por qualquer dos lados – o da tradição da literatura e da crítica brasileiras de se fixar mais no conteúdo do que na forma, e o da própria

⁸⁶ FISCHER, op. cit., p. 149-150.

⁸⁷ FISCHER, op. cit., p. 150.

⁸⁸ FISCHER, op. cit., p. 15.

matéria literária machadiana, que salienta o detalhe e tolhe a visada de conjunto – é compreensível que não tenhamos atentado muito para a estrutura, em nosso caso para a estrutura dos contos. Salvo exceções poucas, estamos, portanto, quanto aos contos numa situação mais moral do que científica, que toma Machado como um comentador da vida, não como um autor de alta literatura que manejou a forma do conto soberanamente, conferindo a ela um caráter específico, de grande rendimento literário e de largo alcance na representação da vida brasileira⁸⁹.

Como acreditamos na importância do tipo de estudo que Fischer classifica como analítico, é com esse enfoque que pretendemos interpretar a contística de Machado de Assis.

⁸⁹ FISCHER, op. cit., p. 150.

CAPÍTULO 2 O DESAFIO TEÓRICO-METODOLÓGICO DOS ESTUDOS LITERÁRIOS AUXILIADOS POR COMPUTADOR

2.1 Um Novo Campo

A especificidade do campo dos estudos auxiliados por computador em Ciências Humanas está intimamente ligada à maneira de configurar o objeto sobre o qual se debruça – em nosso caso, o texto literário – e de obter e analisar dados sobre ele. Com essa finalidade, são desenvolvidos programas especialmente para esse tipo de pesquisa.

As investigações que aliam o uso do computador a métodos estatísticos de análise de dados para o estudo de textos têm quase meio século em países como França, Inglaterra e Estados Unidos. Há uma extensa bibliografia, há um número significativo de programas desenvolvidos e disponíveis no mercado, além dos inúmeros estudiosos que fazem parte dessa comunidade acadêmica.

No Brasil trata-se de um campo incipiente: praticamente não há bibliografia em língua portuguesa⁹⁰, e mesmo a produção do exterior não está disponível nas livrarias do país. Em função disso, apresentamos a seguir um panorama do estado da arte nesse campo no mundo e, a partir daí, um balanço histórico desses estudos e uma relação detalhada das principais obras em duas tradições – de língua inglesa e de língua francesa.

2.1.1 Características

Michel Bernard⁹¹ lista uma série de características dos trabalhos dessa natureza. Uma das principais é a interdisciplinaridade. Esse é o tipo de trabalho que envolve profissionais de várias áreas. No caso dos programas de análise textual (que geralmente são focados na estatística textual), isso acontece desde o primeiro momento, pois, para a concepção e o desenvolvimento da ferramenta, são necessários profissionais da computação e da estatística,

⁹⁰ Os títulos relacionados à área tratam de linguística, não de literatura.

⁹¹ BERNARD, Michel. *Introduction aux études littéraires assistées par ordinateur*. Paris: Presses Universitaires de France, 1999.

bem como da linguística e mesmo da literatura⁹². Além disso, há profissionais especializados para a assistência técnica, assim como para a manutenção dos equipamentos e dos programas. O envolvimento de máquinas sofisticadas e de uma série multidisciplinar de pesquisadores faz com que esse tipo de pesquisa se caracterize também por seu alto custo.

Como a ferramenta de estatística textual é aplicável a textos de qualquer natureza, pesquisadores de várias áreas – como Psicologia, Sociologia, Direito, História, Publicidade, Jornalismo, além da Linguística e da Literatura – acabam por se encontrar, mesmo que por computador, a fim de trocar informações e experiências. Bernard destaca o trabalho em equipe como uma marca dessa nova comunidade científica, unida pela tecnologia e pelas possibilidades abertas pela rapidez e precisão no levantamento de dados proporcionados por ela. O fato de muitos artigos dessa área serem publicados sob autoria coletiva é o exemplo citado por Bernard – será o fim da solidão do pesquisador?

Foi na área da Linguística que se originaram os estudos que unem análise de texto, estatística e, mais recentemente, informática. As possibilidades geradas pela confluência dessas áreas de conhecimento estimularam seus profissionais e pesquisadores a desenvolver conjuntamente tecnologias para levantamento, tratamento e análise de dados textuais – aplicáveis a qualquer texto, independentemente de sua natureza (jurídico, político, jornalístico, etc.).

A natureza interdisciplinar faz com que a estatística textual acabe por ser útil a muitos campos do conhecimento. Salem e Lebart⁹³ dividem essas áreas de aplicação, por tipos de abordagem do texto, em três grupos principais. Um deles é a análise de conteúdo, muito usada nas pesquisas na área de Comunicação, mais especificamente no Jornalismo. Os pioneiros foram B. Berenson e P. F. Lazarsfeld, com o trabalho *The analysis of communications content* (Universidades de Chicago e Nova Iorque), publicado em 1948⁹⁴. Outro é o da inteligência artificial, área mais ligada à indústria, que desenvolve programas de tradução e comandos de voz para facilitar a interface entre usuário e máquina, programas especiais para portadores de deficiência, entre outros produtos. E um terceiro é a Linguística, que engloba lexicologia, lexicometria, morfologia, filologia, etc.

Hoje um dos principais produtos da estatística textual são os bancos de dados textuais: grandes compilações de textos em suporte eletrônico que servem tanto a trabalhos acadêmicos

⁹² É o caso do Hyperbase. O professor Etienne Brunet, que criou e desenvolveu o programa, tem formação na área literária, onde atua, mas é também estatístico e programador autodidata.

⁹³ LEBART, Ludovic; SALEM, Andre. *Statistique textuelle*. Paris: Dunod, 1994.

⁹⁴ LEBART; SALEM, op. cit.

quanto à democratização do acesso do público à literatura e a textos de outra natureza (jurídico, político, etc.).

Quanto aos estudos literários e de linguagem, particularmente, Raymond Siemens divide a aplicação da computação em dois grupos, com base nos resultados: aqueles nos quais o computador é usado para produzir, por meio da manipulação do texto, apoio convencional para pesquisas futuras (dicionários, concordâncias, etc.); e aqueles nos quais o computador é usado para a análise de trabalhos específicos de literatura (análise temática, estudos estilísticos)⁹⁵. Há vasta bibliografia nessa área, principalmente em inglês e em francês, ainda não disponível no Brasil.

2.1.2 Publicações e fontes

No exterior, as publicações mais importantes em inglês são duas revistas especializadas que tiveram origem em associações para o estudo da aplicação da computação (em sentido amplo) na área de Humanas: uma é a *Computers and the Humanities*⁹⁶ (CTH), que acaba de mudar de nome para *Language Resources and Evaluation*, publicação internacional da Association for Computers and the Humanities (ACH), editada na Holanda pela Springerlink. Seu primeiro número foi publicado em 1966 e hoje tem quatro edições por ano, com a opção de versão impressa ou eletrônica. Outra é a *Literary and Linguistic Computing*⁹⁷ (LLC), editada pela Oxford University Press, uma publicação da Association for Literary and Linguistic Computing (ALLC), fundada em 1973. Seu primeiro número saiu em novembro de 1986. Também publica textos de autores do mundo inteiro e tem quatro edições anuais.

As duas associações mantêm uma parceria que fica explícita não só no sítio da LLC como também na promoção de encontros internacionais. O primeiro encontro promovido pela ALLC foi em 1970, em Cambridge, e depois anualmente até 1988, quando firmou uma parceria com a ACH para atuarem em conjunto. O primeiro evento conjunto foi realizado na Universidade de Toronto, no Canadá, em 1989. Desde então as conferências anuais se alternam entre Europa e Estados Unidos.

⁹⁵ SIEMENS, Raymond G. A new computer-assisted literary criticism? *Computers and the Humanities*, n. 36, p. 259-267, 2002, p. 259.

⁹⁶ [http://www.springerlink.com/\(03wgzpidh5yqtlekakscvn4\)/app/home/journal.asp](http://www.springerlink.com/(03wgzpidh5yqtlekakscvn4)/app/home/journal.asp).

⁹⁷ <http://llc.oxfordjournals.org/>.

Na França, também há encontros periódicos a cada dois anos da comunidade acadêmica que trabalha na área. São as Journées d'Analyse Statistique des Données Textuelles (JADT), que tiveram sua primeira edição em Barcelona, na Espanha, em 1992, organizada pelo mesmo grupo de pesquisadores de diversas universidades francesas que editam a revista especializada *Lexicometrica*. Outros periódicos especializados, em francês, são as revistas *Corpus* e *Loxias*, ambas produzidas por pesquisadores da Universidade Sophia Antipolis, de Nice.

Além das revistas, há ainda as listas de discussão especializadas que, com a internet, tornaram-se mais uma fonte de informação para o pesquisador.

2.1.3 No Brasil

As áreas de Letras e Linguística já se apropriaram da tecnologia dos computadores e da internet de variadas formas como ferramentas para seu trabalho. No Brasil, temos uma comunidade científica nessas áreas que vem, há algum tempo, discutindo o papel das novas tecnologias e as mudanças trazidas por elas no âmbito do ensino da literatura, na divulgação de textos literários via rede mundial de computadores (internet), bem como seu papel na produção literária feita em meio digital e pensada para ser consumida nesse tipo de suporte.

Entre as áreas que se destacam no Brasil nesse campo, encontram-se principalmente bases de dados com digitalização de textos (transformação de textos impressos em arquivos eletrônicos), bibliotecas digitais, bancos de dados sobre literatura⁹⁸, edições eletrônicas, produção literária em computador (principalmente para autores iniciantes que não conseguem editoras dispostas a publicar seus escritos). Porém, o uso de programas de computador para auxiliar num trabalho de análise textual de obra literária ainda é praticamente inédito no país.

Na UFSC, há projetos que estão sendo desenvolvidos junto ao Nupill, que, além da parceria com o Centre de Recherche Hubert de Phalèse, da Universidade de Paris III (Sorbonne Nouvelle), trouxe os professores Etienne Brunet (criador do Hyperbase) e Carlos Maciel (criador da base de dados Portext), ambos da Universidade de Nice (Sophia Antipolis), para palestras e cursos sobre o estudo da literatura com as novas tecnologias. Ainda na UFSC,

⁹⁸ Três atividades desenvolvidas pelo Núcleo de Pesquisa em Informática, Linguística e Literatura (Nupill), na UFSC.

há trabalhos sendo desenvolvidos na área da Linguística de *Corpus* que também usam essas tecnologias.

2.2 Histórico

É difícil descobrir quem foi o primeiro estudioso que tentou quantificar os elementos que compõem um texto, literário ou não. Seja por um impulso científico ou apenas por curiosidade, a iniciativa de contar, listar e ordenar elementos é uma das formas mais simples de sistematizar informação sobre determinado objeto. Talvez esse seja o motivo de, na bibliografia especializada, não haver uma data ou nome exato relativo ao primeiro trabalho dessa natureza. Como as maiores fontes de informação sobre os estudos em estatística textual estão em inglês e em francês, e há algumas divergências quanto às origens desses estudos, decidimos dividir nosso histórico em duas versões.

2.2.1 A vertente de língua inglesa

Segundo Anthony Kenny⁹⁹ e Susan Hockey¹⁰⁰, os estudos quantitativos têm sua origem em 1851, quando o professor de matemática da universidade de Londres Augustus de Morgan envia uma carta ao reverendo W. Heald, em Cambridge, propondo a mensuração do tamanho das palavras das epístolas de S. Paulo, a fim de resolver as dúvidas de autoria que pairavam sobre algumas delas. Segundo Hockey, De Morgan argumentava que as palavras usadas na Epístola aos Hebreus pareciam ser mais longas do que nas demais cartas de autoria de Paulo.

Apesar da menção à carta de Morgan, o consenso entre os autores da área é o de que foi o norte-americano T. C. Mendenhall quem teria de fato levado a cabo o primeiro estudo do gênero em 1887. No entanto, Kenny e Hockey contam a história de que Mendenhall houvera adquirido uma cópia da carta escrita por Morgan por volta de 1880. Se foi a carta que lhe deu a ideia de fazer seu estudo não é possível afirmar, mas em 1887 ele publica o artigo “*The*

⁹⁹ KENNY, Anthony. *The computation of style*. Oxford: Pergamon Press, 1982.

¹⁰⁰ HOCKEY, Susan. *Electronic texts in the humanities*. London; New York: Oxford University Press, 2004.

characteristic curves of composition”, em que testa a hipótese de que a extensão das palavras pudesse ser uma característica distintiva de autoria.

Mendenhall¹⁰¹ estudou obras de diversos autores, como Dickens, Thackeray, Shakespeare, Bacon e Marlowe, além do Novo e do Velho Testamento. A partir delas construiu listas de frequência, ou seja, a relação das palavras usadas de acordo com o número de vezes que aparecem no texto. O estudo mais importante do autor foi uma comparação entre os textos de Bacon e Shakespeare. Com a ajuda de um financiador, Mendenhall contratou duas secretárias e até criou uma máquina de contar para ajudá-los a analisar 400 mil palavras de Shakespeare e 200 mil de Bacon. Comparados aos dados que já tinha anteriormente, dos estudos feitos com outros autores de língua inglesa, Mendenhall surpreendeu-se com o fato de que a palavra de maior frequência em Shakespeare tinha quatro letras, dado que ele jamais havia constatado em nenhum dos outros autores. Mendenhall constatou, por fim, que as curvas de extensão de palavra em Bacon e Shakespeare eram de fato discordantes. Porém, para a definitiva queda da hipótese levantada por Morgan, a contagem de apenas esse aspecto da obra não foi suficiente para distinguir autoria: Mendenhall descobriu que a curva característica de Shakespeare coincidia com a de Marlowe, conclusão que derrubou o método que utilizara durante toda a vida.

Enquanto nos Estados Unidos surgiam os estudos de autoria, os pesquisadores europeus desenvolviam técnicas estilométricas, a fim de datar os diálogos de Platão em grego. Em 1867, Lewis Campbell, professor de grego em St. Andrews, publicou uma edição de *Sophist* e *Politicus* acompanhada de testes estilísticos. Para ele, variáveis como a ordem das palavras, o ritmo, a ausência de hiatos e a “originalidade do vocabulário”, medida pela frequência e pelo número de hápax (ou palavras de frequência 1, que aparecem uma só vez no *corpus*). Segundo Kenny, o trabalho de Campbell não repercutiu por 30 anos, até que Ritter, um filólogo alemão, chegasse a conclusões semelhantes a partir de métodos similares, em 1888¹⁰².

Outro pesquisador da obra de Platão foi o polonês W. Lutoslawski, que publicou “*The origin and growth of Plato’s logic*”, em 1897. Kenny considera o método dos pesquisadores de Platão “*much more subtle and sophisticated than the crude measured word-length*” usada por Mendenhall.

Enquanto Campbell e Lutoslawski focavam seu trabalho na quantificação de variáveis de estilo concernentes à obra de um único autor, um filólogo americano chamado L. A.

¹⁰¹ MENDENHALL apud KENNY, op. cit., 1982, p. 3.

¹⁰² KENNY, op. cit.

Sherman pensava em usar método semelhante para o estudo da evolução da língua como um todo, o que foi expresso em seus trabalhos que datam de 1888 e de 1892.

No século XX, um dos nomes de destaque é o de Udny Yule, estatístico da universidade de Cambridge que estudou o tamanho médio de frase de autores como Bacon, Coleridge e Macaulay para verificar as prováveis diferenças entre eles. O clássico de sua autoria na área de estatística textual é *The statistical study of literary vocabulary*.

2.2.2 A vertente francesa

Segundo o professor Henri Béhar, da Université Paris III (Sorbonne Nouvelle), pesquisador de ponta no campo dos estudos literários que usam a estatística e a informática como ferramentas, o primeiro trabalho sistemático que envolve a codificação de uma técnica para desenvolver listas de concordância¹⁰³ data de 1642¹⁰⁴. O texto a ser analisado era nada menos do que a Bíblia, o que deixa bastante claro o grau de dificuldade para a confecção do trabalho: a tarefa de mapear à mão todo o vocabulário de um texto tão longo e complexo em termos de linguagem e ainda listar as ocorrências de cada palavra em seus respectivos contextos. A obra é atribuída a Dom Hubert de Phalèse, um monge beneditino que habitava o mosteiro de Afflighem, na Bélgica. Parece claro que Phalèse tenha usado a mão de obra gratuita dos demais monges, caso contrário, tamanho estudo teria sido inviável.

A partir de 1989, o nome desse estudioso seria adotado como pseudônimo de um grupo de pesquisadores da Universidade de Paris III, para configurar a autoria coletiva das publicações resultantes de suas pesquisas auxiliadas pelas novas tecnologias constituídas por computadores e programas específicos para a análise textual. O motivo da escolha, segundo Béhar, coordenador do grupo, se deveu não só ao fato de Phalèse ter sistematizado uma técnica, mas por sua convicção de que tal trabalho, essencialmente coletivo, não deveria ser atribuído a apenas um autor. A fonte que uniu Phalèse ao grupo francês foi o texto *Bibliorum Sacrorum concordantiae ad recognitionem jussu Sixti V Pontif.max...*, de François Pascal

¹⁰³ O conceito de concordância utilizado por Béhar é o de Pierre Guiraud: “list of all the words in a text in all their uses and in context”.

¹⁰⁴ BÉHAR, Henri. Hubert Phalèse’s, Method. *Literary and Linguistic Computing*, Oxford: Oxford University Press, v. 10, n. 2, 1995. p. 129-134.

Dutripion (1793-1867), publicado em 1838 pela editora francesa Belin Mandar¹⁰⁵, que continha um preâmbulo em latim de autoria do monge.

Em 1991, o grupo publicou seu primeiro trabalho sobre a obra de Huysmans¹⁰⁶. Hoje já são 15 livros publicados, com a análise de obras de autores como Molière, Balzac, Céline, Victor Hugo e Samuel Beckett, em quinze anos de trabalho com a estatística textual. Mas o período que separa o Phalèse original do grupo de Paris contém uma longa história. A necessidade de uma demonstração mais objetiva das hipóteses levantadas no campo da crítica literária levou estudiosos da área a se aproximarem de outros campos do conhecimento para a busca de ferramentas complementares capazes de auxiliar seu trabalho de análise, caminho este percorrido anteriormente por outros ramos da pesquisa científica: física, biologia, medicina, sociologia, psicologia e linguística.

Nos estudos linguísticos, segundo Lebart e Salem¹⁰⁷, os primeiros trabalhos estatísticos foram relativos à transcrição estenográfica, com J. B. Stoup, em 1916, e os estudos de distribuição lexical de abordagem “psicobiológica”¹⁰⁸ da linguagem, com G. K. Zipf, em 1935. Outros nomes importantes da história da estatística textual são Marcel Cohen, que publicou trabalho sobre o tema em 1950, e Pierre Guiraud, que publicou em 1960. Um nome que merece destaque é o do francês Charles Muller, professor da Universidade de Strasbourg, que publicou seu primeiro livro sobre o assunto em 1964, sendo um dos pioneiros no uso e na produção de conhecimento sobre a estatística textual. Atualmente, aos 96 anos, tem um sítio para tirar dúvidas de gramática em francês que já está no ar há oito anos e ainda pesquisa e publica. Depois de várias obras publicadas sobre o tema, Muller considera que a linguística e a literatura perceberam tardiamente, em relação a outras áreas das ciências humanas, o quanto a estatística poderia ser útil para os estudos literários.

O desenvolvimento de trabalhos que aliam critérios qualitativos à possibilidade de uma quantificação precisa, praticamente impossível antes do computador, gerou um ramo nos estudos literários que já faz parte de universidades em várias partes do mundo.

Na França, o nome mais expressivo quando se trata de estatística textual aplicada aos estudos literários é o do professor Etienne Brunet. Além de ser um dos pioneiros nessa

¹⁰⁵ De acordo com o pesquisador Majid Sekhraoui, que usou a obra de Dutripion como referência em seu trabalho de DEA (Diplôme d'Etudes Approfondies), apresentado em 1983, na Université Paris III, o livro foi publicado por outras editoras de várias partes do mundo. O registro oficial mais antigo que encontramos é o de uma edição que faz parte do catálogo da Biblioteca do Vaticano, de 1853 (Paris, Éditions Eugène Belin). Ver: www.vaticanlibrary.vatlib.it/BAVT/home.asp?LANGUAGE=eng&DPT=gen.

¹⁰⁶ PHALÈSE, Hubert de. **Comptes a rebours** : l'oeuvre de Huysmans à travers les nouvelles technologies. Paris: Nizet, 1991. (Collection Cap'agreg).

¹⁰⁷ LEBART; SALEM, op. cit.

¹⁰⁸ LEBART; SALEM, op. cit., p. 16.

metodologia – sua tese de doutorado *Le vocabulaire de Giradoux : structure et evolution* é de 1978 –, tem uma produção invejável. Alguns de seus trabalhos são: *Index de l'Emile* e *Concordance de l'Emile*, ambos de 1980; *Index de l'oeuvre théâtrale et lyrique de J.J. Rousseau*, de 1986; *Le vocabulaire français de 1789 à nos jours*, em 3 volumes (1981); *Le vocabulaire de Proust* (1983); *Le vocabulaire de Zola* (1985); e *Le vocabulaire de Victor Hugo* (1988). Brunet trabalha ainda com o desenvolvimento de programas de computador para estudos literários: *Hyperbase*, lançado em 1999, e *Thief*, além dos cederrons sobre vários autores de língua francesa, como Paul Eluard, de 1996; Balzac, também de 1996; e ainda os sobre Rimbaud, Proust, Pascal e Rabelais, todos editados em 1999.

No Brasil, a estatística textual já tem pesquisadores em algumas universidades, mas na área de literatura o uso dessa metodologia é inédito. O ramo que nos interessa é o dos estudos estilísticos auxiliados por essas novas tecnologias, uma vez que, além de mais precisas do que uma quantificação manual, são muito mais rápidas. A estilometria é área que busca os padrões de repetição de elementos que compõem o texto. São esses padrões que compõem em grande parte a identidade de um autor. A eles se misturam os padrões da língua, as características próprias do gênero literário, do tema e as influências da época, o que torna o desafio maior ainda.

Há obras amparadas na estatística textual de autores de outras nacionalidades como o caso do italiano Roberto Busa. Segundo Rockwell, ele foi o pioneiro no uso da informática no tratamento de textos¹⁰⁹. No fim dos anos 40, Busa tinha como tecnologia disponível tabuladores eletromecânicos que usou na produção de concordâncias para seu trabalho *Index Thomisticus*, sobre as obras de São Tomás de Aquino. Em 1950, Busa migra para os computadores eletrônicos e nos anos 1970 publica seu estudo em suporte impresso. Só nos anos 1990 (1992) sai o cederrom, enfim a versão eletrônica de seu trabalho. Hoje um dos principais prêmios internacionais para pesquisas que apliquem a tecnologia da informação na área de humanas leva seu nome¹¹⁰.

¹⁰⁹ ROCKWELL, Geoffrey. What is text analysis, really? *Literary and Linguistic Computing*, v. 18, n. 2, p. 209-219, 2003.

¹¹⁰ Sítio da ALLC, responsável pelo prêmio triannual Roberto Busa: <http://www.allc.org/awards/busaawd.htm>.

2.3 Principais Contribuições

*What is important is for the scholar to understand what statistics are valid for the data being studied and how to interpret the results*¹¹¹.

Muitas são as abordagens teórico-metodológicas possíveis quando se trata de estatística textual aplicada à literatura. Cada pesquisador desenvolve seu modelo de análise de acordo com seus objetivos e formação, mas todos os trabalhos, mesmo os que não chegaram a resultados satisfatórios, ajudaram a criar os fundamentos para pesquisas posteriores. Por isso, a seguir traçaremos um panorama dos principais estudos de análise literária e de atribuição de autoria, que são fundamentais para um trabalho focado no estilo, como o nosso.

2.3.1 *Junius Letters*

Ellegård teve como objeto de sua investigação as chamadas “*Junius Letters*”, textos publicados entre 1769 e 1772 na *Public Advertiser*. O próprio Ellegård ironizava a importância de seu objeto tratando-o como “*one of the minor mysteries of English literary history*”¹¹².

A fim de distinguir a identidade do autor em questão de seus contemporâneos, o pesquisador desenvolveu um cálculo que seria uma “razão de distinção”, que era a divisão da taxa de frequência nos textos de Junius pela frequência em uma amostra de um milhão de palavras de 59 escritores contemporâneos dele. Desse cálculo Ellegård criou as “categorias” “*Junius plus-expressions*” e “*Junius minus-expressions*”, que mapeavam as palavras e frases mais usadas por ele em relação aos outros autores da época.

Outro recurso metodológico usado pelo pesquisador era um teste feito com um conjunto de sinônimos, para verificar as formas preferidas pelo autor. À época, os computadores não tinham a capacidade de cálculo que têm hoje, logo Ellegård usou um método que Hockey classifica como de tentativa-e-erro. Mesmo assim, essa autora considera as “*plus*” e “*minus expressions*” como a maior contribuição metodológica do pesquisador.

¹¹¹ HOCKEY, op. cit., p. 115.

¹¹² HOCKEY, op. cit., p. 107.

2.3.2 *Federalist Papers*

Os “*Federalist Papers*”, publicados entre 1787 e 1788, eram uma série de 85 ensaios dirigidos à população de Nova Iorque com o objetivo de convencê-la a ratificar a Constituição norte-americana. A autoria era assinada como *Publius*, e era sabido que se tratava de autoria coletiva de três pessoas que se revezavam na escrita: Hamilton, Madison e Jay. A dúvida sobre a autoria era relativa a 12 ensaios, mas provas externas restringiam a disputa apenas entre Hamilton e Madison.

Os primeiros estudos sobre esse caso foram feitos em 1940 por Mosteller, em parceria com outro pesquisador, e tinham seu foco principal em métodos estatísticos. Eles começaram pela análise de extensão de frase, mas depois de muito trabalho de contagem acabaram por concluir que a média em Hamilton e Madison era praticamente a mesma, de 35 palavras por frase. Os resultados desanimaram Mosteller, que parou a investigação para só retomá-la com Wallace na década de 1960 (1964). Dessa vez, optaram por investigar o vocabulário de Hamilton e Madison, partindo das 30 palavras mais usadas nos textos de autoria comprovada de ambos. O método foi eficaz, e os dados apontaram que Madison era o autor dos 12 ensaios.

Segundo Hockey¹¹³, esse estudo transformou-se num clássico dos trabalhos sobre autoria, em parte porque os antecedentes históricos são bem documentados e porque o conjunto de textos não é tão grande para projetos escolares nem tão pequeno para a análise estatística. Para Hockey, as duas maiores contribuições metodológicas desse estudo são:

- a) a investigação e comparação de sinônimos; e
- b) o contexto geral do problema, tendo apenas dois candidatos com farto material para comparação da parte de ambos, de mesmo gênero sobre o mesmo assunto.

Mesmo com a questão da autoria resolvida, outros autores retomaram os *Federalist Papers* para testar novas metodologias. Foram eles: Merriam; Holmes e Forsyth; e Tweedie, Singh e Holmes. Hockey considera os primeiros trabalhos como fundadores, no sentido de que deram a base sobre o que contar. Embora reconheça que Mendenhall encontrou algo de interessante no estudo da extensão das palavras, ela relativiza essa importância usando o argumento de C. B. Williams de que o tamanho das palavras está relacionado ao contexto¹¹⁴.

¹¹³ HOCKEY, op. cit.

¹¹⁴ HOCKEY, op. cit., p. 109.

A mensuração de extensão de frase traz consigo o problema da definição de frase. Para os programas em geral, é a pontuação forte que determina início e fim de frase. Assim, pode-se ter problemas com abreviaturas, que podem ser codificadas (ou lematizadas) à parte. Há outros aspectos a serem observados e vários cuidados a serem tomados para evitar esses problemas. Em nosso caso, porém, como usamos um *corpus* muito grande, esses erros acabam por formar uma fatia insignificante na contagem do todo, e não vale a pena o tempo despendido para procurar cada caso.

A mensuração de extensão de frase, segundo Hockey, começou com W. C. Wake, em 1948, mas foram A. Q. Morton e seus colaboradores quem mais usaram esse recurso em seu estudo sobre a prosa grega. Outros casos do uso desse tipo de recurso deixaram dúvidas quanto a sua eficácia.

J. N. Binongo, graças a um estudo sobre o estilo de um escritor filipino publicado em 1994, concluiu que um mesmo autor pode variar bastante o tamanho de frase de um trabalho para outro. (Acreditamos que em Machado de Assis é o que acontece, ao compararmos o estilo do início de sua carreira, repleto de diálogos, ao estilo maduro, mais narrativo.) O que acabou por ajudar Binongo a distinguir o estilo de seu autor foi a lista de palavras mais usadas – no caso, ele optou por observar as 36 palavras mais frequentes.

A frequência das palavras, particularmente a das “common words”, pode ser considerada como o melhor discriminador para fins de investigação de autoria, porque este tipo de palavra é relativamente independente do tema abordado e elas são frequentes o suficiente para serem observadas mesmo em um *corpus* pequeno¹¹⁵. Como exemplo, Hockey cita os marcadores de autoria dos *Federalist Papers*, em sua maioria palavras comuns, ao contrário das *Junius Letters*, em que Ellegård teria encontrado a maior dificuldade no fato de não ter focado especificamente esse tipo de palavra. O uso desse recurso aparece também no trabalho de Burrows.

Outro tipo de estudo usa muitas palavras comuns, separando os homógrafos e dividindo as palavras de acordo com sua função. Um exemplo é o de Kenny, *The Aristotelian Ethics*. O pesquisador escolheu uma série de discriminadores que acabam por cobrir 60% do vocabulário do *corpus*. São 36 partículas, 19 preposições, vários pronomes, advérbios, demonstrativos e artigos definidos.

¹¹⁵ HOCKEY, op. cit., p. 110.

2.3.3 Análise multidimensional ou multivariada

Ao invés de se basear em um ou dois fatores, a análise multidimensional ou multivariada investiga a relação entre muitos objetos diferentes, com muitas variáveis. Exemplos são os trabalhos de Biber e Frischer e colaboradores.

Hockey¹¹⁶ afirma que algumas técnicas de estudo multidimensional operam mediante a redução das diferentes variáveis a um pequeno número de fatores subjacentes, que são carregados positiva ou negativamente com alguma das variáveis em estudo. O primeiro fator ou dimensão contém a maior quantidade de informação e, dependendo da natureza dos dados, a maior parte das variáveis pode ser representada por um pequeno número de fatores – às vezes, somente dois. É então possível examinar as relações entre os objetos ou variáveis colocando-os em um diagrama que mostre as relações espaciais como acontecem em algumas funções disponíveis no Hyperbase, a exemplo da análise fatorial e seus diagramas e gráficos correspondentes.

Para Hockey, não há regras rígidas para os estudos de autoria, o que vale é a máxima de que “mais é melhor”, ou seja, “mais textos, mais testes”¹¹⁷.

Estelle Irizarry tem um estudo sobre a obra *The misfortune's of Alonso Ramirez*, de Carlos de Sigüenza Góngora. Este autor afirmava que seu livro era a transcrição de uma narrativa (oral) de um marinheiro analfabeto, o próprio Ramirez do título. Para verificar a veracidade da afirmação de Góngora, a pesquisadora comparou o romance com três textos narrativos do autor. Ela constatou que o romance era composto de frases mais curtas, mas que não havia diferenças significativas na extensão das palavras. Além disso, a autora listou as palavras mais e menos usadas e pôde observar também construções sintáticas e morfológicas; para que fossem analisadas, estas precisariam aparecer, no mínimo, cinco vezes em uma mostra de 4 mil palavras. Por fim, Irizarry constatou que várias palavras e expressões ocorriam com maior frequência nas narrativas do que no romance. Após vários testes, a pesquisadora concluiu que é muito provável que Ramirez seja, de fato, uma pessoa, e não apenas personagem.

Dixon e Mannion, um estudioso das Humanas e um matemático, estudaram onze ensaios periódicos publicados postumamente atribuídos a Oliver Goldsmith. Para a investigação de autoria foram usados textos de Goldsmith e de quatro outros autores.

¹¹⁶ HOCKEY, op. cit., p. 116.

¹¹⁷ HOCKEY, op. cit., p. 117.

A questão da autoria envolvendo as Epístolas de S. Paulo virou um clássico da área, assim como os *Federalist Papers*. O caso foi retomado por muitos estudiosos desde a carta de Morgan, em 1851. Há o artigo de W. C. Wake, *The authenticity of the Pauline Epistles*, publicado em 1948, que, por sua vez, tem seu método de análise de extensão de frase retomado por Morton em seu trabalho de 1978, *Literary detection: how to prove authorship and fraud in literature and documents*. O tema das epístolas aparece também em artigo de Morton e McLeman, de 1966, *Some indications of authorship in Greek prose*, no texto de D. L. Mealand, *On finding fresh evidence in old texts: reflections on results in computer-assisted biblical research*, publicado em 1992, e, ainda, no trabalho de K. J. Neumann, *The Authenticity of the Pauline Epistles in the light of stylostatistical analysis*, de 1990.

A seguir, trataremos de mais algumas obras, agrupando-as de acordo com o tipo de elemento a ser investigado ou com a abordagem escolhida, segundo os critérios adotados por Hockey¹¹⁸.

2.3.4 Distinção de gênero literário

Lessard e Hamm estudaram as estruturas repetidas em Stendhal, e a análise do vocabulário pôde também contribuir para a investigação de gênero literário. Hockey afirma que esse tipo de abordagem é metodologicamente muito similar ao trabalho sobre variação em textos falados e escritos feito por Biber, em 1998, embora o contexto (ou análise literária) leve a uma diferente ênfase na interpretação.

Em 1991, Craig também fez um estudo nessa direção. O autor investigou o uso dos pronomes plurais em treze peças de Shakespeare e em treze de Jonson, a fim de ver como eles podem ser usados para distinguir diferentes gêneros literários.

Um dos nomes mais importantes dos estudos literários auxiliados por computador é o de John Burrows, professor da Universidade de Newcastle, na Austrália. Em 1992, ele fez um estudo sobre gênero literário a partir de quatro tabelas contendo as menores contagens das 50 palavras mais comuns em quatro textos. As altas frequências de *I, You, He, She* e do verbo *Said* indicaram corretamente que esses textos eram romances. O exame da segmentação dos textos mostrou onde o diálogo predominava. Burrows também estudou os romances de Jane

¹¹⁸ HOCKEY, op. cit.

Austen, que, conforme Hockey¹¹⁹, é das poucas “full length monographs based on computational approaches” e data de 1987. Nesse caso, o argumento básico é o de que as 30 palavras mais comuns (oito pronomes pessoais, seis formas de verbos auxiliares, cinco preposições, três conjunções, dois advérbios, os artigos definidos e indefinidos e *to*, *that*, *for*, e *all*) teriam sido quase que completamente negligenciados no estudo desses romances, mas que tais palavras poderiam ajudar a esclarecer as relações entre narrativa e diálogo, entre os personagens e entre romances e romancistas.

O estudo baseou-se em concordâncias de seis romances de Austen, acrescidos de *Sanditon and Sanditon by Jane Austen and Another Lady*; *Frederica*, de Georgette Heyer; *The Waves*, de Virginia Woolf; *The Awkward Age*, de Henry James e *Howards End*, de E. M. Forster. Um dos métodos usados por Burrows foi comparar as correlações entre pares de personagens¹²⁰. O autor também fez estudos concentrando parte da investigação no herói e na heroína de cada romance.

Outros autores usaram as ferramentas eletrônicas para estudar questões relativas aos gêneros literários. Merideth estudou o diálogo das heroínas em três romances de Henry James, contrastando a heroína de cada romance com um personagem masculino (*Daisy Miller*, *Portrait of a Lady* e *The Bostonians*). Merideth focou sua análise na incidência de perguntas, imperativos, exclamações, pausas, fragmentos, condicionais, definições, negativas, universais, advérbios e comparações. Irizarry fez um estudo sobre idioletos “gênero-relacionados” em dois autores mexicanos contemporâneos (um homem e uma mulher), a fim de verificar diferenças de escrita relacionada ao gênero (aqui não mais literário) dos respectivos autores.

2.3.5 Análise temática

Fortier e McConnel desenvolveram ferramentas para a análise temática na literatura francesa. Fortier mostra a predominância do tema da violência no capítulo sobre a África em *Voyage au bout de la nuit*, de Céline, em comparação com os capítulos relativos à guerra do mesmo livro. Há outros trabalhos temáticos, como o que estuda as dimensões míticas em Malraux e sobre a associação entre doença e saúde no *L’Immoraliste*, de Gide. Hockey

¹¹⁹ HOCKEY, op. cit., p. 70.

¹²⁰ HOCKEY, op. cit., p. 70.

considera o método aplicado nesses trabalhos similar ao tipo de análise de conteúdo feito nas ciências sociais.

Ide investigou as imagens em *The Four Zoas*, de Blake. A autora institui uma série de categorias de imagem usando listas de palavras e um dicionário de símbolos, para então identificar suas ocorrências no poema. Conforme Hockey, a intenção era tentar identificar as relações espaciais entre as imagens para auxiliar a esclarecer os símbolos e os sistemas de relações em Blake.

J. B. Smith usou ferramentas similares às de Ide para estudar vários temas em Joyce, ligando a densidade do tema ao conteúdo, mas também enfatizando que a crítica precisa interpretar os resultados. Corns também estudou a imagem em Milton, mas Hockey o critica por não ter definido imagens “em termos exatos”. Miall fez estudo sobre as palavras associadas ao emocional em Coleridge, e Rommel estudou Robinson Crusoe.

2.4 Nasce uma Nova Crítica?

*The computer is merely a tool. It can do some things very well, but much should be left to the judgement of the scholar, both to situate the project in the broader range of scholarship on the particular topic and to interpret the results within the context of other research*¹²¹.

Quando tratamos de ferramentas tecnológicas que aliam informática, estatística e literatura, tal mistura suscita muitas discussões. Vamos nos ater aqui a duas posições epistemológicas que dividem as opiniões dos pesquisadores desse novo campo: o uso de programas de computador específicos para a análise textual funda uma nova crítica ou uma nova teoria literária? Veremos aqui duas posições opostas.

A primeira que abordaremos é a linha defendida por Rockwell. Em artigo publicado em 2003¹²², o autor trata da relação entre o conceito e as teorias tradicionais do texto literário e as novas ferramentas tecnológicas para tratamento e análise de texto. O autor aborda a tradição “editorial” impressa e todo um modelo teórico e crítico criado para cuidar do texto impresso. A questão é se o texto em novo suporte (digital) pode ou deve ser analisado da maneira tradicional, já que a própria transposição do papel para o meio digital já configuraria

¹²¹ HOCKEY, op. cit., p. 66-67.

¹²² ROCKWELL, op. cit.

um texto classificado por Rockwell como “híbrido”, tomando emprestado o conceito desenvolvido por Mikhail Bakhtin para tratar da linguagem no romance.

O autor acredita em uma ampliação da capacidade de análise do pesquisador, a partir da adoção de ferramentas tecnológicas (programas de análise de textos), no sentido de possibilitar a abordagem de um maior número de variáveis ou de um número maior de textos – como no nosso caso, em que, graças ao programa Hyperbase, podemos mapear, além do vocabulário, vários outros aspectos dos textos e ainda traçar uma relação estatística entre os vários elementos do *corpus* (ou dentro de cada texto, considerado como um *subcorpus*). Sem o auxílio do computador ou do programa mencionado, seria impossível levantarmos tamanho volume de informação sobre o *corpus*. A alternativa seria termos um grupo muito numeroso de pesquisadores engajados por um longo período sobre esse vasto material (205 contos, quase 3 mil páginas).

Rockwell argumenta que a evolução tecnológica que sofisticou as ferramentas disponíveis para análise textual gerou novas possibilidades de levantamento de dados que vão bem além das concordâncias¹²³. Isso significa que o pesquisador pode formular questões mais complexas, pode refinar não só o dado que pretende extrair por meio dessa ferramenta. Tal mudança atinge também o modo de pensar do pesquisador: ela o provoca, transforma e o desafia a formular questões mais complexas e sofisticadas. O comportamento do pesquisador também se altera, e isso pode redundar em uma mudança para além da tecnologia. Seria uma mudança hermenêutica? A pergunta é: será que o arsenal teórico-crítico do pesquisador também não precisaria se transformar e se sofisticar diante das possibilidades que essa nova ferramenta lhe oferece, em termos de acesso a um número muito maior de informações até então impossíveis de se alcançar com tamanha rapidez e precisão?¹²⁴

A contribuição desse questionamento está centrada na ideia de mudança nos princípios que norteiam a concepção das ferramentas. Os princípios de unidade e coerência que sustentam o conceito tradicional de texto poderiam ser repensados a fim de criar ferramentas baseadas nos princípios de pesquisa como um “jogo disciplinado”¹²⁵, pois, para ele, a interpretação das concordâncias é que daria unidade ao texto, e não a concordância em si mesma. Rockwell sugere um jogo de experimentação com as ferramentas disponíveis, de modo a criar possibilidades de interpretação, mas destaca que, mesmo para experimentar, é

¹²³ ROCKWELL, op. cit.

¹²⁴ ROCKWELL, op. cit., p. 211.

¹²⁵ ROCKWELL, op. cit., p. 213.

necessário que o crítico ou pesquisador esteja munido do que chama de “intuição pré-teórica”¹²⁶.

O principal risco desse posicionamento em relação às ferramentas tecnológicas é o de se acreditar que elas são uma solução mágica, capaz de substituir o olhar crítico do pesquisador. Não há utilidade em uma tecnologia se o pesquisador não tem questões para responder ou se não há uma análise crítica dos dados fazendo a correlação com o contexto de seu lugar e seu tempo, ou se não se conhece o restante da obra do autor. Por isso vamos à outra posição, defendida por Michel Bernard e Susan Hockey.

Bernard evita qualquer mistificação em torno da tecnologia e afirma que o computador tem como função principal facilitar a vida do pesquisador:

Il ne s’agit donc pas de l’irruption dans le champ littéraire d’un intrus exogène et imposé de l’extérieur mais de la rencontre assez naturelle entre les techniques de la recherche littéraire et des outils que la facilitent en la déchargeant de ses tâches les plus ingrates¹²⁷.

Susan Hockey corrobora a opinião de Bernard afirmando que o computador é apenas uma ferramenta¹²⁸.

O computador é de fato uma ferramenta, porém revolucionária. Os programas de análise estatística vieram para cumprir um novo papel que o pesquisador anteriormente não teria condições de fazê-lo. Essas ferramentas não vieram substituir nada: vieram fundar uma metodologia que une várias áreas do conhecimento, a fim de permitir uma exploração do texto impossível de ser feita manualmente, na extensão que essas tecnologias permitem. Mapear dados e buscar elementos, objetos que em conjunto constituem aspectos de uma obra, com precisão e rapidez em *corpus* de mais de 3 mil páginas como o nosso seriam tarefas impossíveis de ser feitas manualmente.

Que o uso dessas tecnologias fundou um novo campo de estudos é uma realidade que o enorme número de pesquisadores, estudos, simpósios e textos publicados sobre o assunto bastam para comprovar. Mas será que toda essa informação por si só é capaz de fundar uma nova crítica?

A revolução trazida pelo computador está na exploração de informações antes inacessíveis, principalmente quando levamos em conta que com ele o pesquisador não tem acesso apenas a dados brutos – listas das palavras ou número de frases ou parágrafos –, mas

¹²⁶ ROCKWELL, op. cit., p. 214.

¹²⁷ BERNARD, op. cit., p. 8.

¹²⁸ HOCKEY, op. cit., p. 66.

às relativizações feitas pelas funções estatísticas que permitem comparar dados de textos de diferentes tamanhos com o percentual relativo ou mesmo a aplicação de modelos de probabilidade que acabam por mostrar ao pesquisador o quanto um dado é relevante em termos numéricos.

No entanto, não pode haver uma crítica que prescindia do conhecimento profundo da obra sem que haja uma teoria que sustente a organização e a interpretação desses dados. Antes mesmo de extraí-los, o pesquisador tem que ter em mente que tipo de informação deseja extrair do *corpus*, de acordo com os objetivos de seu estudo. Dessa posição aparentemente simples surgem as principais questões metodológicas do campo. O que contar? Que recursos estatísticos escolher para organizar o que foi contado e dar uma resposta satisfatória às dúvidas ou questionamento que pode ajudar a solucionar? Os dados coletados são suficientes? Deve-se usar amostra? De que tamanho?

A máquina torna acessíveis recursos inéditos que não servem para nada se o pesquisador não souber o que quer extrair dela e como fazê-lo. Essa última questão é mais fácil de responder: é preciso estudar e familiarizar-se com os conceitos e técnicas das outras áreas envolvidas no processo. Já a primeira dúvida – o que extrair – é mais complexa no sentido de que é a reflexão que deu origem ao trabalho que está em jogo: o que busco? para quê?

A base teórica e a familiaridade do pesquisador com a obra e a crítica sobre ela são insubstituíveis, e não há estudo literário que seja possível sem essa bagagem. Não há ferramenta mágica que seja capaz de dar forma a um trabalho sem que os requisitos acima estejam presentes. Daí a nossa concordância com o pensamento de Bernard:

Il me semble en effet que l'ordinateur ne génère rien qui ressemblerait à une « nouvelle critique », à une approche radicalement originale du phénomène littéraire mais qu'il peut, en revanche, se mettre au service de toutes les lectures, de toutes les pratiques de la recherche, de plus traditionnelles aux plus nouvelles¹²⁹.

A principal função do computador, afirma Hockey, é a de testar hipóteses e intuições do pesquisador¹³⁰. Nos estudos literários, o pesquisador já tem objetivos a serem atingidos e intuições a serem testadas, a partir do instrumental tradicional (obra, bibliografia crítica, informações sobre o contexto, etc.) de que dispõe. Mas, muitas vezes, tudo o que se tem quando se trata do vocabulário de um autor é apenas uma impressão ou uma intuição que, se

¹²⁹ BERNARD, op. cit., p. 16.

¹³⁰ HOCKEY, op. cit.

puderem ser testadas, podem se transformar em um dado relevante para ser trabalhado. Hockey afirma que, nos estudos auxiliados por computador, “*more is more*”, mais amostras e mais testes sempre assegurarão maior probabilidade de obtermos uma informação confiável. É claro que sempre haverá quem conteste o uso dos números nas Humanas, mas acreditamos que, ao afirmarmos algo amparados em um dado que tem um grau de precisão alto, estamos buscando maior rigor ao nosso trabalho: “*Whether or not linguistic habits are measurable in this way is itself debatable, but various projects have shown what it is possible to characterize an author or a set of texts by linguistic fingerprints*”¹³¹. Ou ainda: “[...] *even simple counts can help to reinforce a feeling about a text or show that what intuitively seems rare or very frequent is not in fact so*”¹³².

Outra função é a de traçar um panorama do instrumental, principalmente de vocabulário, utilizado nos textos em estudo. Com esse material em mãos, o pesquisador tem um instrumento eficaz para fins de comparação entre textos, função bastante explorada nas investigações de autoria. Mesmo sendo considerada uma área independente dos estudos literários – pois pode servir à investigação de autoria em várias áreas –, esse tipo de estudo pode nos ser muito útil em termos de método. Um dos motivos é que o foco dos estudos de autoria está centrado no reconhecimento de padrões estilísticos dos autores que são candidatos à autoria do texto em questão. Como neste trabalho estudamos as possíveis variações de padrões estilísticos em um mesmo autor, as técnicas e métodos desenvolvidos para atribuição de autoria e os testes de comparação terminam por se assemelhar.

Como vimos nas seções anteriores, cada caso, em se tratando de estudos literários, exige um modelo de análise específico. A investigação de alguns elementos já se comprovou eficaz na distinção de estilo, mas é na análise dos dados que o pesquisador terá a segurança para decidir se deve continuar na direção que seu conhecimento prévio e sua intuição apontavam ou se deve mudar de rumo.

¹³¹ HOCKEY, op. cit., p. 104.

¹³² HOCKEY, op. cit., p. 66.

CAPÍTULO 3 ESTILO E QUANTIFICAÇÃO

3.1 Questões de Estilo

Primeiramente, é essencial que nos posicionemos acerca da natureza de nosso objeto. Acreditamos que a obra literária seja, antes de tudo, “algo idêntico a si mesmo”¹³³, e a consequência imediata disso é considerarmos a linguagem como seu principal elemento constitutivo. É a linguagem que delimita o campo das significações, das interpretações possíveis de uma obra literária. É ela que dá identidade, possibilitando que todos possam reconhecer uma obra como a mesma, e não outra¹³⁴. É no trabalho de linguagem, no modo como o autor lida com esse instrumento, que está sua arte.

Assim como Ingarden afirma que, na leitura, a linguagem tem a primazia sobre a ideia da obra¹³⁵, Leech e Short descrevem a direção da decodificação do texto a partir do nível grafológico¹³⁶. Como os estudos desses autores partem do ponto de vista do leitor, isso justifica a definição do ponto de partida na observação da linguagem. Porém, a análise estilística de Leech e Short baseia-se num movimento cíclico que não impõe um ponto de partida fixo; ela pode começar tanto da descrição do material linguístico quanto da interpretação literária, retomando o método de Leo Spitzer, do *philological circle* ou *circle of understanding*:

Spitzer argued that the taste of linguistic-literary explanation proceeded by the movement to and from linguistic details to the literary centre of a work or a writer’s art. There is a cyclic motion whereby linguistic observation stimulates or modifies literary insight, and whereby literary insight in its turn stimulates further linguistic observation¹³⁷.

¹³³ “Por causa do papel fundamental que desempenha na economia da obra literária, Ingarden considera a linguagem um estrato constitutivo da sua essência. A obra literária é essencialmente constituída de formações linguísticas (palavras, frases, períodos), que são um todo composto por dois elementos: um material fônico e um sentido. Estes elementos interagem de forma que o material fônico é ‘multidiferenciado’ e ‘diversamente ordenado’ pelo sentido; e o sentido é ligado ao material fônico. O material fônico é a forma que representa um sentido. E esta forma varia de acordo com as mudanças de entonação, timbre e intensidade, que o sentido exigir”. INGARDEN, Roman. *A obra de arte literária*. 2. ed. Tradução de Albin E. Beau, Maria da Conceição Puga e João F. Barrento. Lisboa: Fundação Calouste Gulbenkian, 1965, p. 32.

¹³⁴ FREITAS, Deise J. T. *A revolução do estilo de Machado de Assis: uma análise de quatro contos*. 1998. Dissertação (Mestrado em Teoria Literária) – Curso de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis, 1998, p. 17.

¹³⁵ INGARDEN, op. cit., p. 169.

¹³⁶ LEECH, Geoffrey N.; SHORT, Michael H. *Style in fiction*. New York: Longman, 1990.

¹³⁷ LEECH; SHORT, op. cit., p. 13.

Também focado no problema da linguagem é o ponto de vista de Pierre Guiraud sobre a literatura:

[...] l'essence de l'oeuvre littéraire, est moins dans les idées en elles-mêmes que dans la réalisation linguistique qui les manifeste.

Il y a une sorte d'inversion des fonctions ordinaires du langage: dans la communication littéraire, la langue n'est plus un moyen mais un fin. Toutes les expériences de la poésie, du théâtre, du roman modernes sont fondées sur cette propriété; l'oeuvre cesse d'être la simple traduction d'une expérience pour être réalisation des virtualités expressives de la langue [...]

Bref, la littérature, la philosophie, les arts, sans parler des sciences humaines, apparaissent aujourd'hui comme une méditation sur la nature et les fonctions du langage.

De ces oeuvres nouvelles, la critique doit être linguistique. Et cette reconversion des critères traditionnels nous engage, em même temps, à une réévaluation des oeuvres anciennes¹³⁸.

Como nosso trabalho é focado no problema do estilo de um autor em determinado gênero e nas suas transformações dentro deste, é importante precisar o conceito de estilo que adotamos – uma entre várias abordagens possíveis do tema.

Para Guiraud, o estilo – que tem a origem de seu nome em *stilus*, espécie de estilete que servia para escrever em tabuinhas, antes do papel e da pena de ganso” – “é a maneira de escrever, a utilização pelo escritor dos meios de expressão para fins literários, distinguindo-se, portanto, da gramática, que define o sentido e a correção das formas”. E “é só a língua literária que interessa ao estilo, especialmente o seu rendimento expressivo, o “colorido”, como se dizia, próprio para convencer o leitor, agradá-lo, manter vivo o seu interesse, impressionar-lhe a imaginação mediante formas vivas, pitorescas, elegantes e estéticas”¹³⁹.

Leech e Short apontam dois conceitos de estilo: um em sentido amplo, que é o de escolha linguística; e outro em sentido estrito, em que considera a possibilidade de se distinguir entre *o que* o autor quer dizer e *o modo* que ele escolhe para fazê-lo¹⁴⁰. O conceito mais amplo nos remete à afirmação de que estilo é um desvio em relação a uma norma, mesma premissa utilizada por Leech e Short para falar do estilo pessoal como *idioleto* em relação à norma ou *dialeto*:

Linguistics have used the term DIALECT for varieties of language which are linguistically marked off from other varieties and which correspond to geographical, class, or other divisions of society. A DIALECT is thus the

¹³⁸ GUIRAUD, op. cit., p. 14.

¹³⁹ GUIRAUD, Pierre. *A estilística*. Tradução de Miguel Maillet. São Paulo: Mestre Jou, 1970, p. 17.

¹⁴⁰ “Alternative ways of rendering the same subject matter”. LEECH; SHORT, op. cit., p. 39.

particular set of linguistic features which a defined subset of the speech community shares; IDIOLECT refers, more specifically, to the linguistic “thumbprint” of a particular person: to the features of speech that mark him off as an individual from those around him¹⁴¹.

Se o estilo pertence à “escolha linguística”, se é, antes de tudo, um padrão de escolhas, e transparece em preferências repetidas no texto, consideramos necessário um meio de medir essas preferências¹⁴². Tal necessidade, por sua vez, remete à estatística, afirma Leech:

Aesthetic terms used in the discussion of style (exuberant, vigorous, plain) are not directly referable to any observable linguistic features of texts [...] the more a critic wishes to substantiate what he says about style, the more he will need to point to the linguistic evidence of texts; and linguistic evidence, to be firm, must be couched in terms of numerical frequency¹⁴³.

No entanto, a quantificação envolve uma série de problemas. O primeiro se refere à natureza estética do objeto literário. O caráter objetivo da estatística pode levar o pesquisador a buscar dados completos da obra, a buscar um padrão da língua como referência de comparação do que segue a norma linguística e do que pode configurar um desvio. Esse tipo de esforço, segundo os autores, não apresenta resultados. Primeiro, porque é impossível mapear todos os recursos linguísticos utilizados numa obra, e depois porque nem tudo o que aparece na obra é importante. Também não se pode limitar a observação a uma só característica. Deve-se relacioná-la a outras contra um pano de fundo de tendências de preferências repetidas no texto e medir sua frequência. Para isso, contudo, é necessário estabelecer um critério de seleção para as amostras a serem comparadas.

Se o estilo deve ser medido em termos de desvios e se o uso dos dados numéricos deve ser adaptado às necessidades literárias, para a estilística mais importante do que a frequência de preferência por certo recurso linguístico é a determinação do lugar que ocupa no texto. Então, a principal função da estilística quantitativa é a de confirmar ou negar as intuições ou *insights* do pesquisador e levá-lo a novos *insights*, complementares¹⁴⁴.

A ideia de norma e desvio vem originalmente de Valéry, como explica Maciel:

¹⁴¹ LEECH; SHORT, op. cit., p. 13. (Grifo dos autores).

¹⁴² FREITAS, op. cit., p. 39-40.

¹⁴³ LEECH; SHORT, op. cit., p. 46.

¹⁴⁴ “Stylistics often uses not categories of the language as such, but special stylistics categories, derived, by abstraction and combination, from more basic linguistic categories. It seems, then, that the list of linguistic features is indefinitely large, if we want a quantitative description of a text to have a fine enough mesh to catch the linguistic details which contribute to reader’s feeling for differences of style. The quest for a completely objective measurement of style must be abandoned on this score, as well as on the determining frequencies for the language as a whole”. LEECH; SHORT, op. cit., p. 46-47.

C'est en effet à Paul Valéry – statisticien lui-même – que nous devons cette définition du style, reprise ensuite par de nombreux chercheurs linguistes et littéraires, dont Pierre Guiraud, selon laquelle “le style est un écart par rapport à une norme”. Entendons-nous toutefois – et immédiatement! – sur l'importance qu'elle peut avoir pour nous: cette définition a pour nous l'intérêt de reprendre ces termes-clés des calculs de probabilité qui sont “norme” et “écart” [...] Et c'est bien là le sens de notre démarche et le but du calcul statistique: l'étude de la structure théorique et les écarts provenant de la comparaison entre les données théoriques – ou *modèle* – et les données réelles, c'est-à-dire l'*observation*. On ne saurait donc accorder au mot “norme” un sens différent de celui de “moyenne des usages dans le corpus considéré”! En conséquence – et cas par cas – si l'effectif réel s'écarte beaucoup de celui résultant d'une répartition théorique, nous dirons qu'il y a eu des changements dans le style¹⁴⁵.

Se o estilo é um desvio em relação a uma norma, é possível quantificá-lo. Isso não significa dizer que seja possível, por meio da quantificação de traços estilísticos, esgotar uma obra – já que as partes contadas fazem parte de um todo concatenado de modo intencional e tal ordem é parte do contexto que modifica o sentido que as palavras têm se analisadas isoladamente. Daí a necessidade de o pesquisador conhecer muito bem seu objeto¹⁴⁶. Como Machado faz um largo uso da ironia, ninguém seria ingênuo de acreditar que a análise do material linguístico, fora do contexto, vá dar conta da complexidade da obra. No entanto, se esse método não comporta determinados aspectos da obra, para outros ele serve de excelente instrumento de exploração e análise.

Vejamos alguns argumentos sobre a validade do método da estatística textual, dentro, é claro, dos limites a ele impostos pelo contexto interno e externo à obra. Para Guiraud,

la linguistique statistique est aujourd'hui une branche importante de l'étude du langage, c'est une science qui, convenablement utilisée, peut apporter de nombreuses réponses à la critique historique et philologique des textes¹⁴⁷.

Desde que dirigida a determinados aspectos da obra,

Problèmes purement formels traités indépendamment du contenu et du sens, et qui trouvent dans l'analyse quantitative leur critères les plus efficaces. La statistique, en effet, est la science des écarts et la spécificité de l'oeuvre est entre autre chose un écart par rapport à une norme qu'il reste, évidemment, à définir¹⁴⁸.

¹⁴⁵ MACIEL, Carlos Alberto Antunes. *Richesse et evolution du vocabulaire d'Érico Veríssimo (1905-1975 – Porto Alegre, Brésil)*. Paris: Champion; Genève: Slaktine, 1986, p. 24-25.

¹⁴⁶ Sobre “Um cão de lata ao rabo”, conto em que Machado brinca com três estilos diferentes de escrever, Sá Rego observa como Machado tinha “uma concepção da arte literária como forma, e está claramente baseado na idéia central de parodiar estilos narrativos prototípicos”. SÁ REGO, op. cit., p. 148.

¹⁴⁷ GUIRAUD, Pierre. *Essais de stylistique*, p. 16.

¹⁴⁸ GUIRAUD, Pierre. *Essais de stylistique*, p. 16.

O grande perigo é o de se perder na montanha de dados numéricos e perder a dimensão do todo da obra. Sobre isso, Guiraud, além de alertar para o problema, aponta um caminho para superá-lo:

L'étude des écarts stylistiques, et en particulier l'étude statistique de ces écarts, tend, en effet, à réduire l'oeuvre à une somme de traits spécifiques pris em eux-mêmes et séparément. Mais à partir du moment où le linguiste conçoit la langue d'une oeuvre comme un code particulier, il y voit non un simple inventaire de formes (le lexique, la grammaire ou la rhétorique de l'auteur) mais un système de valeurs, dans lesquels les signes fonctionnent em opposition et tirent leurs sens de leurs relations réciproques au sein de l'ensemble.

Alors, il apparaît que les grandes oeuvres ont non seulement un langage particulier, mais que ce langage est organisé¹⁴⁹.

Aparece aqui a principal questão sobre a qual é preciso refletir, antes de tudo, para viabilizar a abordagem estatística de um texto e torná-la válida: quais critérios adotar para definir a norma? Em nosso caso específico, o ponto de partida foi o autor, mas, mesmo no universo de um mesmo autor, diferenças se impõem de acordo com o gênero literário¹⁵⁰ e com a época. Machado de Assis escreveu poesia, crônica, crítica, teatro, romance e conto. Obviamente há diferenças impostas pela natureza dos textos. (Como já é sabido, diante de tal universo, optamos pelo conto, mas de nossa opção decorreram mais algumas questões a resolver, como a cronologia e o problema das fases colocado pela tradição crítica.)

Esses critérios não são aleatórios, eles se baseiam, antes de tudo, na ideia de que um estudo comparativo só pode ser válido se compara objetos de natureza semelhante. Como afirma Guiraud,

Il est clair, en effet, que pour chaque individu, chaque milieu, chaque époque, chaque genre il exist des mots, des figures, des constructions privilégiées, comme les métaphores de Victor Hugo ou les imparfaits de Flaubert; et que non moins significatives sont les formes absentes ou anormalement rares, telle la réduction du vocabulaire et l'absence des termes concrets dans la tragédie classique.

Que le roman traditionnel soit généralement écrit au passé et la poésie lyrique le plus souvent au présent, cela constitue un fait d'observation; et la conformité à cette norme ou, au contraire, tout ce qui s'en écarte, peut être considéré comme significatif et répondant à une conception et une fonction particulière de l'oeuvre¹⁵¹.

¹⁴⁹ GUIRAUD, Pierre. *Essais de stylistique*, p. 16.

¹⁵⁰ Sobre a importância do gênero no estudo do estilo, ver RASTIER, François. *Arts et sciences du texte*. Paris: PUF, 2001, p. 178-179.

¹⁵¹ GUIRAUD, Pierre. *Essais de stylistique*, p. 62.

Outro autor que ratifica a necessidade de homogeneidade nos critérios é o pioneiro da área, Charles Muller:

Suivant que l'on considère deux textes (ou fragments, ou échantillons...) très proches par leur sujet et leur style, et composés à des dates rapprochées par un même auteur, ou au contraire deux textes divers par l'auteur, le thème et le style, on conçoit que la comparaison de leurs vocabulaires donnera des résultats très différents ; à l'intérieur d'une même oeuvre, une comparaison semblable peut être instituée pour deux fragments. Dès l'instant où des oeuvres littéraires sont indexées, il y a là une mine pour les recherches comparatives¹⁵².

Por isso é necessário ter em mente que, se por um lado, a estatística traz uma segurança ao pesquisador, por outro, como afirma Muller, “elle doit surtout leur imposer ses principes de rigueur, d'attitude critique à l'égard des données numériques et d'analyse minutieuse et progressive des résultats”¹⁵³.

Se um dos motivos que levam pesquisadores da área de Humanas a adotar e adaptar métodos de outras áreas como a estatística é a busca de determinado rigor, como o apontado acima por Muller, é preciso atenção e humildade para não confundi-lo com a “falácia da objetividade”, como dizem Leech e Short a respeito da quantificação¹⁵⁴. Nosso objetivo é nos afastar ao máximo dos impressionismos, mas não temos ilusão quanto ao caráter subjetivo – e, esperamos, intersubjetivo – do método que adotamos.

Se o exame da linguagem no texto literário pode ser um meio para entender a obra de modo mais completo, como afirmam os autores¹⁵⁵, um instrumento capaz de permitir o acesso a informações sobre a totalidade de uma obra ou a um grupo de obras, em termos dos elementos linguísticos que a constituem, não pode ser desprezado. Para fins comparativos, pois, não só a estatística textual auxiliada por computador é capaz de mapear quase que totalmente o material como também, em termos de contagem, é mais rápida e precisa do que se o mesmo processo fosse feito à mão.

Embora com outro sentido, para nós esse é mais um instrumento que pode transformar o estudo do estilo numa “aventura da descoberta”¹⁵⁶. Como bem afirma Muller, as ferramentas da estatística textual “permettent d'arracher aux textes des secrets qu'aucune lecture ne

¹⁵² MULLER, Charles. *Initiation à la statistique linguistique*. Paris: Larousse, 1968, p. 210.

¹⁵³ MULLER, op. cit., p. 209.

¹⁵⁴ LEECH; SHORT, op. cit., p. 47.

¹⁵⁵ “[...] examining the language of a literary text can be a means to a fuller understanding and appreciation of the writer's artistic achievement”. LEECH; SHORT, op. cit., p. 1.

¹⁵⁶ LEECH; SHORT, op. cit., p. 6.

pouvait déceler, et d'ajouter aux connaissances traditionnelles du linguiste et du littéraire des données nouvelles"¹⁵⁷.

O importante é ter em mente, antes de tudo, o princípio de Guiraud de jamais confundir quantitativo e qualitativo, “qui sont certes complémentaires et peuvent souvent converger vers des conclusion similaires, mais constituent deux moments de l'analyse que doivent être distincts”.

3.2 A Ferramenta Hyperbase

O programa Hyperbase para análise textual de *corpus* de grande extensão teve sua primeira versão disponível em 1997. Embora possa ser usado para o estudo de textos de qualquer natureza, a maior vantagem desse programa em relação aos demais é que ele foi concebido tendo como objeto específico o texto literário, enquanto a maioria dos programas de estatística textual é voltada para a chamada estatística linguística. Seu idealizador, Etienne Brunet, é Ph.D. em Literatura mas, para poder concretizar o projeto de programa que tinha em mente, tornou-se estatístico e programador autodidata. Cada função, seja documental ou estatística, colocada no Hyperbase foi pensada a partir dos problemas específicos da linguagem literária ou da pesquisa nessa área.

O primeiro passo para quem quer usar um programa de estatística textual é buscar fontes seguras de onde possa retirar uma versão eletrônica que seja fiel a sua matriz impressa. Em nosso caso optamos por duas fontes. A primeira é o sítio de nosso núcleo de pesquisa, o Nupill, que digitalizou e publicou gratuitamente na internet a versão eletrônica das *Obras completas de Machado de Assis*, publicadas pela Nova Aguilar. Como esses volumes, na realidade, não contêm toda a obra de Machado, tivemos que buscar outra fonte para completar os contos que faltavam àquela publicação – o sítio de Cláudio Weber Abramo, que se utilizou de várias fontes secundárias para poder reunir todos os contos do autor.

Embora ambas as fontes sejam confiáveis, coube à pesquisadora fazer uma revisão das versões eletrônicas a fim de verificar se restou algum problema de escaneamento que tenha escapado à revisão dos próprios sítios. Foi o início de uma longa e árdua tarefa: a preparação

¹⁵⁷ MULLER, Charles. Des participes, de leurs accords, et des inquiétudes qu'ils occasionnent. In: MELLET, Sylvie et alii. *Mots chiffrées et déchiffrées* : mélanges offerts à Etienne Brunet. Paris: Honoré Champion, 1998, p. 288.

dos textos para que pudessem passar pelo tratamento inicial feito pelo Hyperbase. A primeira providência tomada foi a transformação dos arquivos. Em geral eles vêm das fontes (no nosso caso, dos sítios do Nupill e de Cláudio W. Abramo) em formatos específicos para sua colocação na rede, e o programa só consegue ler arquivos em formato *somente texto*. Feitas as devidas revisões e uma vez limpo de qualquer formatação ou edição, o *corpus* ficou pronto para ser submetido ao tratamento inicial do Hyperbase.

Na fase inicial de criação da base de dados, o programa segue uma série de 12 etapas, a fim de extrair do *corpus* toda a informação necessária para adequá-lo à exploração de todas as funções disponíveis¹⁵⁸.

3.2.1 Conceitos básicos

Inicialmente, apresentamos alguns conceitos básicos utilizados na estatística textual. Como o computador trabalha com séries de códigos numéricos, ele não é capaz de diferenciar uma palavra ou sinal gráfico de outra sequência qualquer de caracteres. Para que o programa “leia” um texto, é necessário que este passe por um processo de codificação. Quando um texto é escaneado, transformado num arquivo eletrônico e gravado no disco de um computador, ele já passou pela primeira fase dessa codificação. O momento seguinte é o tratamento feito pelo programa de análise textual, que vai transformar os textos que formam nosso *corpus* em uma base de dados textuais. Assim, o *corpus* está preparado para que possamos tirar todas as informações dele, a partir das funções do programa.

Para separar uma palavra de outra, ou uma *forma gráfica* de outra, o usuário deve lançar mão de alguns caracteres delimitadores, que serão pedidos pelo programa na fase inicial de tratamento do texto. Daí o conceito de *forma*. Forma é tudo o que aparece como unidade: pode ser um sinal de pontuação, pode ser uma palavra, seja ela com ou sem carga semântica (preposições, artigos, etc.). A maioria das palavras, principalmente em *corpus* de grande extensão, pode se repetir em um texto. O número de vezes que tal palavra se repete é sua *frequência*. Se ela aparece dez vezes no *corpus*, ela é uma palavra de frequência 10. A cada frequência, obviamente, corresponde o contexto de sua aparição; a este contexto em que podemos localizá-la chamamos concordância. Quando uma palavra é de frequência 1, ou seja,

¹⁵⁸ Ver BRUNET, op. cit.

só aparece uma vez em todo o *corpus*, ela ganha uma denominação especial: ela é um hápax, outro conceito importante para nosso programa. O número de hápax em um *corpus* pode denotar a riqueza ou a pobreza de seu vocabulário.

O Hyperbase tem suas funções divididas em dois tipos: as de natureza documental e as estatísticas. As funções documentais são as listas de palavras contidas no *corpus* ou dicionários que contêm todas as palavras que nele aparecem, seja por ordem de frequência, seja por ordem alfabética. Há também o *index* das palavras mais frequentes; em geral, a lista padrão contém as 100 primeiras formas mais repetidas, aí incluídos os sinais de pontuação. Esse *index* também pode ser visualizado por ordem decrescente de frequência ou por ordem alfabética.

Ainda nas funções documentais, podemos fazer a leitura de qualquer texto do *corpus* em sua íntegra, ou ainda ler as partes que o programa seleciona a partir de critérios de probabilidade estatística como as mais significativas de cada parte do *corpus*. Outra função documental importante é a listagem de concordâncias. Basta selecionar uma palavra ou um lema (um radical de verbo, por exemplo), que podemos ter informação sobre o número de vezes que a forma ou o lema aparece no total do *corpus* e em cada parte, bem como acessar cada local em que eles ocorrem, ou seja, ver as concordâncias.

Quanto às funções estatísticas, elas permitem a produção de gráficos a partir dos números relativos dos principais dados documentais, como análise em árvore e os quadros resultantes de análise fatorial.

3.3 O *Corpus*

O nosso *corpus* de trabalho é composto de todos os 205 contos escritos por Machado de Assis, de 1858 a 1906¹⁵⁹. Os contos foram processados por um programa de computador especialmente criado para tratar grandes grupos de textos. Além de compor um dicionário de todas as palavras, separando-as por formas (os vocábulos) e ocorrências (os vocábulos

¹⁵⁹ Quanto à data de produção do último conto, há discordâncias entre alguns autores. Gledson e Cunha datam *O escritor Coimbra* de 1907, enquanto Cláudio Weber Abramo afirma que o conto foi publicado originalmente no Almanaque Brasileiro Garnier, em 1906. Como a fonte de nossos contos, juntamente com o sítio do Nupill, foi o sítio de Weber Abramo, optamos por adotar a cronologia deste. GLEDSON, op. cit., p. 35; CUNHA, op. cit., p. 205.

repetidos com suas respectivas frequências), o programa dispõe de funções estatísticas capazes de relativizar os valores brutos, a fim de tornar possível a comparação entre textos de diferente extensão.

Uma das etapas de nosso trabalho foi a de submeter o *corpus* ao tratamento estatístico feito pelo computador. Com os dados em mãos, buscamos verificar a existência de padrões linguísticos, comuns ou exclusivos de determinada parte do *corpus*. Passamos, então, à análise, a fim de descobrir se os primeiros constituíam uma marca permanente que perpassasse o todo dos contos de modo significativo, e que tipo de explicação (cronológica, temática, etc.) seria possível para os últimos. A partir do que foi possível verificar quanto à natureza e à organização do material linguístico que compõe a contística machadiana, pretendia-se constatar se o que de fato predomina, de uma fase a outra, é a continuidade, o processo, ou a ruptura.

Os critérios que nortearam nossas opções quanto à divisão do *corpus* foram: em primeiro lugar o gênero literário, a cronologia, a seleção de contos feita pelo próprio autor (para publicação em livro) e a divisão em fases feita pela crítica. O resultado foi a separação dos contos avulsos (não publicados em livro), de acordo com a cronologia da publicação dos livros de contos. As datas-chave que marcam a transição foram então: 1882, para os contos, ano da publicação de *Papéis avulsos*; e 1881, data de publicação de *Memórias póstumas de Brás Cubas*, para demarcar a transição nos romances.

Embora este trabalho analise todos os contos de Machado de Assis, para atingirmos nosso objetivo de mapear os padrões do estilo maduro, vários recortes tiveram que ser feitos. O principal é de viés qualitativo. Como muitos contos foram deixados na gaveta pelo próprio autor e foram publicados postumamente, preferimos focar nossa pesquisa apenas nas obras publicadas em livro por Machado, durante sua vida, com seu aval. Por isso, o foco central de nosso trabalho se dirige a esses sete volumes.

Como nossa intenção era verificar se as mudanças estilísticas do conto machadiano devem-se mais a uma ruptura ou a um processo contínuo de amadurecimento, nos valem dos marcos de transição de fases apontados pela tradição crítica, que estabelece como paradigmas da qualidade estilística do conto machadiano os textos contidos nos cinco volumes publicados a partir de *Papéis avulsos* (1882) – aí incluídos *Histórias sem data*, de 1884, *Várias histórias*, de 1896, *Páginas recolhidas*, de 1899, e *Relíquias de casa velha*, de

1906¹⁶⁰. Então, tomamos como parâmetro oficial da primeira fase da contística machadiana os dois primeiros volumes de contos publicados pelo autor: *Contos fluminenses*, de 1870, e *Histórias da meia-noite*, de 1873. Além destes, há ainda 129 contos entre publicados em periódicos e não-publicados. A todos eles chamaremos *avulsos*. Estes também foram submetidos ao critério cronológico da divisão em fases. A primeira começa em 1858, ano em que o autor escreve seu primeiro exemplar do gênero, e vai até 1880, contemplando o que é considerada a fase inicial do autor. A segunda começa em 1881 e vai até 1906, data de produção de seu último conto. O segundo grupo começa por essa data porque é o ano da publicação de *Memórias póstumas de Brás Cubas*, romance que marca a transição estilística no romance, que igualmente é prosa de ficção, e também porque *Papéis avulsos*, que marca a transição no conto, é publicado apenas um ano depois.

O segundo momento, quantitativo, subdivide-se em duas partes: na primeira, o conjunto dos cinco volumes publicados forma um *subcorpus* que teve suas características analisadas em conjunto; na segunda, cada um dos cinco volumes de contos escolhidos foi tratado individualmente, comparando-se um a um com a base composta pelos cinco volumes, a fim de verificar se os elementos estilísticos destacados do primeiro conjunto (macro) são válidos no conjunto dos contos considerados isoladamente (micro). Para garantir que os dados extraídos dessa análise fossem significativos, usamos critérios próprios da estatística textual, itens que são explicados mais adiante.

Como o programa Hyperbase só pode criar bases de, no máximo, 76 unidades de texto por vez, independentemente da extensão deles, foi preciso dividir o *corpus* em várias bases. A divisão forma duas grandes partes: uma desde o primeiro conto, de 1858 até 1880; e outra de 1881 até o último conto do autor, que data de 1906. Esses dois *subcorpora* sofreram mais uma divisão interna, cujo critério é o da publicação: os que não foram publicados em livro em vida pelo autor chamamos *Avulsos*; e os outros foram agrupados em duas grandes bases correspondentes aos volumes publicados em vida, uma com os dois volumes anteriores a *Papéis avulsos* e a outra com os cinco volumes restantes.

Como os sete volumes publicados são nosso material principal, cada volume também formou uma base individual, a fim de verificar se os dados dos conjuntos (formados pelos dois primeiros volumes e outro pelos cinco últimos), divididos cronologicamente, eram compatíveis com os dados dos volumes individuais.

¹⁶⁰ Na seleção, tanto para a edição de antologias bem como para estudos críticos e teóricos, e na cronologia de produção da obra machadiana –, optamos por privilegiar aquela que é considerada (por essa mesma tradição) a fase madura do autor, o auge de sua escrita, de seu estilo, no conto.

Uma vez tratados pelo Hyperbase, os grupos de textos transformam-se em bases de dados textuais, por isso chamamos as subdivisões de nosso *corpus* de bases. Temos as seguintes bases já constituídas, a partir dos critérios expostos anteriormente:

a) publicadas em livro:

- i. uma base formada por todos os volumes de contos publicados em vida pelo autor (sete volumes) (identificada pela sigla 7Vol);
- ii. uma base formada pelos dois primeiros volumes de contos publicados em vida pelo autor (5Vol);
- iii. uma base formada pelo conjunto dos outros cinco volumes de contos publicados em forma de livro durante a vida do autor (63 contos) (FluMeia); e
- iv. uma base para cada volume de conto publicado individualmente, totalizando sete bases (PA, PR, HMN, RCV, HSD, VH, CFLU).

b) avulsos:

- i. uma base de contos avulsos (não publicados em forma de livro) escritos até 1880, constituída por 76 contos (AVG1); e
- ii. uma base de contos avulsos pós-1881, constituída por 53 contos (AVG2).

Os dados extraídos da análise exploratória das bases formadas por esses cinco volumes e por cada um deles em separado são comparados entre si e com três grupos de controle: o conjunto dos demais contos escritos por Machado; os contos dos dois volumes publicados antes de *Papéis avulsos*; e a base de dados Portext¹⁶¹. Utilizamos a base Portext para comparar as características do conto de Machado com as de outros textos literários da época. A base contém textos em língua portuguesa, de várias áreas de conhecimento, escritos nos séculos XIX e XX.

3.3.1 Características do *corpus*

A fim de viabilizarmos a visualização dos dados, principalmente nos gráficos, foi necessária a codificação de todos os contos que compõem o *corpus*, bem como dos grupos ou bases. O Quadro 1 apresenta a lista de todos os contos e seus respectivos códigos, aqui separados por volumes publicados e avulsos.

¹⁶¹ Desenvolvida pelo Prof. Dr. Carlos Maciel, da Universidade Sophia Antipolis, de Nice (França), e pesquisador do mesmo laboratório que desenvolveu o Hyperbase.

Volume/Conto	Base/Código
CONTOS FLUMINENSES (1870)	CFLU
1) Miss Dollar	MiDo
2) Luís Soares	LuSo
3) A mulher de preto	MPre
4) O segredo de Augusta	SAug
5) Confissões de uma viúva moça	CoVM
6) Linha reta e linha curva	LRLC
7) Frei Simão	FrSi
HISTÓRIAS DA MEIA-NOITE (1873)	HMN
1) A parasita azul	PAzu
2) As bodas de Luís Duarte	BoLD
3) Ernesto de Tal	ETal
4) Aurora sem dia	AuSD
5) O relógio de ouro	ReOu
6) Ponto de vista	PVis
PAPÉIS AVULSOS (1882)	PA
1) O alienista	Alie
2) Teoria do medalhão	TMed
3) A chinela turca	ChTu
4) Na arca	Arca
5) D. Benedita	Bene
6) O segredo do bonzo	SBon
7) O anel de Polícrates	APol
8) O empréstimo	OEmp
9) A sereníssima República	Ssma
10) O espelho	OEsp
11) Uma visita de Alcibiades	VALc
12) Verba testamentária	VTes
HISTÓRIAS SEM DATA (1884)	HSD
1) A igreja do Diabo	IgrD
2) O lapso	OLap
3) Último capítulo	UCap
4) Cantiga de esponsais	CEsp
5) Singular ocorrência	Sing
6) Galeria póstuma	GaPo
7) Capítulo dos chapéus	CapC
8) Conto alexandrino	CAle
9) Primas de Sapucaia!	PrSa
10) Uma senhora	USra
11) Anedota pecuniária	AneP
12) Fulano	Fula
13) A segunda vida	SegV
14) Noite de almirante	NAlm
15) Manuscrito de um sacristão	MSac
16) <i>Ex cathedra</i>	ExCa
17) A senhora do Galvão	SGal

18) As academias de Sião	Acad
VÁRIAS HISTÓRIAS (1896)	VH
1) A cartomante	Cart
2) Entre santos	ESan
3) Uns braços	UBra
4) Um homem célebre	HCel
5) A desejada das gentes	DesG
6) A causa secreta	CauS
7) Trio em lá menor	Trio
8) Adão e Eva	Adao
9) O enfermeiro	OEnf
10) O diplomático	ODip
11) Mariana	Marl
12) Conto de escola	CoEs
13) Um apólogo	UApo
14) D. Paula	DnPa
15) Viver!	Vivr
16) O cônego ou metafísica do estilo	OCon
PÁGINAS RECOLHIDAS (1899)	PR
1) O caso da vara	Vara
2) O dicionário	ODic
3) Um erradio	Erra
4) Eterno!	Eter
5) Missa do galo	Galo
6) Ideias de canário	IdCa
7) Lágrimas de Xerxes	LaXe
8) Papéis velhos	PaVe
RELÍQUIAS DE CASA VELHA (1906)	RCV
1) Pai contra mãe	PaiM
2) Maria Cora	Cora
3) Marcha fúnebre	MFun
4) Um capitão de voluntários	CVol
5) Suje-se gordo!	SuGo
6) Umas férias	UFer
7) Evolução	Evol
8) Pílates e Orestes	PiOr
9) Anedota do cabriolet	ACab
AVULSOS 1 (Contos de 1858 a 1881)	AVG1
1858 Três tesouros perdidos	3TPe
1862 O país das Quimeras	PaQu
1864 Virginius	Virg
1864 Casada e viúva	CaVi
1864 O anjo das donzelas	AnDo
1864 Questão de vaidade	QVai
1865 Cinco mulheres	5Mul
1866 Uma excursão milagrosa	UEMi

1866	Diana	Dian
1866	Felicidade pelo casamento	FeCa
1866	Fernando e Fernanda	FeFe
1866	O oráculo	Orac
1866	O pai	OPai
1866	A pianista	APia
1866	O que são as moças	QSMo
1867	Francisca	Fran
1867	História de uma lágrima	HsLa
1867	Onda	Onda
1867	Possível e impossível	PImp
1867	O último dia de um poeta	UPoe
1868	O carro nº 13	Ca13
1868	Não é mel para a boca do asno	Asno
1869	O anjo Rafael	Rafa
1870	O Capitão Mendonça	CpMe
1870	O rei dos caiporas	ReiC
1870	A vida eterna	ViEt
1871	Almas agradecidas	AlAg
1871	Ayres e Vergueiro	AyVe
1871	O caminho de Damasco	CaDa
1871	Mariana	Mari
1872	Uma águia sem asas	Agui
1872	Uma loureira	Lour
1872	Qual dos dois?	Qud2
1872	Quem não quer ser lobo...	Lobo
1872	Ruy de Leão	Leao
1873	Decadência de dois grandes homens	D2GH
1873	Um homem superior	HoSu
1873	Nem uma nem outra	NemU
1873	Quem conta um conto...	QCon
1873	Tempo de crise	TeCr
1874	Miloca	Milo
1874	Um dia de estruendo	DEnt
1874	Muitos anos depois	MuAD
1874	Os óculos de Pedro Antão	Anta
1874	Valério	Vale
1875	Antes que cases...	AQCa
1875	Brincar com fogo	Fogo
1875	A mágoa do infeliz Cosme	Cosm
1875	Onze anos depois	Onze
1875	A última receita	UIRe
1875	Um esqueleto	Esqu
1875	O sainete	OSai
1875	Casa, não casa	Casa
1876	Longe dos olhos...	Long
1876	O astrólogo	OAst
1876	D. Mônica	DMca
1876	Encher tempo	EnTe
1876	O passado, passado	Pass
1876	Sem olhos	SemO
1876	<i>To be or not to be</i>	ToBe
1876	História de uma fita azul	Fita

1877	Um almoço	Almo
1877	Um ambicioso	Ambi
1877	A melhor das noivas	AMNo
1877	Silvestre	Silv
1878	Antes da missa	AnMi
1878	O califa de platina	CPla
1878	Um cão de lata ao rabo	Lata
1878	Conversão de um avaro	Conv
1878	Dívida extinta	DivE
1878	Elogio da vaidade	EVai
1878	Filosofia de um par de botas	Filo
1878	A herança	AHer
1878	O machete	OMac
1878	Folha rota	Folh
1880	A chave	Chav
AVULSOS 2 (Contos de 1881 a 1906)		AVG2
1881	O caso da viúva	CViu
1881	A mulher pálida	MuPa
1882	O imortal	Imor
1882	Letra vencida	LVen
1883	O programa	OPro
1883	A ideia de Ezequiel Maia	Maia
1883	Médico é remédio	MeRe
1883	História comum	HsCo
1883	O destinado	Odes
1883	Troca de datas	TrDa
1883	Questões de maridos	QuMa
1883	Três conseqüências	3Con
1883	Vidros quebrados	ViQu
1883	Cantiga velha	CaVe
1883	Metafísica das rosas	Rosa
1884	Trina e una	TUna
1884	O contrato	OCnt
1884	A carteira	ACar
1884	O melhor remédio	OMRe
1884	A viúva Sobral	VSob
1884	Entre duas datas	E2Da
1884	Vinte anos! Vinte anos!	ViAn
1884	O caso do Romualdo	CaRo
1884	Uma carta	UCar
1885	Só!	Sooo
1885	Habilidoso	Habi
1885	Viagem à roda de mim mesmo	Roda
1886	Casa velha	CVel
1886	Curta história	CHis
1886	Um dístico	UDis
1886	Astúcias de marido	AsMa
1886	Pobre cardeal!	Poça
1886	Terpsicore	Terp
1887	Antes a rocha Tarpeia	ARTa
1887	Identidade	Iden
1887	Sales	Sale

1889	D. Jucunda	DJuc
1890	Como se inventaram os almanaques	Alma
1891	Pobre Finoca!	PoFi
1892	O caso Barreto	CaBa
1892	Uma partida	UMAP
1892	Um sonho e outro sonho	Sonh
1893	Um quarto de século	QuSe
1893	Vênus, divina Vênus	Venu
1894	A inglesinha Barcelos	IngB
1894	João Fernandes	João
1895	Uma noite	Noit
1895	Orai por ele!	PEle
1896	Um incêndio	UInc
1897	Uma por outra	UmPO
1897	Flor anônima	Flor
1904	Jogo do bicho	Jogo
1906	O escrivão Coimbra	ECoí

Quadro 1 – Códigos dos contos de Machado de Assis por volume e bases integrantes do *corpus*

A extensão dos contos, para fins de análise quantitativa, se dá em termos de *formas* (V), que são as palavras diferentes ou vocábulos que constituem o texto ou grupo de textos analisados; e de *ocorrências* (N) que são as palavras contadas em todas as suas repetições, ou seja, acrescidas de sua frequência. Hápx são as palavras de frequência 1, isto é, as que aparecem apenas uma vez no grupo ou subgrupo respectivo.

A extensão total de nosso *corpus*, contando-se os 205 contos, chega a 1.147.244 ocorrências. Essas ocorrências estão distribuídas nas Tabelas 1 e 2.

Tabela 1 – Indicadores de extensão do *corpus* em todos os contos de Machado de Assis

Grupos	N	N'/N	V	<i>f</i>	V ₁	P ₁
Total	1.147.244	1,0000	68.680	16,7041	32.652	0,4754
FLUMEIA	119.545	0,1042	10.781	11,0884	5.687	0,5275
AVG1	547.611	0,4773	23.591	23,2127	10.291	0,4362
5VOL	249.300	0,2173	18.703	13,3294	9.015	0,4820
AVG2	230.788	0,2011	15.605	14,7893	7.659	0,4908

N = número de ocorrências

N'/N = Número de ocorrências de cada subgrupo, dividido pelo total de ocorrências

V = Número de vocábulos

f = Frequência média das palavras, resultado da divisão N/V

V₁ = Número de palavras de frequência 1 ou hápx.

P₁ = Proporção de palavras de frequência 1, por subgrupo.

Tabela 2 – Indicadores de extensão do *corpus* nos cinco volumes de contos publicados em livro por Machado de Assis

Volumes	N	N'/N	V	f	V ₁	P ₁
Total	369.119	1,0000	22.678	16,2765	10.639	0,4691
CFLU	69.439	0,1881	10.781	6,4408	4.108	0,3810
HMN	50.106	0,1357	6.704	7,4740	3.833	0,5717
PA	65.716	0,1780	9.166	7,1695	5.274	0,5753
HSD	63.233	0,1713	8.516	7,4819	4.898	0,5751
VH	55.570	0,1505	7.399	7,5104	4.274	0,5776
PR	28.186	0,0763	4.797	5,8757	2.848	0,5937
RCV	36.595	0,0914	5.292	6,9151	2.997	0,5663

N = número de ocorrências

N'/N = Número de ocorrências de cada subgrupo, dividido pelo total de ocorrências

V = Número de vocábulos

f = Frequência média das palavras, resultado da divisão N/V

V₁ = Número de palavras de frequência 1 ou hápax.

P₁ = Proporção de palavras de frequência 1, por subgrupo.

Como grupo de controle, adotamos a base Portext¹⁶², formada por 76 textos clássicos da língua portuguesa, entre poemas, cartas, peças de teatro, contos, romances, livros de memórias, crônicas e outros. O Quadro 2 apresenta a lista de textos, o período de produção, os autores e seus respectivos códigos e gêneros.

Obra (ano) – Autor	Código (Gênero*)
1) Obra poética – Gregório de Matos	GrMA (P)
2) Sermões (1679) – Padre Antônio Vieira	SeAV
3) Caramuru (1781) – Santa Rita Durão	Cara (P)
4) O Uraguai (1769) – Basílio da Gama	Urag (P)
5) Poemas escolhidos	PoEs (P)
6) Cartas chilenas – Tomás Antônio Gonzaga	CaCh (P)
7) Memórias e cotidiano do Rio de Janeiro no tempo do rei (1811-1821) – Luís Joaquim dos Santos Marrocos	MeRJ (Ca)
8) Macário (1852) – Álvares de Azevedo	Macá (T)
9) Noite na taverna (1855) – Álvares de Azevedo	NoTa (C)
10) Poemas – Álvares de Azevedo	PoAA (P)
11) Poemas – Fagundes Varela	FaVa (P)
12) A moreninha (1843) – Joaquim Manuel de Macedo	MORE (R)
13) O moço loiro (1845) – Joaquim Manuel de Macedo	MoLo (R)
14) Luneta mágica (1869) – Joaquim Manuel de Macedo	LuMá (R)
15) Dirceu de Marília (1845) – Joaquim Norberto de Sousa e Silva	DiMA (P)
16) Poesias completas	LaRi (P)
17) Suspiros poéticos e saudades (1836) – Gonçalves de Magalhães	SuPo (P)
18) Memórias de um sargento de milícias (1853) – Manuel Antônio de Almeida	SaMi (R)
19) A viúvinha (1857) – José de Alencar	Aviu (R)
20) Cinco minutos (1856) – José de Alencar	CMJA (R)

¹⁶² MACIEL, Carlos Alberto Antunes. Da base Portext ao CD-ROM de textos jurídicos. In: SARDINHA, Tony Berber (Org.). *A língua portuguesa no computador*. Campinas, SP: Mercado das Letras; São Paulo: Fapesp, 2005, p. 185-195.

21) Diva. Perfil de mulher (1864) – José de Alencar	DIVA (R)
22) Iracema (1865) – José de Alencar	IRAC (R)
23) A pata da gazela (1870) – José de Alencar	PATA (R)
24) Lucíola (1862) – José de Alencar	Lucí (R)
25) Senhora (1875) – José de Alencar	SENH (R)
26) Encarnação (1877) – José de Alencar	Enca (R)
27) Espumas flutuantes (1870) – Castro Alves	CAAI (P)
28) Miss Dollar – Machado de Assis	CoFl (C)
29) A mão e a luva (1874) – Machado de Assis	LUVA (R)
30) Esaú e Jacó (1904) – Machado de Assis	JACÓ (R)
31) Iaiá Garcia (1878) – Machado de Assis	IAIÁ (R)
32) Memórias póstumas de Brás Cubas (1881) – Machado de Assis	BRÁS (R)
33) Dom Casmurro (1899) – Machado de Assis	DOMC (R)
34) Memorial de Aires (1908) – Machado de Assis	AIRE (R)
35) Americanas – Machado de Assis	AMER (P)
36) Casa velha (1886) – Machado de Assis	CaVe (C)
37) Crisálidas – Machado de Assis	CRIS (P)
38) Ocidentais – Machado de Assis	Ocid (P)
39) Falenas – Machado de Assis	FALE (P)
40) O alienista – Machado de Assis	ALIE (C)
41) O Cabeleira (1876) – Franklin Távora	Cabe (R)
42) Inocência (1872) – Visconde de Taunay	INOC (R)
43) Minha formação (1898) – Joaquim Nabuco	MiFo (M)
44) O ermitão de Muquém (1858 ou 1864?) – Bernardo Guimarães	ERMI (R)
45) O ateneu (1888) – Raul Pompeia	ATEN (R)
46) O mulato (1881) – Aluísio de Azevedo	Mula (R)
47) Girândola de amores (1882) – Aluísio de Azevedo	GiAm (R)
48) Filomena Borges (1884) – Aluísio de Azevedo	FiBo (R)
49) Casa de pensão (1883) – Aluísio de Azevedo	CaPe (R)
50) O cortiço (1890) – Aluísio de Azevedo	CORT (R)
51) Livro de uma sogra (1895) – Aluísio de Azevedo.	LISO (R)
52) A mortalha de Alzira (1891) – Aluísio de Azevedo	MoAl (R)
53) O escravocrata (1884) – Artur Azevedo	Escr (T)
54) A capital federal (1893) – Coelho Neto	CaFe (R)
55) A normalista (1891) – Adolfo Caminha	NORM (R)
56) Contos para velhos – Olavo Bilac	CoVe (C)
57) O morto (1898) – Coelho Neto	MORT (R)
58) A intrusa (1905) – Júlia Lopes de Almeida	INTR (R)
59) O momento literário (1905) – João do Rio	MoLi (O)
60) À margem da história (1909) – Euclides da Cunha	MaHi (O)
61) Contrastes e confrontos (1907) – Euclides da Cunha	CoCo (Cr)
62) Peru versus Bolívia (1907) – Euclides da Cunha	PeBo (O)
63) Os sertões (1902) – Euclides da Cunha	Sert (R)
64) Obras seletas 6 (1889) – Rui Barbosa	ObS6 (O)
65) Obras seletas 7 (1883)	ObS7 (O)
66) Obras seletas 8 (1889)	ObS8 (O)
67) O noviço (1845) – Martins Pena (publicado em 1853)	Novi (T)
68) A poesia interminável – Cruz e Sousa	PoIn (P)
69) Recordações do escrivão Isaías Caminha (1907) – Lima Barreto	ISAÍ (R)
70) O triste fim de Policarpo Quaresma (1911) – Lima Barreto	PoQu (R)
71) Os Bruzundangas (1923) – Lima Barreto	BRUZ (Cr)
72) A nova Califórnia (1910) – Lima Barreto	NoCA (C)
73) Clara dos Anjos (1904) – Lima Barreto	CIAn (R)

74) O cemitério dos vivos – Lima Barreto	CeVi (M)
75) Diário íntimo – Lima Barreto	DiÍn (M)
76) Contos gauchescos (1912) – Simões Lopes Neto	deSimõesLNeto (C)

* P – poesia Ca – cartas T – teatro C – conto R – romance M – memórias Cr – crônicas O – outros

Quadro 2 – Lista de textos e autores da base Portext e respectivos códigos e gêneros

3.3.2 Léxico, convenções e dificuldades

A quantificação implica uma série de opções em termos dos elementos a serem analisados, e estes, por sua vez, trazem consigo as dificuldades relativas a sua natureza. De acordo com nossos objetivos e com as possibilidades e limites do programa que adotamos, escolhemos as variáveis do *corpus* a serem descritas e comparadas. Como o trabalho investiga os padrões estilísticos do conto machadiano, a partir da afirmação da crítica de que houve uma ruptura em *Papéis avulsos*, o critério cronológico é de fundamental importância. É ele que determina nossas opções tanto na escolha de variáveis como nas divisões do *corpus* para a respectiva comparação.

Os elementos que escolhemos para análise são os seguintes: a pontuação, o vocabulário e os verbos. A opção foi motivada pelo exame das altas frequências, que mostraram uma predominância das palavras funcionais, além de outros vocábulos, como pronomes, advérbios e conjunções. Dada a extensão do *corpus*, uma análise comparativa que privilegiasse apenas esse tipo de palavra já seria suficiente para uma tese à parte. Portanto, optamos por agrupá-las como *outros*¹⁶³. Como na lista das maiores frequências, depois desse grupo, os que mais se destacam em termos quantitativos são os sinais de pontuação, os verbos e os substantivos, resolvemos priorizá-los.

Um dos principais problemas enfrentados pelo pesquisador que opta pela estatística textual é o da ambiguidade – palavras homógrafas, de classes gramaticais e funções diferentes só podem ser separadas pela análise de seu contexto. Um exemplo simples é o da forma *que*. Com milhares de ocorrências, só a consulta a cada aparição individual e sua análise e contagem em separado seriam capazes de transformar o dado bruto total em diferenças. Num *corpus* de mais de 1 milhão de ocorrências, com a ferramenta de que dispomos atualmente, essa é uma missão impossível.

¹⁶³ Ver MACIEL, Carlos Alberto Antunes. *Richesse et évolution du vocabulaire d'Erico Veríssimo...*, p. 41.

Os substantivos e verbos, devido a suas flexões, também exigem um trabalho manual, a fim de agrupar essas diferentes formas sob um só tipo de verbo ou substantivo. Uma alternativa para solucionar esse problema seria a lematização¹⁶⁴, porém, em nosso caso, ela teria de ser feita à mão, o que seria igualmente inviável¹⁶⁵.

Os nomes próprios foram tratados à parte. No entanto, o critério utilizado pelo programa para definir o que é nome próprio é que comece com letra maiúscula. Logo, foi preciso outro trabalho de edição das listas geradas, a fim de separar as palavras que começam uma frase dos nomes próprios de fato.

Os sinais de pontuação também são vítimas da ambiguidade. Enquanto o ponto, além de sinalizar o fim de frase, serve para as abreviações de todo o tipo, o hífen e o travessão compartilham o mesmo sinal gráfico. Para sabermos quantos se referem a diálogos, a separação de verbo e pronome ou a intercalação de frase, só é possível por meio da adoção do procedimento de consultar as concordâncias individualmente.

De acordo com as categorias em análise, extraímos dados de tipos diferentes, com opções de visualização adequada a cada caso (listas, gráficos em bastão ou em árvore). Em princípio, os dados observados por categoria foram assim organizados:

- a) pontuação: sinais de pontuação forte; pontuação e extensão de frase;
- b) léxico: evolução do vocabulário; riqueza lexical; distribuição das frequências;
- c) verbos; e
- d) altas frequências e hápax.

Os procedimentos específicos relativos a cada categoria escolhida para análise são detalhados em suas respectivas seções. Como a extensão do *corpus* é muito grande e o foco central de nosso trabalho está voltado para os sete volumes publicados, a análise mais aprofundada desses elementos é feita sobre essa base.

3.3.3 Ordem dos textos no Hyperbase

Por alguma razão não explicada no manual de referência¹⁶⁶, o programa Hyperbase alterou a ordem dos textos que compõem cada base – exceto a dos romances machadianos. Por isso, para a interpretação dos gráficos apresentados nos próximos capítulos e no apêndice, consideramos útil indicar, no Quadro 3, a ordem dos textos nas bases determinada pelo programa, por código. Todos os gráficos em barras seguem tal ordem.

¹⁶⁴ Sobre lematização, ver LEBART; SALEM, op. cit., p. 36-37.

¹⁶⁵ Há um programa específico para esse fim que acompanha as novas versões do Hyperbase. A versão que usamos para este trabalho ainda não dispunha desse recurso.

¹⁶⁶ BRUNET, op. cit.

7VOL MiDo LuSo LRLC FrSi CoVM SAug MPre ReOu ETal AuSD BoLD PVis PAzu VALc SBon OEsp OEmp APol Alie Arca Bene Ssma VTes ChTu TMed UCap Sing PrSa OLap NALm MSac GaPo Fula ExCa CAle CapC CEsp Acad AneP SGal SegV IgrD USra UBra HCel Trio OEnf ODip OCon Mar1 ESan DnPa CoEs	Adao DesG CauS Vivr Cart UApo PaVe ODic Vara Galo LaXe IdCa Erra Eter UFer CVol SuGo PiOr PaiM MFun ACab Evol Cora 5VOL VALc SBon Espe OEmp APol Alie Arca Bene Ssma VTes ChTu TMed UCap Sing PrSa OLap NALm MSac GaPo Fula ExCa CAle CapC CEsp Acad AneP SGal SegV IgrD	USra UBra HCel Trio OEnf ODip OCon Mar1 ESan DnPa CoEs Adao DesG Caus Vivr Cart Apol PaVe ODic Vara Galo LaXe IdCa Erra Eter UFer CVol SuGo PiOr PaiM MFun ACab Evol Cora Flumeia MiDo LuSo LRLC FrSi CoVM SAug MPre ReOu ETal AuSD BoLD PVis PAzu	AVG1 AlAg Fogo Rafa AnDo CPla CaDa CpMe Ca13 Lata Águi AnMi AQCa 5Mul D2GH Dian EVai EnTe DEnt FeCa FeFe Filo Fran HsLa Long AMNo MuAD Asno Orac OPai Pass Onda Anta Lour APia QSMo ReiC PImp QuD2 Lobo QVai Leao Silv ViEt UPoe 3TPe UIRe Casa Mari Milo OMac OSai TeCr Esqu	UEMi DMca Virg AyVe CaVi Conv DivE Fita NemU Onze QCon ToBe HoSu AHer Cosm Folh OAst PaQu SemO Almo ambi Vale Chav AVG2 ARTa AsMa Alma CaBa CViu ECoI IngB Joao MeRe Rosa UMAP MuPa PEle PoCa Terp QuSe Sonh TUna Maia VSob CVel DJuc Flor Habi Iden Jogo CaRo OCnt Sale	Sooo UDis Noit UmPO Venu Roda ViQu ACar CHis PoFi UCar Imor OMRe OPro QuMa 3Con TrDa UInc ViAn ODEs HsCo LVen CaVe E2Da Portext GrMA SeAV Cara Urag PoEs CaCh MeRJ Macá NoTa PoAA FaVa MORE MoLo LuMá DiMA LaRi SuPo SaMi Aviu CMJA DIVA IRAC PATA Lucí SENH Enca CAAI CoFI	LUVA JACÓ IAIÁ BRÁS DOMC AIRE AMER CaVe CRIS Ocid FALE ALIE Cabe INOC MiFo ERMI ATEN Mula GiAm FiBo CaPe CORT LISO MoAl Escr CaFe NORM CoVe MORT INTR MoLi MaHi CoCo PeBo Sert ObS6 ObS7 ObS8 Novi PoIn ISAI PoQu BRUZ NoCA CIAn CeVi DiIn deSimõesLNeto
---	---	--	---	---	--	---

Quadro 3 – Ordem dos textos nas bases (por código)

CAPÍTULO 4 AS ALTAS FREQUÊNCIAS

4.1 Altas Frequências de Nomes Próprios, Verbos, Substantivos e Outros

A primeira curiosidade do pesquisador ao quantificar um texto é relativa à natureza das palavras usadas por determinado autor e sua respectiva distribuição: descobrir se são os substantivos, os nomes próprios ou os adjetivos que predominam; se, como leitores, fomos bons observadores, não deixando nossa memória confundir os elementos que se destacam pela quantidade ou pela repetição com os que jamais esquecemos por outras razões, temáticas ou afetivas, com as quais nos identificamos.

Para a análise do léxico de um *corpus*, o ideal seria podermos recensear todo o vocabulário, para depois separar as palavras por classe gramatical; assim teríamos um quadro mais preciso de como o léxico se constitui. Porém, ainda não dispomos de um programa de lematização¹⁶⁷ para língua portuguesa que possa ser utilizado junto com o Hyperbase. Como nosso *corpus* é composto de 205 contos, que totalizam 1.147.244 ocorrências, esse mapeamento, se tiver que ser feito a mão, é definitivamente inviável.

Para obtermos um panorama da natureza das palavras que compõem a contística machadiana, utilizamos as listas referentes às altas frequências, que contêm as 100 formas mais repetidas em cada grupo de texto. Embora nosso foco principal esteja nos sete volumes de contos publicados por Machado de Assis e tenhamos uma base constituída com os 76 textos respectivos, optamos por ampliar um pouco mais essa lista, pois acreditamos que nos restringirmos a 100 formas seria insuficiente para compor uma visão panorâmica do tipo de palavra usada pelo autor num *corpus* tão extenso. Para isso, utilizamos as listas das altas frequências, subtraindo delas os itens relativos aos sinais de pontuação. Temos então 11 listas: quatro relativas aos grupos macro Flumeia, AVG1, 5VOL e AVG2; e sete relativas a cada volume: CFLU, HMN, PA, HSD, VH, PR e RCV¹⁶⁸.

A partir das listas chegamos às categorias mais repetidas, que dividimos em verbos, substantivos, nomes próprios e outros¹⁶⁹. Esta última agrupa as palavras funcionais, os

¹⁶⁷ Ver item 3.3.2 deste trabalho.

¹⁶⁸ A íntegra das listas encontra-se no apêndice.

¹⁶⁹ Seguimos aqui o modelo de repartição usado por Maciel em seu estudo sobre Érico Veríssimo: MACIEL, Carlos Alberto Antunes. *Richesse et évolution du vocabulaire d'Érico Veríssimo...*, p. 41.

advérbios, as conjunções, as interjeições, etc. Optamos por tal divisão em razão do alto grau de ambiguidade dessas palavras, devido às múltiplas funções que podem exercer no texto, o que teria de ser checado caso a caso em milhares de exemplares. Os adjetivos merecem uma explicação à parte. Nas listas, aparecem apenas dois: *verde*, em PA (que, na verdade, faz parte do nome próprio Casa Verde, o hospício de “O alienista”), e *grande*, em HSD, que se distribui praticamente por todos os contos do volume. Ambos os adjetivos se repetem nas listas dos grupos macro a que pertencem. Outra palavra muito repetida nos contos de Machado que dá margem para dúvidas é *só*. Como nos sete volumes ela tem pouco mais de 600 aparições, foi possível verificar nas listas de concordância os respectivos contextos. Como sua função predominante é a de advérbio, ela está classificada como *outros*.

Amostras, em análise de texto, não são os melhores dados que podemos utilizar. No entanto, como em nosso caso é nossa única opção para ter uma ideia da distribuição das categorias gramaticais no *corpus*, preferimos optar por analisar esses dados, mesmo compreendendo-os como parciais. Amparamos nossa decisão na afirmação de Charles Muller de que um pequeno número de palavras forma uma grande parte de todo discurso:

on estime que le 50 unités les plus fréquentes, dans un idiome quelconque, couvrent 50% du texte; en fait, et surtout pour un idiome comme le français où les mots de relation sont indispensables et fréquents, il suffit parfois moins encore. Le fait que les 1.000 unités les plus fréquents couvrent environ 85% du texte a conduit à rechercher pour divers idiomes quelles sont ces 1.000 unités, qui forment une sorte de vocabulaire fondamental¹⁷⁰.

Também amparamos a decisão nos dados utilizados por Carlos Maciel sobre o vocabulário de Érico Veríssimo, tomados do trabalho de John Duncan (*A frequency dictionary of Portuguese words*): “en langue portugaise, les dix premiers vocables couvrent 36,20% des occurrences alors que les 1.000 premiers couvrent 84,57% du total des occurrences recensées”¹⁷¹. Um dos testes que fizemos para verificar se nossa amostra não era por demais insuficiente foi a soma do número de ocorrências que compõem as listas de alta frequência. Tomando CFLU como exemplo, que tem um dos volumes mais limitados (com 69.439 ocorrências) e menor riqueza lexical, se somarmos até a palavra que ocupa o quinquagésimo lugar na lista, já ultrapassamos 50% do total de ocorrências.

Vamos então ao Quadro 4, com as categorias em percentual.

¹⁷⁰ MULLER, op. cit., p. 162.

¹⁷¹ MACIEL, Carlos Alberto Antunes. *Richesse et évolution du vocabulaire d’Erico Verissimo...*, p. 43, nota 3.

Base	Substantivos	Verbos	Nomes próprios	Outros
CFLU	13,04	10,86	8,69	67,39
HMN	11,23	12,35	7,86	68,53
PA	8,88	11,11	2,22	77,77
HSD	13,18	10,98	1,09	75,81
VH	11,11	13,33	0,00	75,55
PR	12,08	12,08	3,29	72,52
RCV	10,98	12,08	5,49	71,42

Quadro 4 – Evolução do uso de substantivos, verbos, outros e nomes próprios nos volumes de contos publicados por Machado de Assis (%) – a partir da lista de 100 maiores frequências

A superioridade numérica da categoria *outros* é evidente. Além de já começar alta, cronologicamente segue em crescimento até pouco antes da metade do período, para logo depois ter uma queda gradual e estabilizar-se. Outro dado interessante é a quantidade de *nomes próprios* nas altas frequências dos dois primeiros volumes (CFLU, de 1870, e HMN, de 1873) em discrepância com quatro dos restantes. Depois de um total desaparecimento em VH (1896), eles voltam a ter frequência significativa no último volume publicado pelo autor, em 1906. A divisão percentual em categorias é mais adequadamente visualizada no Gráfico 1.

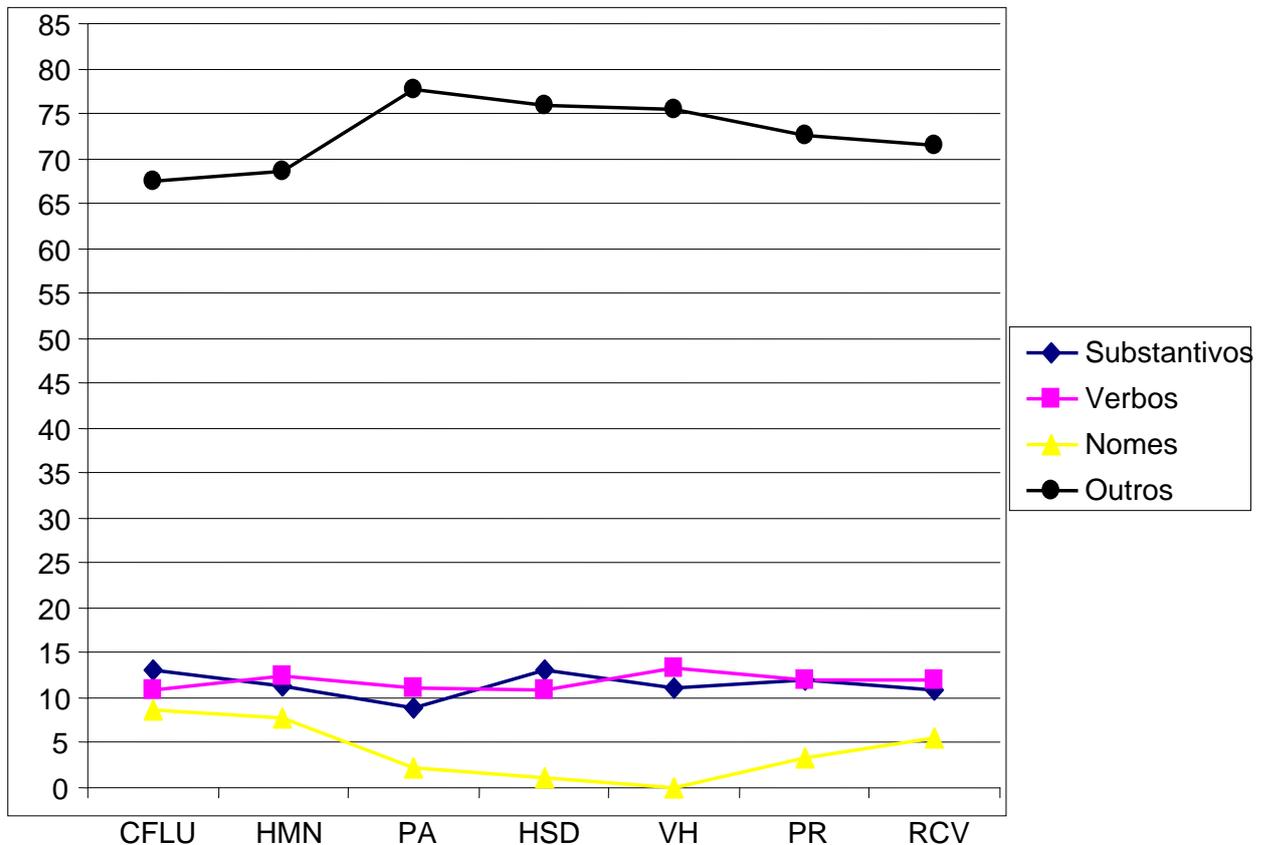


Gráfico 1 – Evolução do uso de substantivos, verbos, outros e nomes próprios nos volumes de contos publicados de Machado de Assis (%) – a partir da lista de 100 maiores frequências

A predominância da categoria *outros* sobre as demais é bastante forte. Mesmo que esse tipo de palavras seja frequente na língua, a diferença mais significativa no conto machadiano está na evolução cronológica: há um forte crescimento, que acontece simultaneamente à redução dos *verbos* e *substantivos*, o que está relacionado a um alongamento das frases. (Sobre essa relação falaremos no Capítulo 6, quando chegarmos aos dados específicos sobre essa reestruturação da extensão de frase.) Mesmo que a categoria *outros* decresça e atinja um equilíbrio de HSD (1884) em diante, ela continua acima do percentual das primeiras obras – CFLU e HMN.

Nos grupos macro, a distribuição das frequências sofre alterações, conforme indica o Quadro 5.

Texto/Item	Substantivos	Verbos	Outros	Nomes próprios
FLUMEIA	12,08	12,08	68,13	7,69
AVG1	14,28	10,98	74,72	0,00
5VOL	10,98	12,08	76,92	0,00
AVG2	13,33	10,00	76,66	0,00

Quadro 5 – Evolução do uso de substantivos, verbos, outros e nomes próprios nos volumes de contos agrupados de Machado de Assis (%) – a partir da lista de 100 maiores frequências

Tomando dois a dois, por ordem cronológica, os grupos Flumeia e AVG1 comparados a 5VOL e AVG2 têm praticamente as mesmas diferenças percentuais quanto aos *substantivos* e aos *verbos*. Quanto à categoria *outros*, os dois primeiros têm uma diferença significativa, com o aumento das palavras funcionais sugerindo maior extensão de frase em AVG1 em relação a Flumeia, além da ausência de *nomes próprios* nas altas frequências. 5VOL e AVG2 ficam praticamente iguais nas duas categorias.

As palavras mais frequentes em um texto podem nos dar uma ideia tanto da temática quanto da estrutura de um texto. As categorias passíveis de lematização¹⁷², a partir da lista das maiores frequências, deram origem aos Quadros 6 e 7.

¹⁷² Os substantivos e verbos foram lematizados, ou seja, foram contabilizadas à mão todas as flexões, tanto dos verbos quanto dos substantivos, a partir do *index* das palavras que compõem cada subgrupo, gerado pelo programa Hyperbase.

Volume / Item	Flumeia (1870-1873)	AVG1 (1858-1880)	5VOL (1882-1906)	AVG2 (1881-1906)
Substantivos	Casa Tempo Carta Rapaz Dia Moça Olhos Homem Amor Coração Marido	Casa Moça Dia Tempo Olhos Coração Amor Coisa Pai Homem Noite Rapaz Vida	Casa Olhos Tempo Dia Anos Homem Rua Vida Cousa Noite	Casa Dia Tempo Olhos Pai Dias Rua Moça Cousa Marido
Verbos	Ser Dizer Ter Haver Estar Perguntar	Ser Dizer Ter Estar Haver	Ser Ter Dizer Estar Haver Poder	Ser Dizer Ter Estar Poder Haver

Quadro 6 – Substantivos e verbos mais usados por Machado de Assis nos volumes agrupados (em ordem de relevância) – a partir das listas de altas frequências de cada grupo

Volume / Item	CFLU (1870)	HMN (1873)	PA (1882)	HSD (1884)	VH (1896)	PR (1899)	RCV (1906)
Substantivos	Casa Carta Dia Amor Tempo Rapaz Homem Marido Olhos Verdade Cousa Mulher	Moça Tempo Rapaz Olhos Casa Homem Coração Carta Dia Tenente Pai	Casa Alienista Dia Homem Olhos Tempo Coisa Amor	Casa Anos Tempo Olhos Dia Cousa Homem Vida Rua Verdade Marido Noite	Olhos Tempo Casa Dia Vida Cousa Amor Rua Homem Noite	Casa Tempo Homem Olhos Amigo Rua Mundo Anos Noite Marido Dia	Casa Tempo Rua Dia Tia Olhos Vida Noite Cousa Amigo
Verbos	Ser Dizer Ter Estar Perguntar Saber	Ser Dizer Estar Haver Ter	Ser Dizer Ter Haver Estar Poder	Ser Ter Dizer Ir Estar Haver	Ser Ter Estar Dizer Ir Haver Poder Ver	Ser Ter Dizer Estar Haver Ver	Ser Dizer Ter Ir Estar Poder

Quadro 7 – Substantivos e verbos mais usados por Machado de Assis nos contos publicados em livro (em ordem de relevância) – a partir das listas de altas frequências de cada grupo

Tanto os substantivos quanto os verbos apresentam uma semelhança significativa, que podemos visualizar mais claramente no Quadro 8.

Volume/ Palavra	CFLU	HMN	PA	HSD	VH	PR	RCV	AVG1	AVG2
Casa	X	X	X	X	X	X	X	X	X
Carta	X	X							
Dia	X	X	X	X	X	X	X	X	X
Amor	X		X					X	
Tempo	X	X	X	X	X	X	X	X	X
Rapaz	X	X						X	
Homem	X	X	X	X	X	X		X	
Marido	X			X		X			X
Olhos	X	X	X	X	X	X	X	X	X
Verdade	X			X					
Cousa	X		X	X			X	X	X
Mulher	X								
Moça		X						X	X
Coração		X						X	
Tenente		X							
Pai		X						X	X
Alienista			X						
Vida				X	X		X	X	
Rua				X	X	X	X		X
Noite				X	X	X	X	X	
Amigo						X	X		
Mundo						X			
Tia							X		

Quadro 8 – Substantivos mais frequentes nos volumes de contos publicados por Machado de Assis e nos avulsos – a partir das listas de altas frequências de cada grupo

As palavras *casa*, *dia*, *tempo* e *olhos* estão presentes em todas as listas de maiores frequências, tanto dos volumes publicados quanto dos avulsos, formando as constantes do conto machadiano. *Homem* e *cousa* vêm logo em seguida, presentes em seis e cinco dos grupos respectivamente. *Rua* e *noite* também aparecem em mais da metade dos grupos.

Nos grupos macro, as palavras exclusivas dos primeiros Flumeia e AVG1, até 1880, são *rapaz*, *moça* e *coração*, sopros de romantismo que perpassam a juventude do autor e de sua escrita. Já os substantivos comuns exclusivos de 5VOL e AVG2, pós-1880, conferem maior complexidade aos contos, pela divisão do foco em dois tipos de realidade: a exterior e a interior. Como as duas almas de Jacobina, em “O Espelho”, os *anos* remetem ao indivíduo e a sua relação com a constante *tempo*, ambos voltados para dentro; já a *rua* nos remete ao mundano, ao Rio de Janeiro – ao Ouvidor –, à vida urbana, remetendo o foco para fora.

Haverá, pois, algum ponto de intersecção entre as duas “fases” quanto aos substantivos predominantes? Uma só palavra une os contos publicados: *homem*, a matéria-prima de toda a obra machadiana. O fenômeno se repete quando cotejamos os avulsos dos dois períodos, que têm em comum *moça* e *pai*. Entre publicados e avulsos, também há

conexões exclusivas, como *noite* e *vida* (entre AVG1 e 5VOL) e *marido* (entre Flumeia e AVG2).

Quanto aos verbos, dois fenômenos nos chamam a atenção. O primeiro é relativo ao verbo *perguntar*, que, em se tratando das altas frequências, é exclusivo dos volumes publicados. Outro é relativo aos grupos macro pós-1880, que gozam da exclusividade do destaque do verbo *poder*, como é possível observar no Quadro 9.

Verbo	CFLU	HMN	PA	HSD	VH	PR	RCV	AVG1	AVG2
Ser	X	X	X	X	X	X	X	X	X
Ter	X	X	X	X	X	X	X	X	X
Dizer	X	X	X	X	X	X	X	X	X
Estar	X							X	X
Perguntar	X	X	X	X	X	X	X		
Saber	X								
Haver		X	X	X	X	X	X	X	X
Poder			X		X		X		
Ir				X	X				X
Ver					X	X			

Quadro 9 – Verbos mais frequentes nos volumes de contos publicados por Machado de Assis e nos avulsos – a partir das listas de altas frequências de cada grupo

O Quadro 10 apresenta os verbos mais utilizados pela literatura de língua portuguesa, resultado de pesquisas pioneiras de lexicologia¹⁷³, utilizadas originalmente por Carlos Maciel em seu estudo sobre o vocabulário de Érico Veríssimo. Quase todos os 40 verbos fazem parte dos contos de Machado de Assis: a única exceção é o verbo *benzer*, o que se explica pela especificidade temática de seu texto de origem, a obra *Auto da Compadecida*.

Achar	Dever	Ir	Saber
Andar	Dizer	Levar	Sentir
Benzer	Dormir	Morrer	Ser
Cantar	Encontrar	Olhar	Sonhar
Chamar	Entrar	Parecer	Subir
Chegar	Estar	Passar	Surgir
Chorar	Falar	Pensar	Ter
Começar	Fazer	Poder	Ver
Dar	Ficar	Pôr	Vir
Deixar	Haver	Querer	Voltar

Quadro 10 – Quarenta verbos mais usados na língua portuguesa

Fonte: MACIEL, Carlos. *Richesse et évolution du vocabulaire d'Erico Verissimo* (1905-1975 – Porto Alegre, Brésil). Paris: Champion; Genève: Slatkine, 1986.

¹⁷³ EMORINE, Jacques. *Lexique et analyse lexicale de l'Auto da Compadecida*. Étude statistique, Thèse de Doctorat, Université de Toulouse, Toulouse, 1968. DUNCAN JR., John C. *A frequency dictionary of Portuguese words*. Michigan-London: University Microfilms International, 1971. Vocabulário poético: Alphonsus de Guimaraens, Cruz e Sousa e Edgard Mata (CUNHA DE SOUZA, Cilene. *Um método quantitativo para a análise lexical*. Rio de Janeiro: Tempo Brasileiro; Brasília: MEC, 1979.

Para completar essa etapa da análise, identificamos os 50 verbos mais utilizados por Machado de Assis nos volumes de contos publicados, exceto os 40 mais comuns da língua (Quadro 11).

Acabar	Correr	Estudar	Mostrar	Rir
Aceitar	Contar	Explicar	Mudar	Sair
Almoçar	Crer	Fechar	Negar	Sentar
Bater	Deitar	Fugir	Ouvir	Servir
Buscar	Descer	Imaginar	Parar	Supor
Cair	Descobrir	Jantar	Parecer	Tirar
Casar	Despedir	Lembrar	Pedir	Tocar
Comprar	Entender	Ler	Pensar	Vestir
Confessar	Escrever	Levantar	Perder	Visitar
Conhecer	Esperar	Meter	Receber	Viver

Quadro 11 – Verbos no infinitivo utilizados em todos os volumes de contos publicados por Machado de Assis (excluídos os 40 verbos do Quadro 10)

4.2 Análise Fatorial e em Árvore dos Substantivos

A análise fatorial¹⁷⁴ faz parte do grupo dos métodos de estatística descritiva multidimensional. Seguindo princípios básicos desse tipo de método de análise de dados, cada uma das dimensões de um quadro retangular de dados numéricos permite definir distâncias entre elementos de outra dimensão. Desse modo, o conjunto das colunas permite calcular, a partir de fórmulas apropriadas, as distâncias entre as linhas, como afirmam Lebart e Salem¹⁷⁵:

Les méthodes factorielles, largement fondées sur l’algèbre linéaire, produisent des représentations graphiques sur lesquelles les proximités géométriques usuelles entre points-lignes et entre point-colonnes traduisent les associations statistiques entre lignes et entre colonnes.

A partir desses quadros de distância, foram desenvolvidas as representações geométricas que aplicamos ao nosso *corpus* de trabalho. Para Lebart e Salem, o maior problema com o qual o pesquisador tem que lidar é relativo às perdas que podem resultar da

¹⁷⁴ Tipo de método multidimensional disponível no Hyperbase, que tem como base estatística a análise de correspondência que segue o algoritmo proposto por Jean-Paul Benzécri, cuja adaptação para Windows foi efetuada por André Salem. Ver BRUNET, op. cit., p. 65-66.

¹⁷⁵ LEBART; SALEM, op. cit., p. 80.

transposição do quadro de linhas e colunas – que abarca toda a informação disponível sobre as variáveis em questão – para a representação gráfica, em que a visualização se torna mais acessível, mas também mais parcial ou incompleta. A melhor alternativa seria a de combinar mais de um método¹⁷⁶, já que, segundo os autores, eles podem ser aplicados sobre os mesmos quadros de dados, um complementando o outro. A maior vantagem desses métodos de representação gráfica (análise fatorial, histogramas, diagramas em bastão), afirmam eles, está no fato de não haver regras de interpretação prefixadas, bastando o que chamam de uma “*apprentissage rudimentaire*”¹⁷⁷.

Em nosso caso, o apelo às representações gráficas é obrigatório, pela extensão do *corpus*. A partir de uma base como 7VOL, composta de 76 textos de extensão variada, um quadro que cruze informações sobre todos eles, comparados dois a dois, seria tão extenso quanto ilegível: ao invés de facilitar o acesso aos dados, tornaria mais complicado o trabalho. Seguimos aqui as regras de Charles Muller para o trabalho de estatística textual: “*Simplicité et constance de la norme, garants de son efficacité, valent bien quelques sacrifices à l’idéal, un peu utopique, d’une norme hautement scientifique*”¹⁷⁸.

Esse tipo de quadro forma uma “nuvem”¹⁷⁹, em que podemos observar a distribuição das variáveis que, ao se afastarem do ponto mais concentrado, afirmam sua especificidade¹⁸⁰.

Como a combinação de mais de um método acaba por ser complementar, tornando a análise mais eficaz, adotamos outro método de representação gráfica: a análise em árvore, desenvolvida por Xuan Luong, que critica as lacunas da análise fatorial:

L’analyse factorial des correspondances (AFC) offre, certes, **représentation simultanée** et **directions principales**, mais ne donne pas le détail et les **hiérarchies** entre proximités. Elle peut même conduire à des erreurs d’interprétations (projections, voisines sur un plan, de points éloignés dans l’espace), ce qui est impossible dans une arbre (où les distances se lisent en

¹⁷⁶ Outro método multidimensional (além da análise fatorial) é o que agrupa a informação em famílias de classes hierarquizadas ou em classes de linhas e colunas. LEBART; SALEM, op. cit., p. 80.

¹⁷⁷ “L’interprétations des histogrammes, les diagrammes en bâtons, les graphiques de séries chronologiques ne nécessitent qu’un apprentissage rudimentaire; alors que dans le cas de l’analyse des correspondances, par exemple, il sera nécessaire de connaître des règles de lecture des résultats plus contraignantes que ne le laisse croire le caractère souvent suggestif des représentations obtenues.” LEBART; SALEM, op. cit., p. 81.

¹⁷⁸ “Principes et méthodes de statistique lexicale”. MULLER apud MACIEL, Carlos Alberto Antunes. *Richesse et évolution du vocabulaire d’Erico Veríssimo...*, p. 31, nota 1.

¹⁷⁹ Análise fatorial como “nuage de points”. Ver MULLER, op. cit., p. 127.

¹⁸⁰ MACIEL, Carlos Alberto Antunes. *Richesse et évolution du vocabulaire d’Erico Veríssimo...*, p. 41.

suivant les arcs et non pas en mesurant visuellement des distances entre feuilles)¹⁸¹.

Voltamos então às lacunas apontadas por Lebart e Salem relativas à análise fatorial. A redução da abrangência da variância¹⁸² por termos que optar por uma combinação de dois entre três fatores torna a análise fatorial menos interessante do que a análise em árvore, que, por sua vez, daria uma noção mais completa sobre os dados que lhe dão origem, como explica Brunet¹⁸³:

L 'algorithme produit des graphes qui rendent compte de la proximité des objets étudiés (ici des textes) à partir d'une distance (ici celle de Labbé¹⁸⁴). L'avantage de cette technique, par rapport à l'analyse factorielle, est qu'on n'a plus à distinguer et à croiser des facteurs, dont chacun n'explique qu'une partie de la variance. Tout l'explication se résume ici en une seule représentation graphique, qui peut prendre deux formes: rectangulaire ou radiale.

É a combinação desses dois métodos que adotamos nesta seção sobre os substantivos. Os dados submetidos às análises fatorial e em árvore são os substantivos extraídos das listas individuais de cada base. A interpretação dos dados de cada base é apresentada a seguir, após o par de gráficos respectivo.

¹⁸¹ LUONG, Xuan; NOVI, Michel. Représentations arborées de données textuelles. In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. *Actes...*, Nice, 1986. p. 577-586, p. 583. (Grifos dos autores).

¹⁸² Ver BARBETTA, Pedro Alberto. *Estatística aplicada às Ciências Sociais*. 5. ed. rev. Florianópolis: Ed. da UFSC, 2005, p. 103.

¹⁸³ BRUNET, op. cit., p. 71.

¹⁸⁴ Sobre este método, “Labbé propose une méthode particulièrement simple destinée à mettre en évidence ce qu'il appelle *l'univers lexical* d'une forme donné. Pour chaque forme *formel* du corpus, l'ensemble des phrases du corpus peut être divisé en deux sous-ensembles : *PI*, sous-ensemble de celles qui contiennent *formel* et *P0*, sous-ensemble des unités desquelles *formel* est absente”. LEBART; SALEM, op. cit., p. 70.

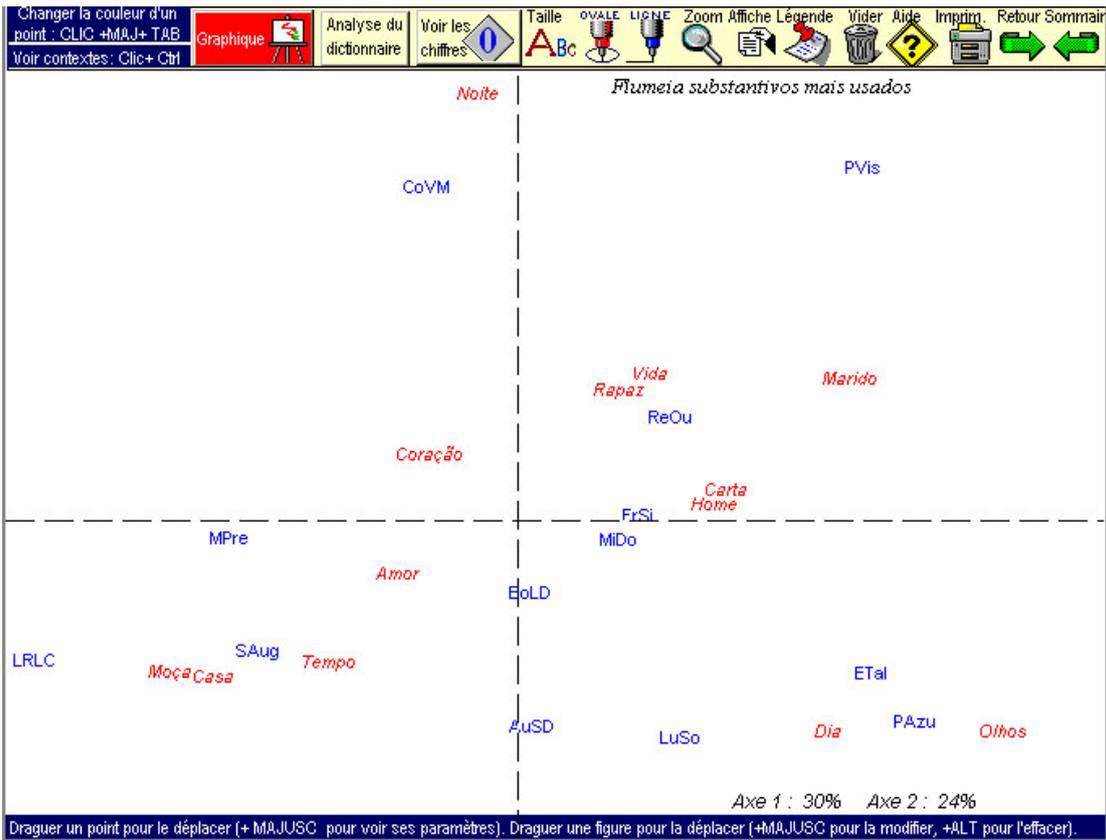


Gráfico 2 – Análise fatorial dos substantivos na base Flumeia

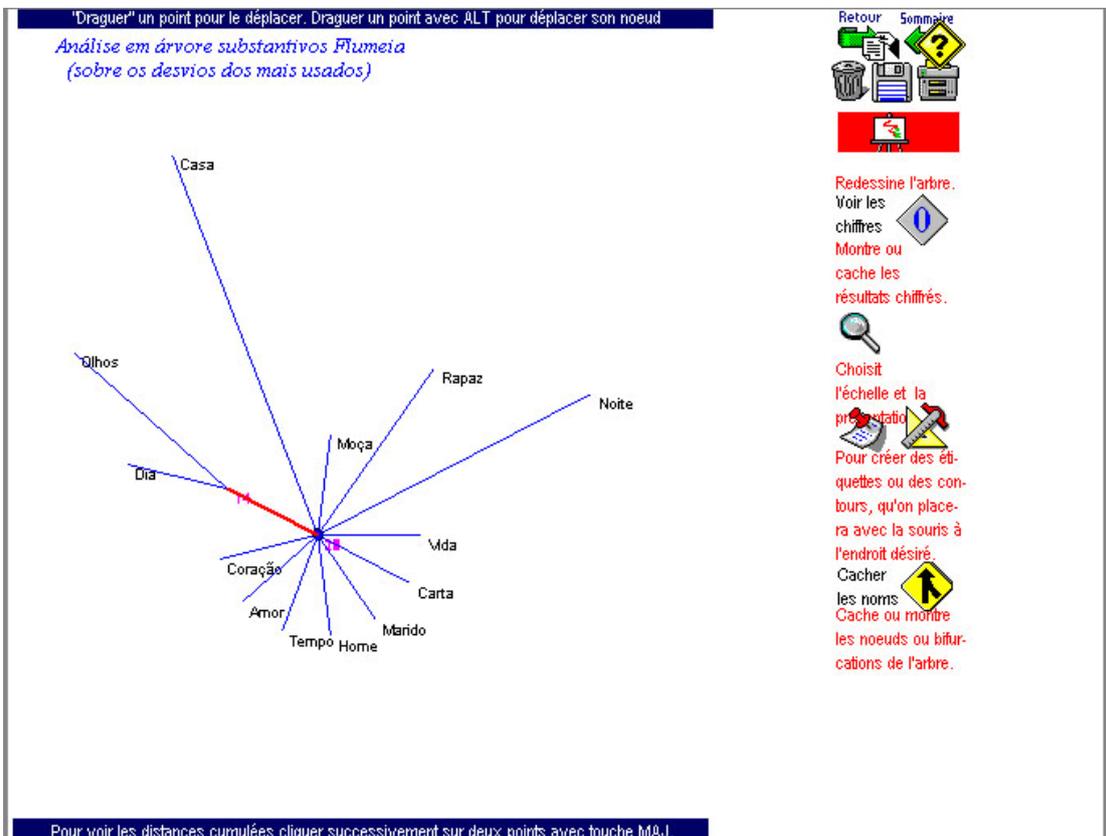


Gráfico 3 – Análise em árvore dos substantivos na base Flumeia

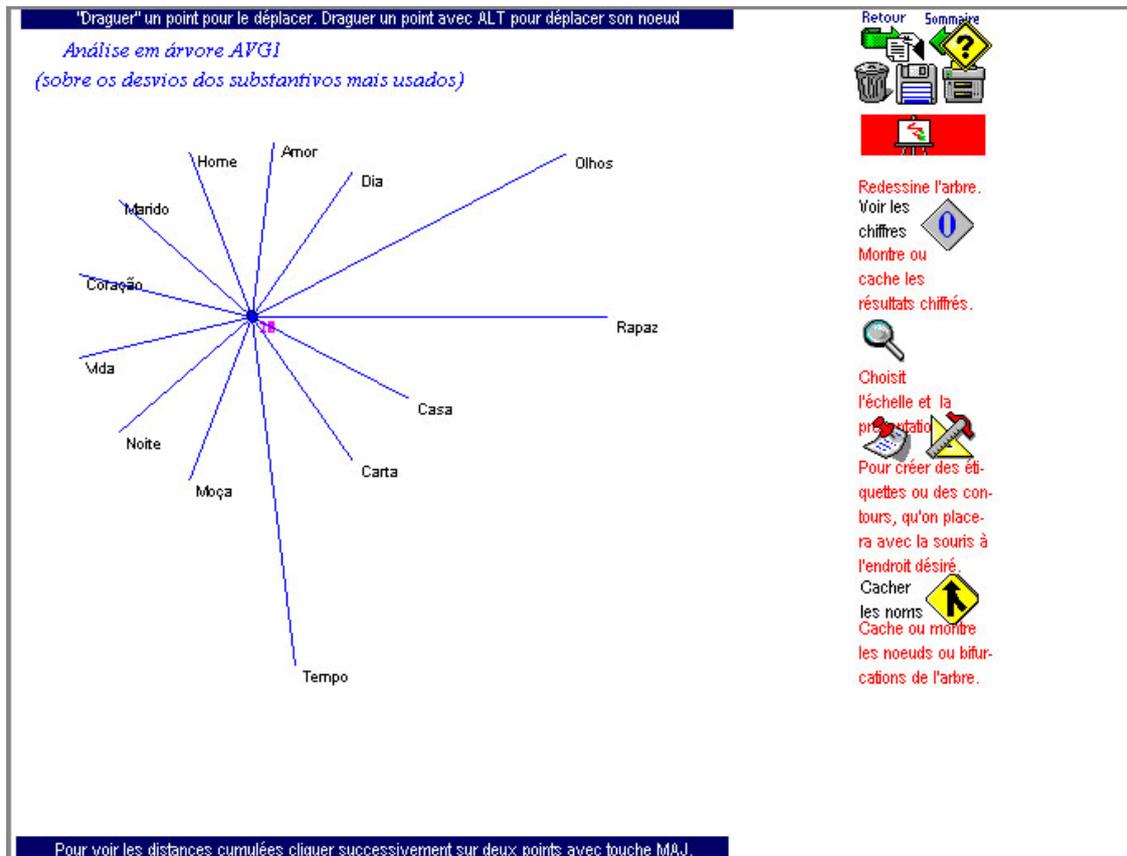


Gráfico 5 – Análise em árvore dos substantivos na base AVG1

Temos novamente discrepâncias entre os dois métodos, na avaliação da base AVG1. Na análise fatorial, a palavra *casa* se afasta de modo significativo de nuvem, que dessa vez está extremamente concentrada. Já na análise em árvore, *tempo* e *olhos*, que na análise fatorial aparecem na área mais concentrada, aqui compõem o conjunto das mais remotas. Também nesse grupo há dispersão na árvore, porém nenhum galho foge ao nó central.

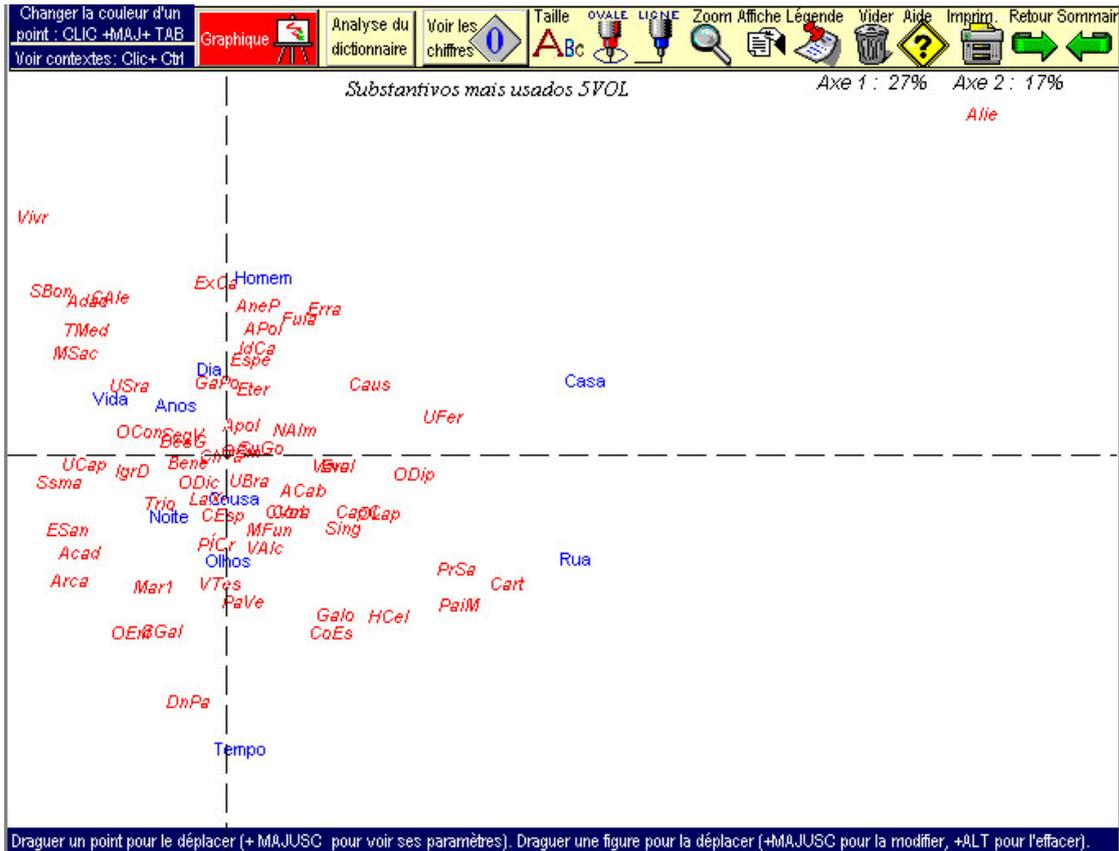


Gráfico 6 – Análise fatorial dos substantivos na base 5VOL

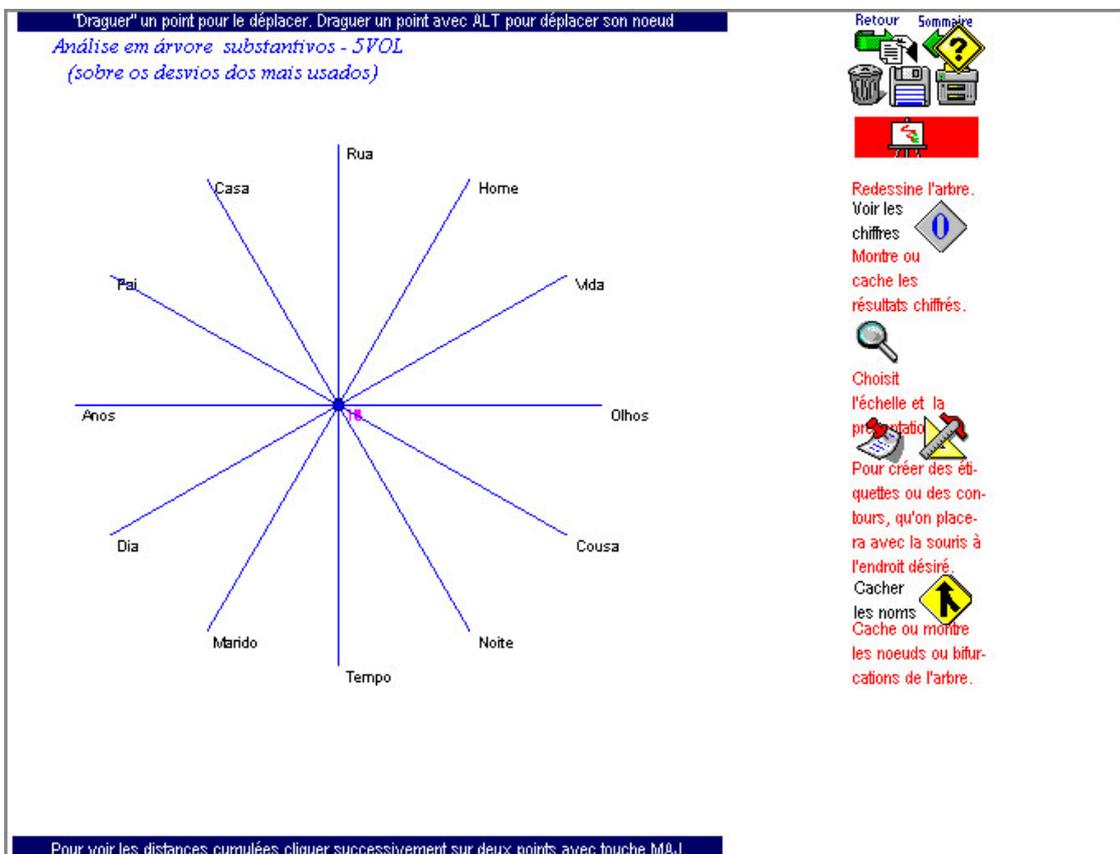


Gráfico 7 – Análise em árvore dos substantivos na base 5VOL

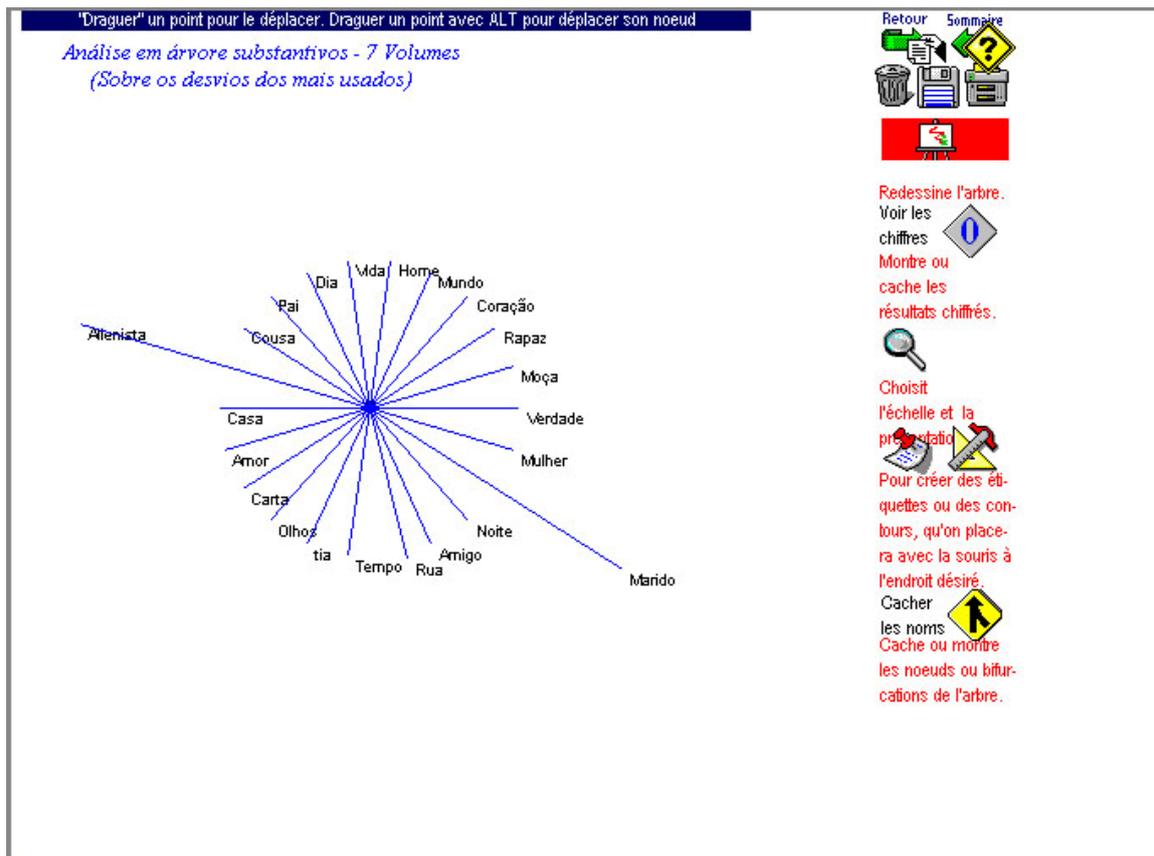


Gráfico 11 – Análise em árvore dos substantivos na base 7VOL

Com a nuvem bastante concentrada, a base 7VOL tem nas palavras *alienista*, *tia* e *casa* as maiores distâncias. A palavra *rapaz* aparece associada a *Luís Soares* de CFLU, e os outros três contos ao redor pertencem a HMN, todos tendo em comum a época. Outra concentração acontece próximo a *marido*, com os contos *Senhora do Galvão*, *D. Paula* e *Confissões de uma viúva moça*, todos de volumes diferentes, mas com tema em comum.

Mais uma vez o que se vê na análise fatorial só se confirma em parte na análise em árvore, com um gráfico equidistante, com todas as folhas partindo do mesmo ponto, e apenas dois afastamentos importantes: *alienista* e *marido*.

Na análise fatorial, como o conto *O alienista* faz parte da base Portext, ele é o único a se aproximar da palavra que lhe dá título. O outro destaque é *cousa*. Palavra bastante usada por Machado (ver a primeira seção deste capítulo), não se estranha que, mesmo afastada do grupo, ela venha acompanhada de dois romances do autor: *Iaiá Garcia* e *Memorial de Aires*. A obra mais afastada dos substantivos mais usados nos contos publicados de Machado são os *Sermões*, de Vieira.

Mais uma vez uma impressão falsa nos é passada pela análise fatorial. Embora com uma nuvem de boa concentração, ela não aponta nem remotamente toda a dispersão que a análise em árvore evidencia. Os quatro galhos, que comportam onze folhas, são apenas a metade ou menos dos outros grupos que podemos identificar, o que não é de se estranhar pela mistura de autores, gêneros e datas dos textos que compõem essa base.

Em termos de análise fatorial, temos em Flumeia a maior dispersão, enquanto AVG1 apresenta a maior concentração. A falta de unidade de Flumeia em relação ao aspecto tratado é confirmada pela análise em árvore. No entanto, a concentração de AVG1 é desmentida, sendo ele o segundo mais disperso nesse tipo de representação, enquanto a maior unidade está em 5VOL.

Neste capítulo tratamos das altas frequências dos verbos e dos substantivos isoladamente. No próximo aplicaremos novamente ao *corpus* análises de tipo multidimensional, com a diferença de que o próximo tipo leva em conta o conjunto do dicionário extraído dos conjuntos das bases pelo programa. A partir da análise do dicionário, focaremos mais no conjunto, principalmente nos sete volumes que mais nos interessam.

CAPÍTULO 5 RIQUEZA LEXICAL E HÁPAX¹⁸⁵

5.1 Riqueza Lexical

No capítulo anterior nos concentramos nas altas frequências e nas listas de substantivos e verbos que selecionamos a partir delas, focando, assim, aspectos mais específicos de forma isolada. Neste capítulo, o objetivo é focar o conjunto quanto ao léxico. Para isso, aplicamos ao *corpus* o recurso do Hyperbase relativo à riqueza lexical, que oferece duas opções de análise: uma sobre as formas; e outra sobre os hápax ou palavras de frequência 1. Como as informações extraídas por meio dessas opções são de naturezas diferentes, optamos por usar as duas, de modo a observar com maior exatidão a evolução do autor no aspecto lexical.

A riqueza lexical, em termos estatísticos, refere-se exclusivamente ao aspecto quantitativo, como explica Maciel:

La notion de «richesse lexicale» [...] est donc indépendante de la présence ou non dans l'oeuvre littéraire de tel ou tel vocable considéré comme étant savant ou rare – et elle ne peut pas être comprise ou confondue avec un jugement de valeur. [...] Notre définition de «richesse lexicale» correspond ainsi uniquement a des données quantitatives ; elle est un élément de la structure du texte et traduit des faits de style¹⁸⁶.

Dado um conjunto formado por textos de tamanhos diferentes, a riqueza lexical é expressa pela extensão do vocabulário (V). De acordo com o contingente de formas (V) que apresenta, um texto pode ser considerado mais ou menos “rico”. As diferenças de extensão dos textos que formam o conjunto são relativizadas pelo programa¹⁸⁷ mediante a aplicação da lei binomial¹⁸⁸, o que torna possível tal comparação.

Como a riqueza lexical calculada pelo programa toma como parâmetro o conjunto dos textos que formam a base de dados, optamos por focar nos grupos macro de nosso *corpus*: 7VOL (todos os volumes publicados em livro), Flumeia, AVG1, 5VOL, AVG2. Para fins de comparação externa, usamos uma base com todos os romances de Machado e a base Portext.

¹⁸⁵ “Du grec hapax legomenon, chose dite une seule fois”, nota em LEBART; SALEM, op. cit., p. 46.

¹⁸⁶ MACIEL, Carlos Alberto Antunes. *Richesse et évolution du vocabulaire d'Érico Veríssimo...*, p. 75.

¹⁸⁷ O cálculo da riqueza lexical baseia-se no quadro de distribuição de frequências e na extensão relativa dos textos, seguindo a lei binomial (método Muller). Ver BRUNET, op. cit., p. 58.

¹⁸⁸ Ver MULLER, op. cit., p. 38-42; ou BARBETTA, op. cit., p. 139-146.

Começamos com a comparação interna, ou seja, entre os grupos que foram comparados entre si de acordo com a cronologia, a fim de detectar o caráter de ruptura ou continuidade estilística de Machado no aspecto lexical.

Tanto a base Flumeia quanto a base AVG1 pertencem ao início da carreira do contista. A primeira compreende os dois primeiros volumes de contos – *Contos fluminenses*, de 1870, e *Histórias da meia-noite*, de 1873; e a segunda, os contos avulsos escritos entre 1858 e 1880¹⁸⁹. Veremos a seguir os gráficos relativos à riqueza lexical, a partir de V.

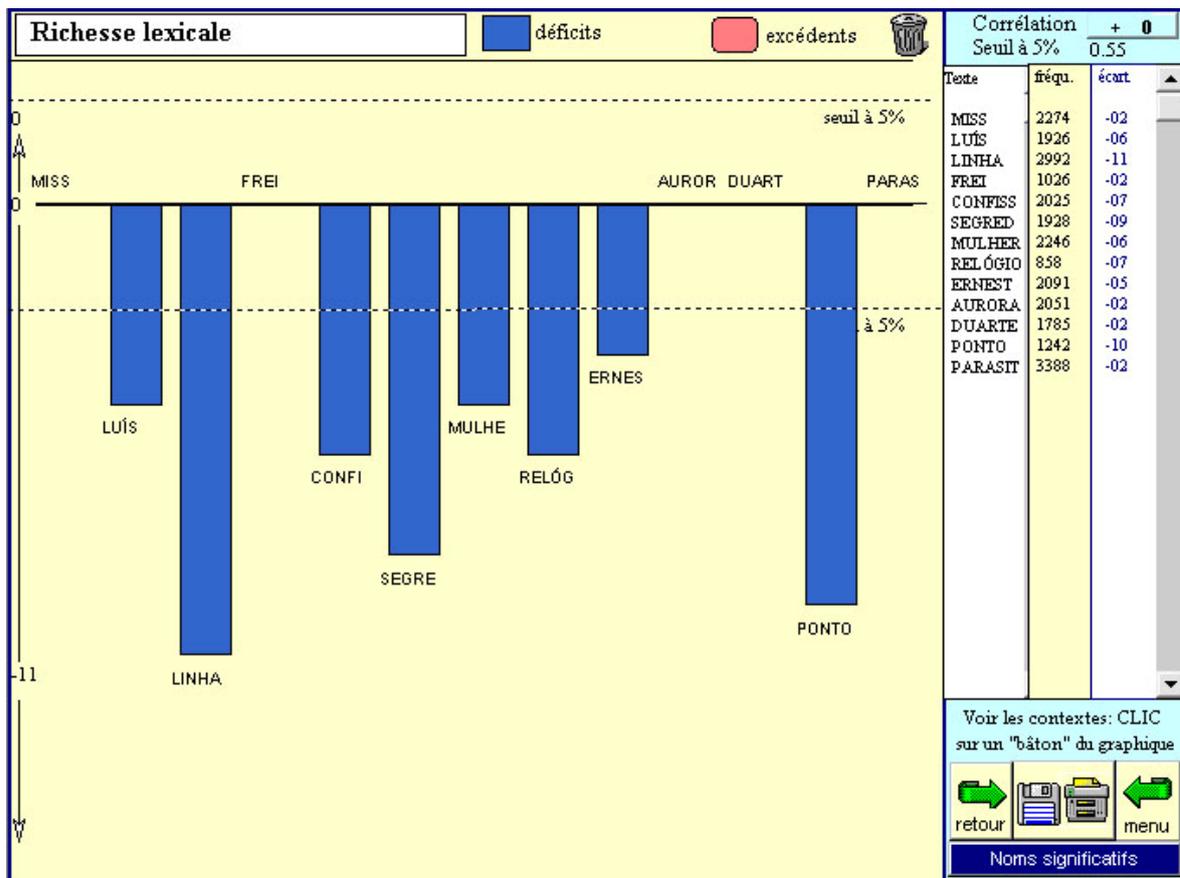


Gráfico 14 – Riqueza lexical na base Flumeia

A base Flumeia é formada pelos dois primeiros volumes de contos publicados por Machado de Assis, *Contos fluminenses* (1870) e *Histórias da meia-noite* (1873), somando treze contos, sete do primeiro e seis do segundo respectivamente. No gráfico, os contos estão distribuídos de acordo com a ordem em que se encontram nos volumes publicados. A ordem cronológica segue somente a da publicação dos volumes (não por ordem cronológica de escrita).

¹⁸⁹ Os contos estão na ordem determinada automaticamente pelo programa no momento da criação da base.

A riqueza lexical é calculada levando-se em conta apenas o conjunto ou *corpus* que forma a base de dados. Portanto, a comparação aqui é apenas entre esses dois volumes.

A leitura do gráfico sugere que nesse ponto da carreira do contista Machado, não havia grande riqueza lexical, pois os déficits são mais numerosos. Os contos mais pobres em termos de vocabulário são *Linha reta e linha curva* e *Ponto de vista*¹⁹⁰. Não há nenhum texto que, de acordo com o cálculo, esteja com excedente nesse aspecto. Os contos de vocabulário mais rico estão na média. Nota-se um leve aumento da riqueza no segundo volume (a partir de “O relógio de ouro”). Dos seis contos, metade está na média de riqueza lexical, enquanto no volume anterior apenas dois de sete estavam na mesma situação.

O movimento de aumento da riqueza lexical de acordo com a cronologia segue a normalidade, uma vez que o autor, no decorrer do tempo, vai experimentando e acrescentando palavras até consolidar um vocabulário próprio.

Vejamos como se comporta a base AVG1, que é composta dos contos avulsos que vão de 1858 a 1880.

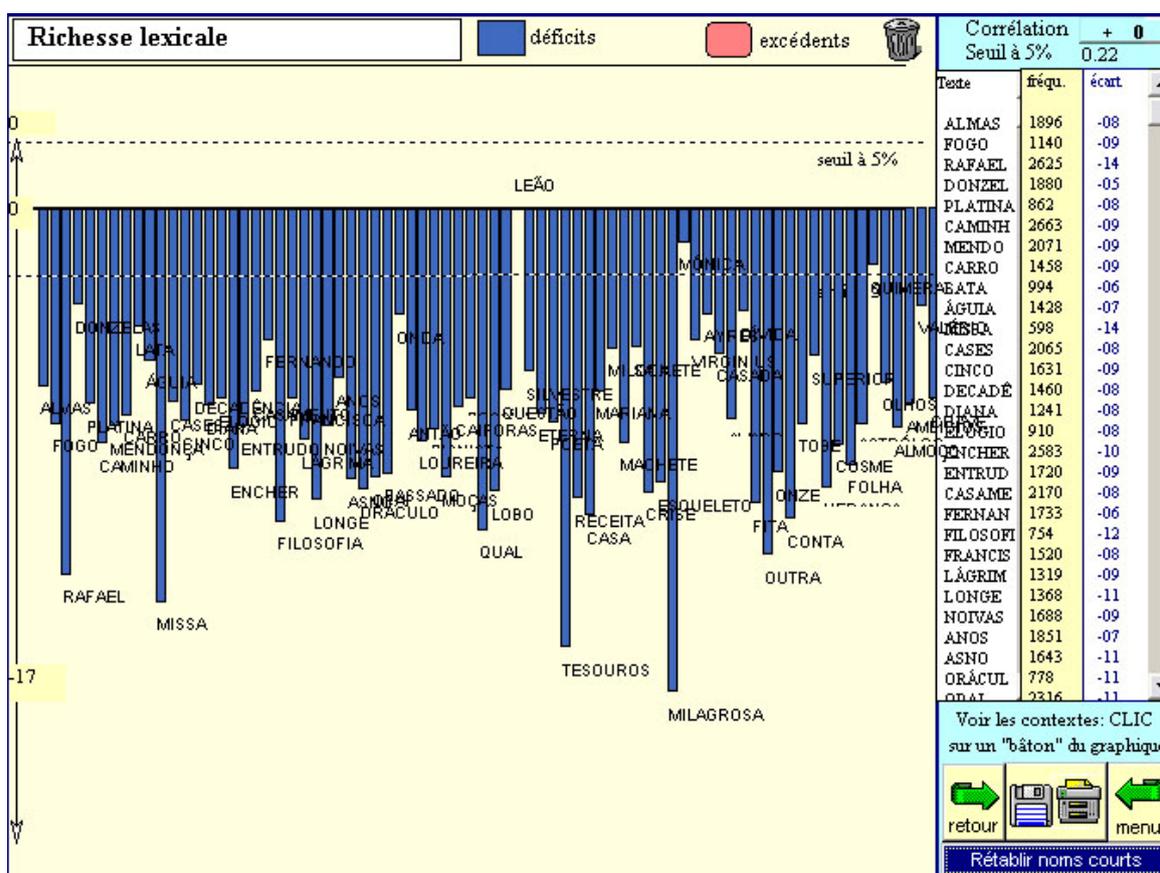


Gráfico 15 – Riqueza lexical na base AVG1

¹⁹⁰ Com data de escrita de 1865 e 1873 respectivamente. Ver CUNHA, op. cit., p.198-199.

O comportamento de AVG1 quanto à riqueza lexical também é predominantemente deficitário. Os contos que se destacam por serem mais “pobres” são: *Uma excursão milagrosa* (1866), *Três tesouros perdidos* (1858), *Antes da missa* (1878) e *O anjo Rafael* (1869). Os de maior riqueza lexical, ou, no caso, os únicos não deficitários são *Ruy de Leão* (1872), *D. Mônica* (1876) e *O país das quimeras* (1862).

Neste grupo podemos observar que o grau de riqueza lexical não está necessariamente associado à época dos contos, ou ao menos que seu crescimento não segue a ordem cronológica, pois há tanto textos mais ricos no início da escrita do autor (1862) quanto mais pobres escritos depois (1878). Vejamos se, com a chegada do ano de 1881, que supostamente marcaria o fim da “primeira fase”, altera-se o quadro.

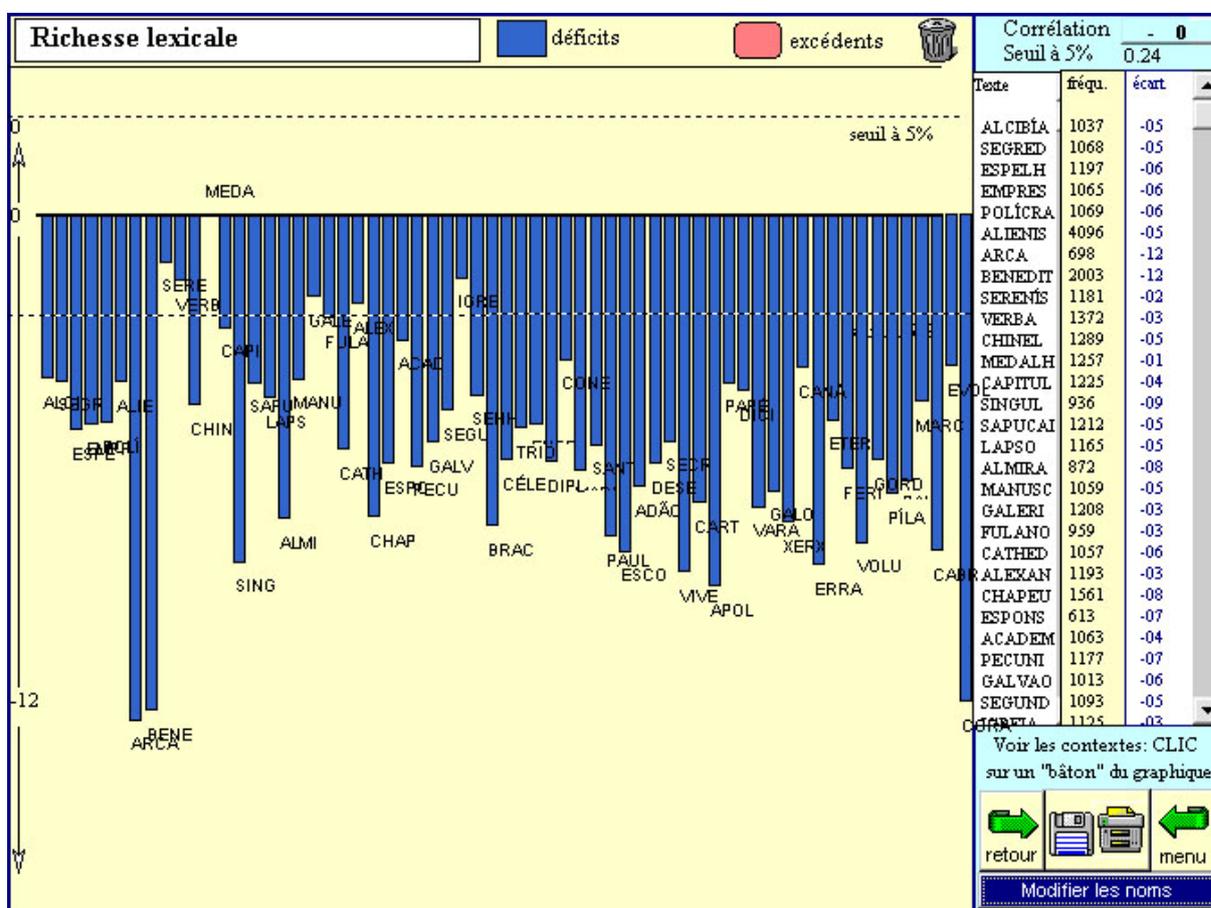


Gráfico 16 – Riqueza lexical na base 5VOL

O amadurecimento estilístico que se traduz na qualidade dos contos de *Papéis avulsos* não aparece ligado necessariamente a um aumento da riqueza lexical. Embora sejam menos deficitários que os demais, uma parte importante dos contos desse volume continua deficitária. Os maiores destaques, em termos de déficit lexical, são os contos *Na arca* e *D.*

Benedita, ambos do volume *Papéis avulsos* (1882), e *Maria Cora*, de *Relíquias de casa velha* (1906). Os mais “ricos” são *Teoria do medalhão*, *A sereníssima República*, *Verba testamentária* e *A igreja do Diabo*.

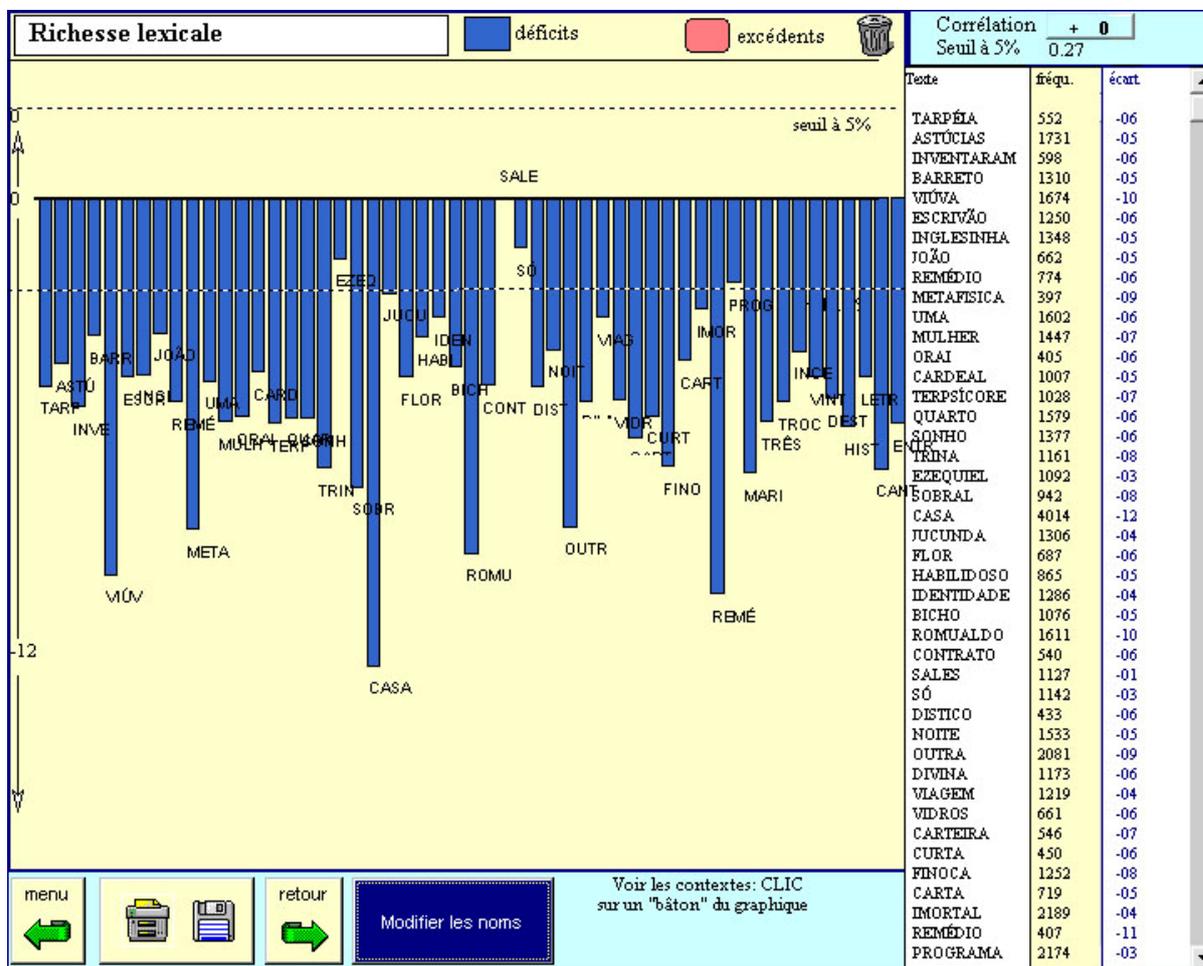


Gráfico 17 – Riqueza lexical na base AVG2

Como podemos ver, os déficits parecem ser o padrão nesse aspecto dos contos machadianos. Em AVG2, *Casa velha*, *O melhor dos remédios* e *O caso da viúva* são os mais “pobres”, enquanto *Sales*, *Só*, *O programa* e *A ideia de Ezequiel Maia* (1883), os de maior riqueza.

Tomando nossas quatro bases macro, chegamos a algumas conclusões. A primeira é a de que há uma semelhança muito forte entre todos os grupos em termos de riqueza lexical e sua tendência é para o déficit. Independentemente de serem avulsos ou publicados, os conjuntos mantêm essa semelhança. Outro ponto importante é que não constatamos nenhuma mudança radical de 1881 em diante nesse aspecto.

Passemos então apenas ao conjunto dos publicados, para verificar se o quadro se mantém o mesmo.

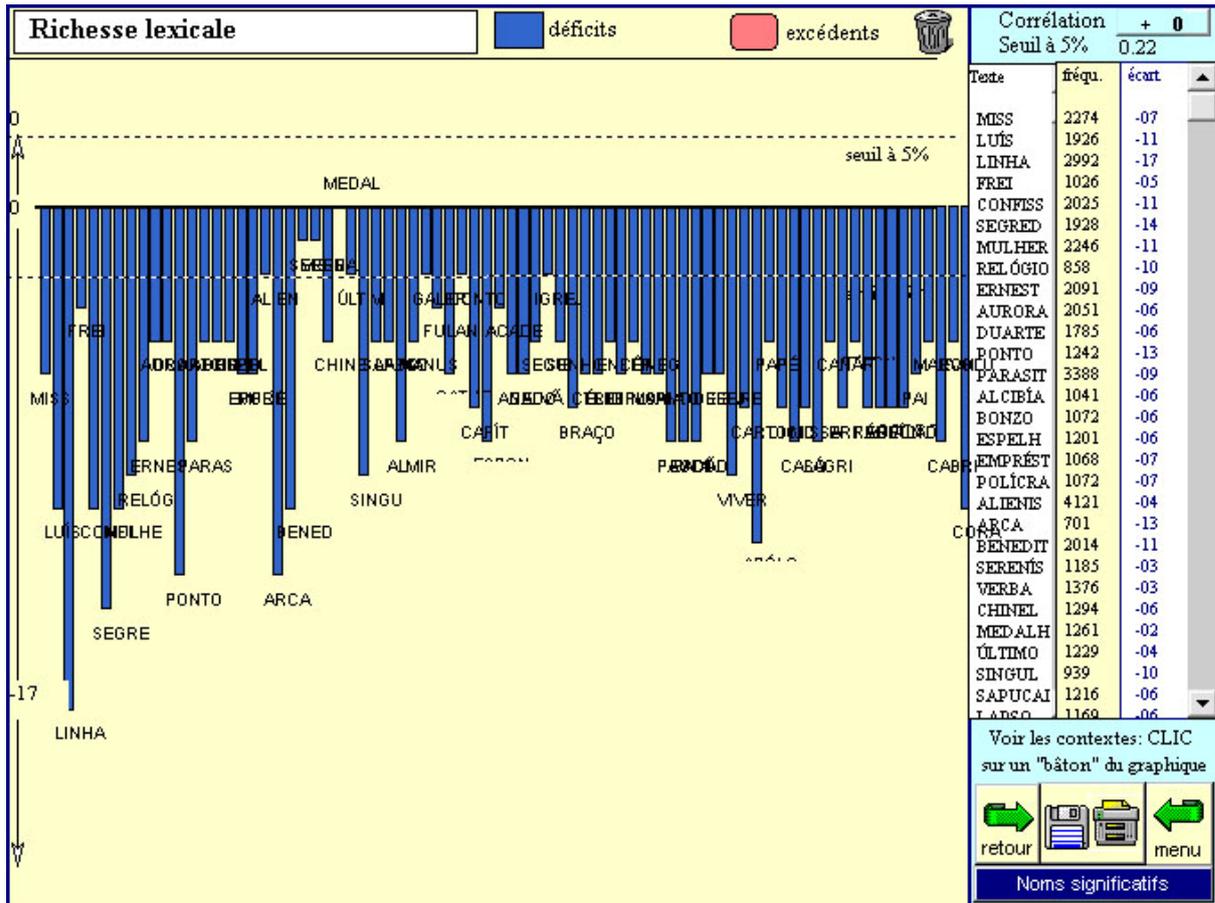


Gráfico 18 – Riqueza lexical na base 7VOL (todos os volumes publicados)

Com base nos dados do Gráfico 18, vejamos agora, em conjunto, como fica a classificação dos contos quanto aos mais “ricos” e aos mais “pobres”, enfocando apenas os mais significativos:

- maior riqueza lexical: *Teoria do medalhão* (1882) PA; *Verba testamentária* (1883) PA; *Sereníssima República* (1883) PA; *O alienista* (1883) PA; *Último capítulo* (1884) HSD; *Galeria póstuma* (1884) HSD; *A igreja do Diabo* (1884) HSD e *Conto alexandrino* (1884) HSD; e
- menor riqueza lexical: *Linha reta e linha curva* (1870) CFLU; *O segredo de Augusta* (1870) CFLU; *Ponto de vista* (1873) HMN; *Na arca* (1882) PA; *Um apólogo* (1896) VH; *Maria Cora* (1906) RCV; *Luís Soares* (1870) CFLU e *D. Benedita* (1882) PA.

O gráfico confirma, em primeiro lugar, o caráter deficitário em termos de vocabulário. Podemos observar que Machado, nos primeiros contos, tem o vocabulário mais pobre. Aos poucos, este começa a aumentar e atinge seu ápice com a *Teoria do medalhão*, em 1882, para voltar a diminuir, aumentar e, enfim, atingir um equilíbrio. Parece-nos um movimento natural, já que no início o escritor está em busca de seu vocabulário específico, das palavras com que

lida melhor. A experimentação segue até ele atingir uma maturidade artística. Daí em diante, seu universo vocabular está definido, e as variações que eventualmente quebrem o equilíbrio ou a estabilidade desse movimento passam a ser exceções.

Para fins de comparação quanto à riqueza lexical, criamos uma base com todos os romances do autor. Escolhemos os romances, pois, como também fazem parte do grupo da prosa de ficção, as diferenças de gênero são atenuadas.

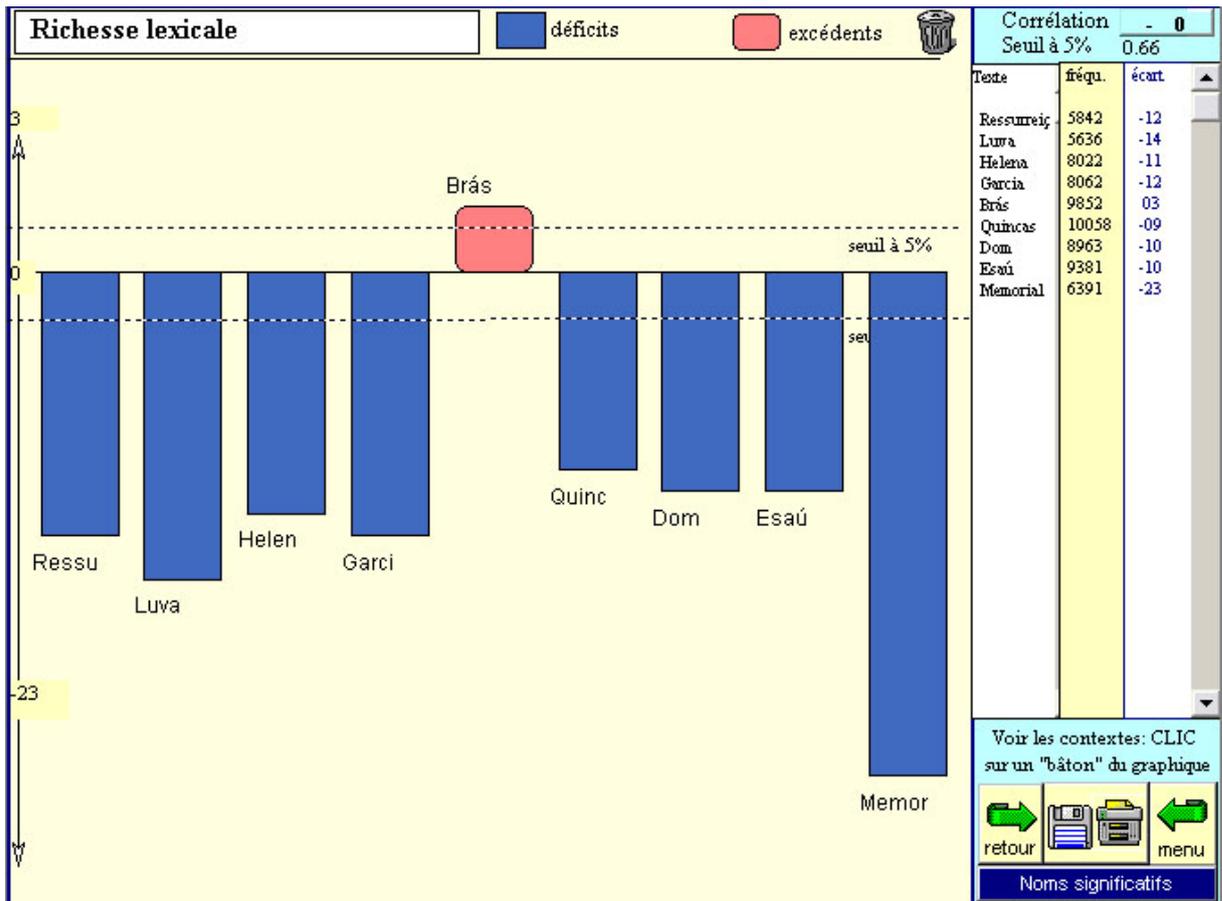


Gráfico 19 – Riqueza lexical na base Romances

A análise dos romances, no Gráfico 19, confirma um padrão do autor em relação à riqueza lexical. Apenas *Brás Cubas* ultrapassa a média da riqueza, atingindo um superávit. Quanto ao movimento, é praticamente o mesmo dos contos, publicados ou não: uma baixa inicial, crescimento, ápice e equilíbrio, tendo no *Memorial* uma exceção para o período.

Se Machado não se destaca pela riqueza lexical, qual sua posição, nesse aspecto, em relação a outros escritores de nossa língua? Vejamos como sua obra se comporta na base Portext.

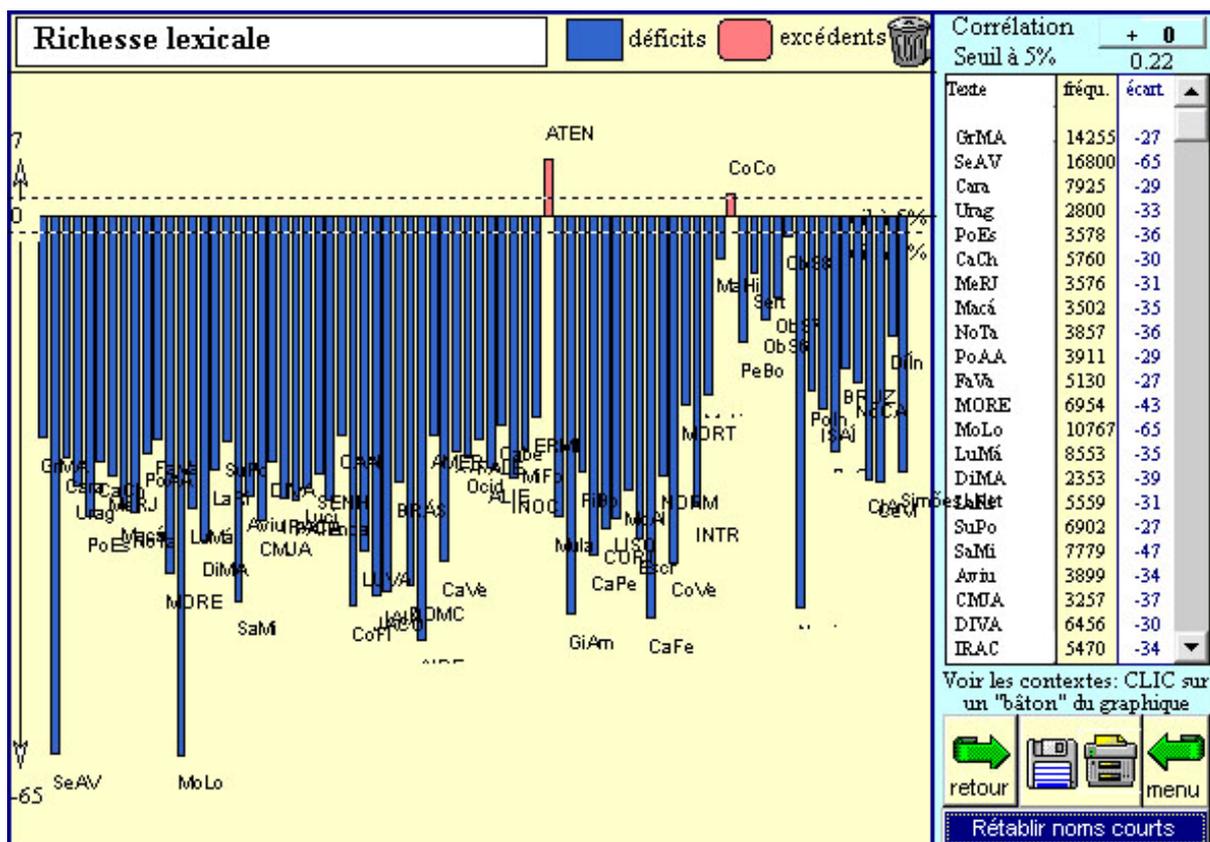


Gráfico 20 – Riqueza lexical na base Portext

O que parecia muito deficitário em Machado se torna relativo ao analisarmos o Gráfico 20. Independentemente do gênero ou da época ou do autor, há um predomínio dos déficits, que na obra machadiana não costuma descer tanto abaixo da média quanto a maioria dos textos dessa base.

5.2 O Crescimento Lexical

Além de uma análise comparativa quanto à riqueza do vocabulário, vimos que o movimento de crescimento deste está ligado à evolução cronológica da escrita do autor. O crescimento lexical, portanto, é um dado que pode contribuir para a análise do estilo, como nos ensina Muller:

A condition d'être comparé à un modèle mathématique et non apprécié au hasard, l'accroissement du vocabulaire est certainement un excellent moyen

de decelerar as variações estilísticas e por vezes temáticas à l'intérieur d'un texte¹⁹¹.

O programa que utilizamos dispõe de um recurso capaz de aferir especificamente esse crescimento, mas, como o cálculo toma como parâmetro os textos que formam a base – e esta não pode exceder 76 textos –, optamos por fazer um quadro de crescimento lexical apenas do grupo que mais nos interessa: o dos sete volumes publicados (7VOL).

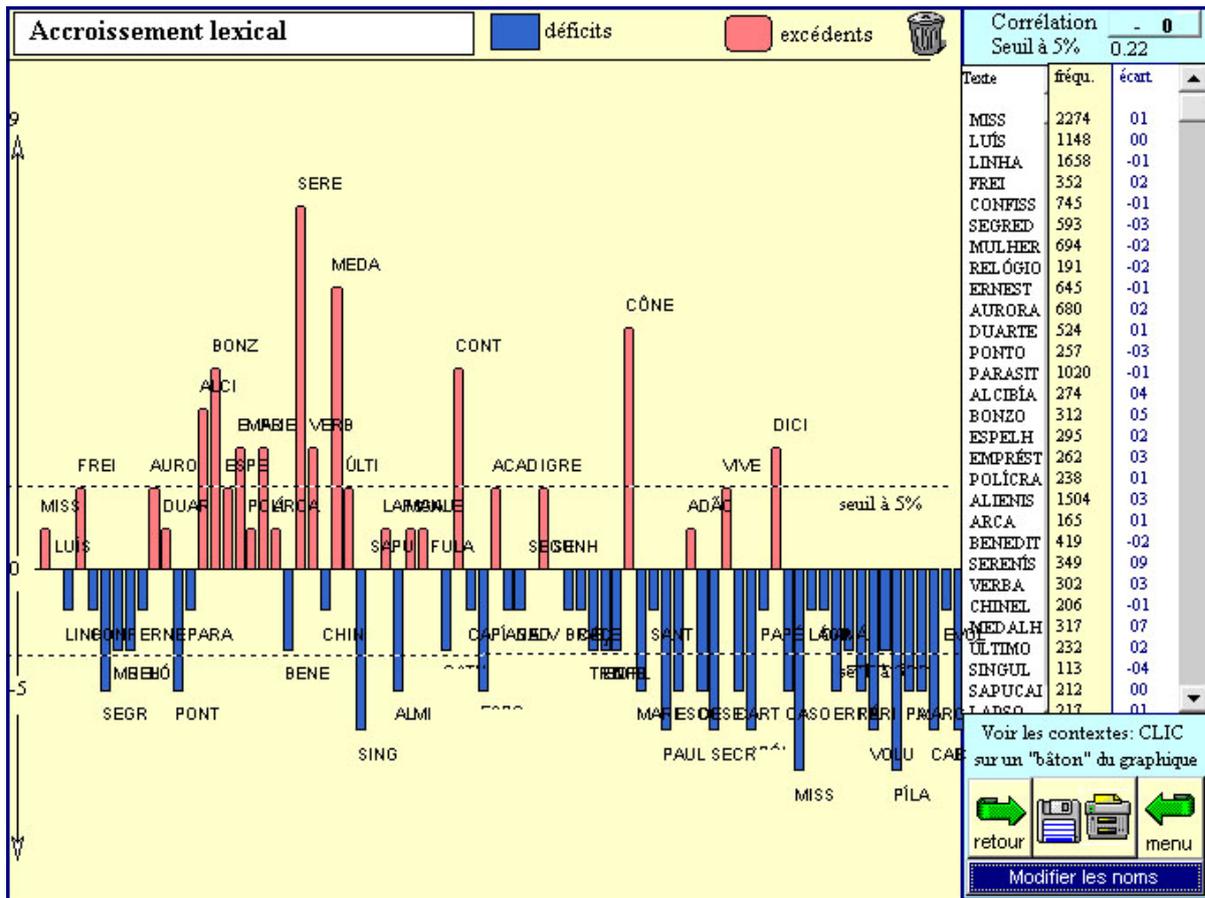


Gráfico 21 – Crescimento lexical na base 7VOL

Comparando o Gráfico 20, de riqueza lexical dessa base, com o Gráfico 21, podemos confirmar o movimento descrito anteriormente. Há uma instabilidade no primeiro volume, CFLU, com alguns contos em crescimento e outros em decréscimo, o que se repete em HMN. O crescimento fica mais significativo a partir de *Aurora sem dia* (1873), e PA traz um movimento ascendente mais duradouro, que mantém o vocabulário acima da média até *Uma senhora* (1884). Em outras palavras, temos uma década de crescimento antes de o vocabulário se estabilizar. No entanto, não podemos deixar de salientar que no conto *O dicionário*, de

¹⁹¹ MULLER, op. cit., p. 190.

1899, o vocabulário de Machado de Assis ainda é capaz de crescer – 26 anos após *Aurora sem dia* e 29 anos depois do primeiro volume publicado, o bruxo continua com fôlego para se renovar.

Vamos, a seguir, continuar a análise da riqueza lexical, só que a partir das palavras de frequência 1.

5.3 Hápax

Nesta seção, tratamos dos dados relativos à riqueza lexical calculados sobre os hápax. Nesse caso, o cálculo usado para a relativização da extensão dos textos segue a lei normal¹⁹². As tabelas com os dados de base (incluindo riqueza lexical) estão reunidas no apêndice.

Segundo Muller, “les causes stylistiques agissent surtout par élimination, les causes thématiques par sélection”. O autor aqui trata do conceito de léxico de situação¹⁹³, que engloba dois tipos de elementos: um ligado ao interlocutor e ao efeito que o locutor quer produzir; e outro que se refere ao que o locutor quer comunicar, ao “conteúdo da mensagem”. O primeiro é de ordem estilística, e o segundo, de ordem temática. Ou seja, as palavras mais raras acabam mais ligadas ao assunto tratado do que o estilo, que estaria centrado na rejeição de determinadas palavras, expressões e construções linguísticas.

O contingente de hápax tem forte relação com o estilo e suas transformações, como nos afirma o autor:

[...] dans une tranche du texte, l’effectif des vocables de fréquence ou de sous-fréquence 1 donne une image statique d’un fait dynamique, à savoir l’entrée dans le texte des lexèmes non encore employés depuis le début du discours, donc du courant est d’autant plus intense que le lexique est plus riche.

L’effectif des vocables de fréquence (ou de sous-fréquence) 1, figé dans un tableau de distribution à côté des autres classes, est donc à traiter comme un élément stylistique, et à mettre en relation avec la richesse du lexique de situation; c’est une des observations qui permettent une appréciation quantitative sur la langue¹⁹⁴.

¹⁹² Ver BRUNET, op. cit., p. 58.

¹⁹³ “Elle englobe deux sortes d’éléments: les uns sont liés à l’interlocuteur (dans le cas de l’écrivain, à son public réel ou imaginé) et à l’effet que le locuteur veut produire; les autres sont liés à ce que le locuteur veut communiquer, au ‘contenu du message’. Les premiers sont d’ordre stylistique, les seconds d’ordre thématique”. MULLER, op. cit., p. 138-139.

¹⁹⁴ MULLER, op. cit., p. 191.

Vejamos, pois, como se dá a evolução da frequência 1 na contística machadiana, a começar da base Flumeia (Gráfico 22).

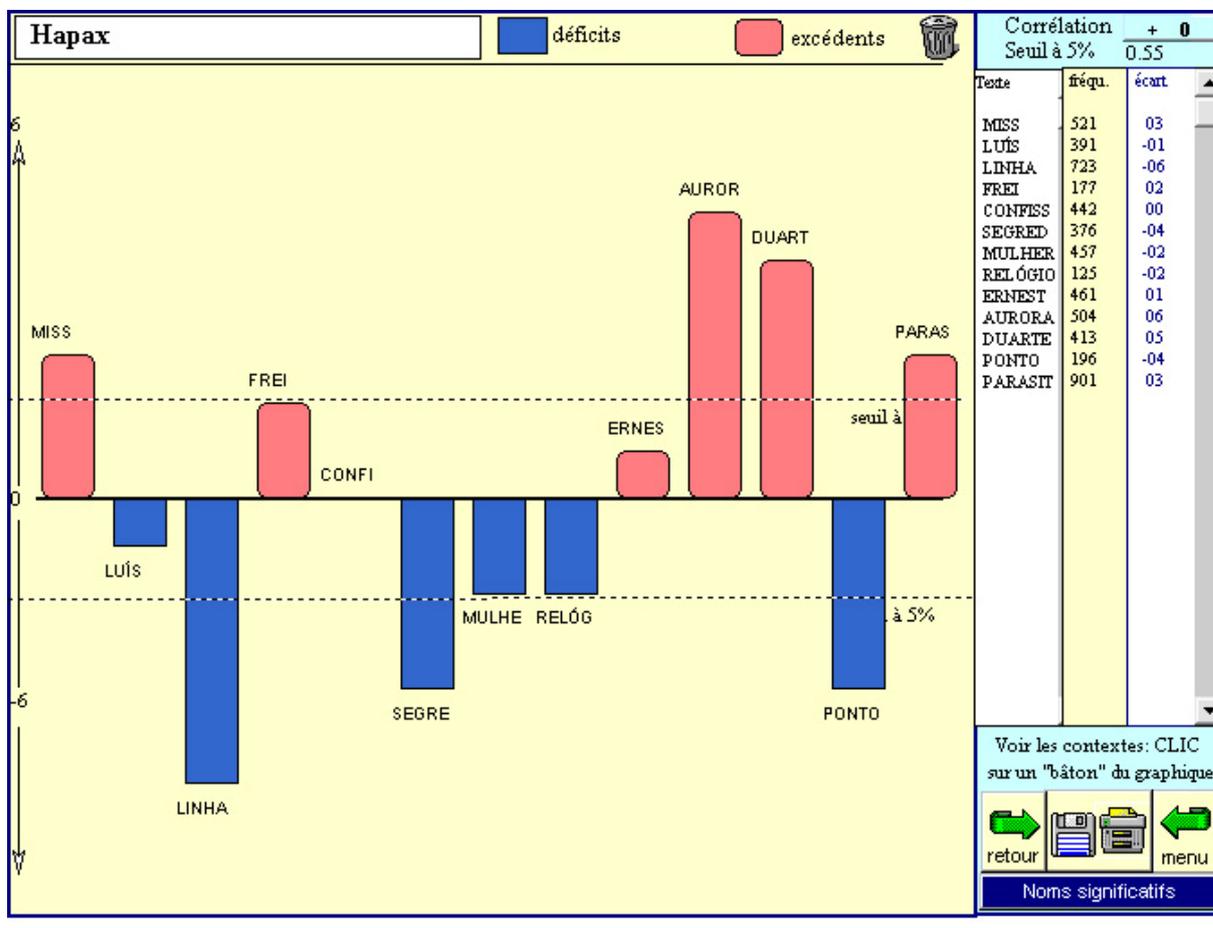


Gráfico 22 – Histogramas da riqueza lexical sobre hápax em Flumeia

Como o autor, aqui, está no início de sua carreira de contista, consideramos natural que as palavras novas sejam experimentadas e acrescentadas gradualmente.

Um pouco diferente de Flumeia, AVG1 aponta para uma maior experimentação em termos do acréscimo de palavras de frequência 1, havendo mais excedentes do que déficits. Vamos aos números mais significativos: *Um cão de lata ao rabo* (1878); *Elogio da vaidade* (1878); *Ruy de Leão* (1872) e *O califa de platina* (1878).

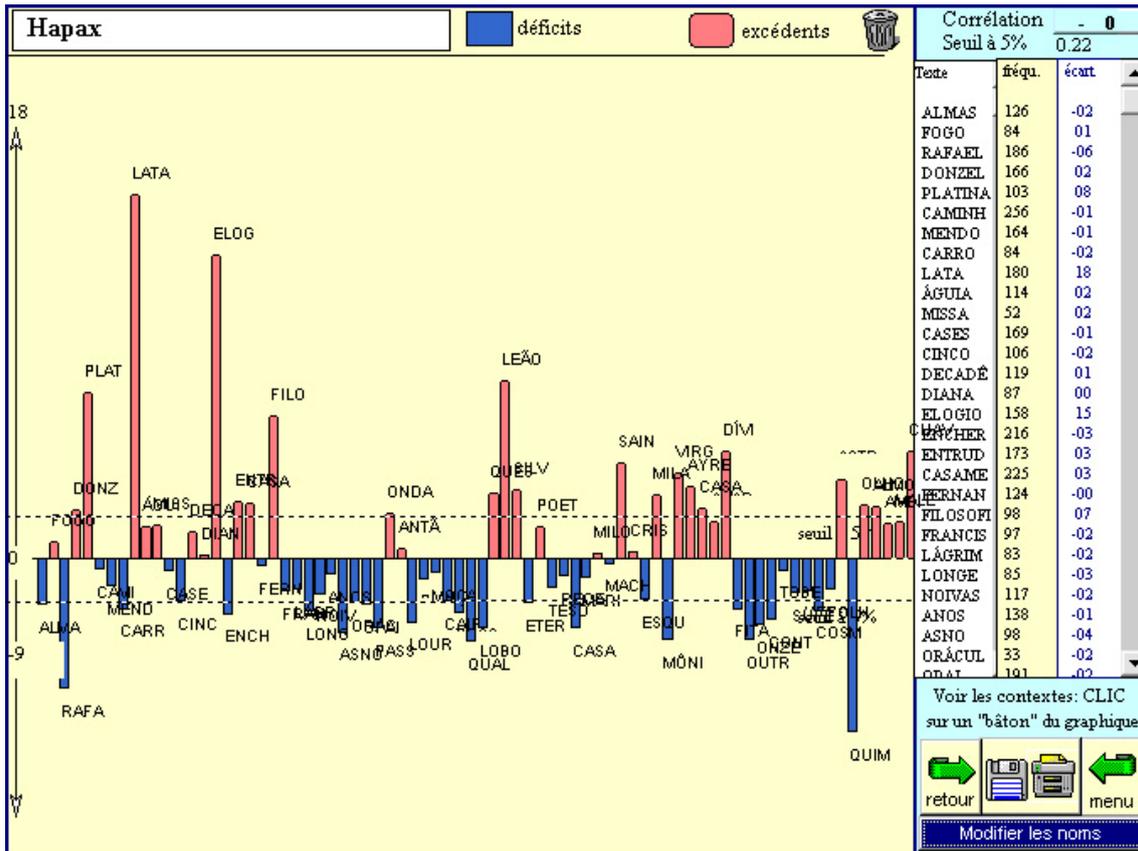


Gráfico 23 – Histogramas da riqueza lexical sobre hápax em AVG1

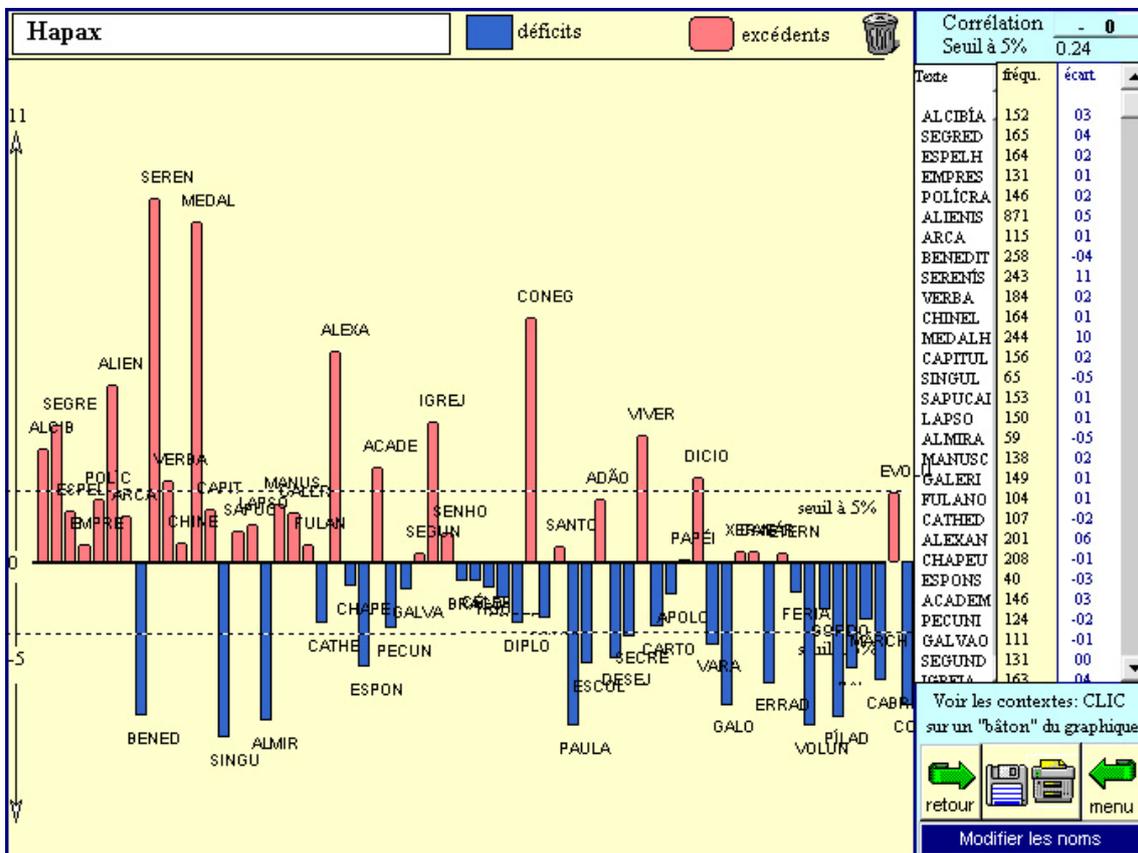


Gráfico 24 – Histogramas da riqueza lexical sobre hápax em 5VOL

O movimento de crescimento do vocabulário que vimos anteriormente, até 1884, que é sugerido tanto pelo gráfico de riqueza sobre o vocabulário quanto pelo do crescimento lexical, se confirma aqui, com o período que abrange PA e HSD como os mais ricos em hápax.

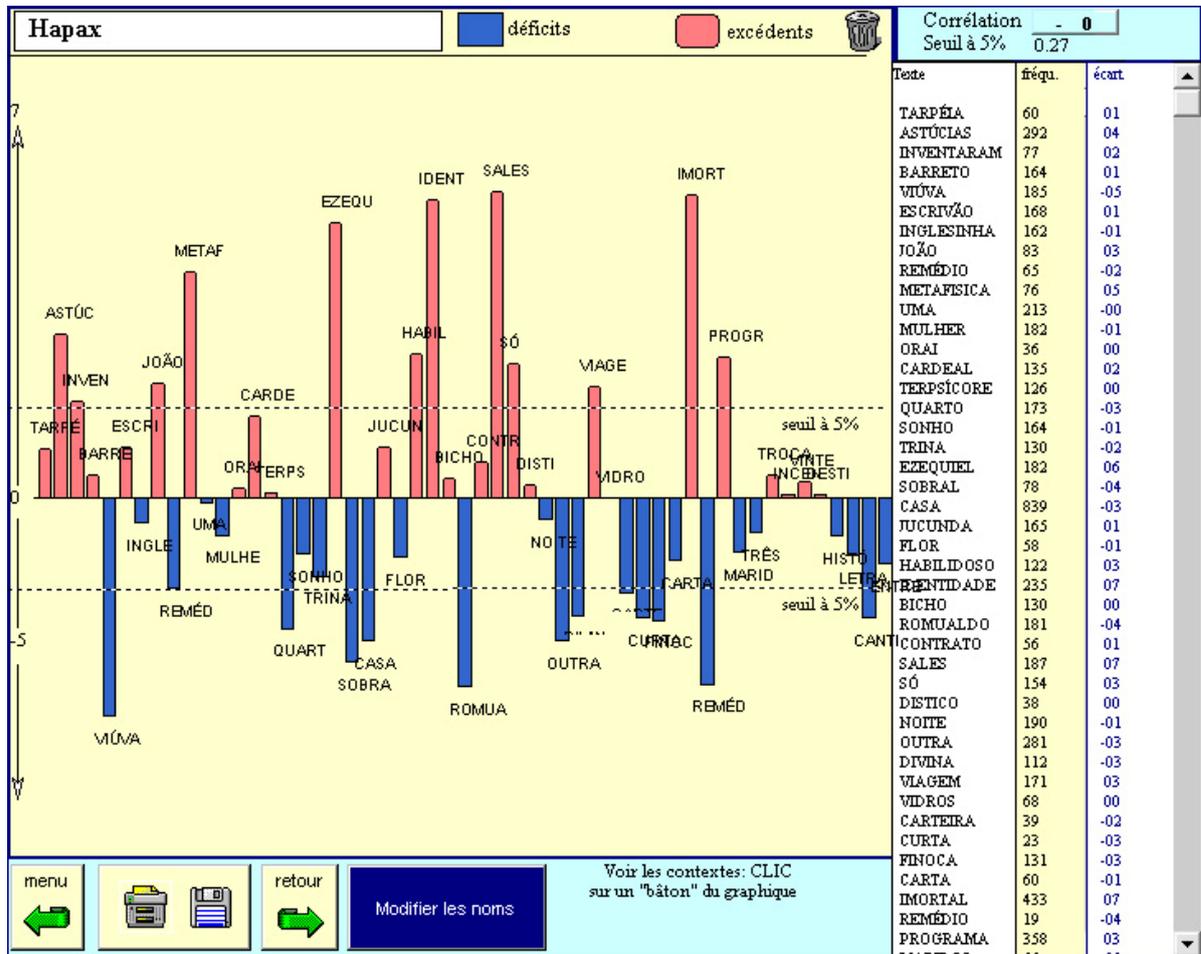


Gráfico 25 – Histogramas da riqueza lexical sobre hápax em AVG2

Como observamos na comparação entre Flumeia e AVG1, há uma abundância de hápax nos avulsos em relação aos publicados. Na comparação entre 5VOL e AVG2, observamos o mesmo fenômeno, só que de modo mais evidente. Para nós isso constitui mais um indício de que os avulsos serviram como laboratório do perfeccionista Machado, para chegar aos 76 contos que decidiu transformar em livro.

Vejamos quais os contos mais significativos quanto ao excedente de hápax: “Sales” (1887); “O imortal” (1882); “Identidade” (1887); “A ideia de Ezequiel Maia” (1883); “Metafísica das rosas” (1883) e “Astúcias de marido” (1886).

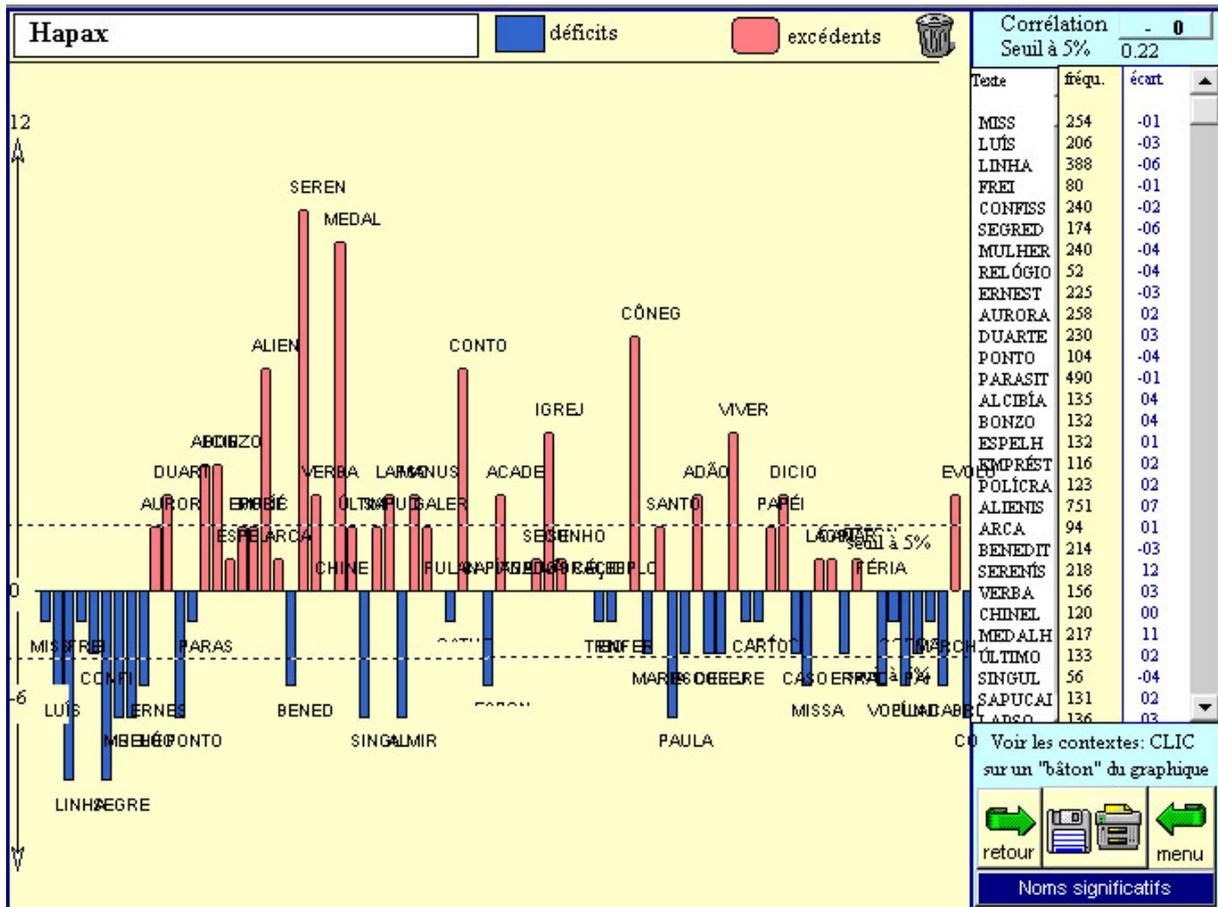


Gráfico 26 – Histogramas da riqueza lexical sobre hápax em 7VOL

Se tomarmos o Gráfico 26 e o compararmos com o de riqueza lexical da mesma base (Gráfico 18), veremos com clareza como os altos índices de hápax coincidem com os menores déficits de riqueza, ou seja, os mais “ricos”. Confirma-se, pois, nos contos publicados a complementaridade desses dados. Os mais significativos desse grupo são: “A sereníssima República” (1883) PA; “Teoria do medalhão” (1882) PA; “O cônego” (1896) VH; “O alienista” (1882) PA; e “Conto de escola” (1896) VH.

Vamos aos romances, para saber como sua evolução se comporta em relação à do conto.

O histograma de hápax dos romances (Gráfico 27) coincide com o do Gráfico 19, de riqueza lexical sobre o vocabulário relativo a essa base. Confirma-se o período mais rico de Machado entre *Iaiá Garcia* e *Dom Casmurro*, de 1878 a 1899, tendo seu auge na década de 1880, como acontece nos contos – aqui com *Brás Cubas*, de 1881, seguido por *Quincas Borba*, exatamente uma década depois, em 1891.

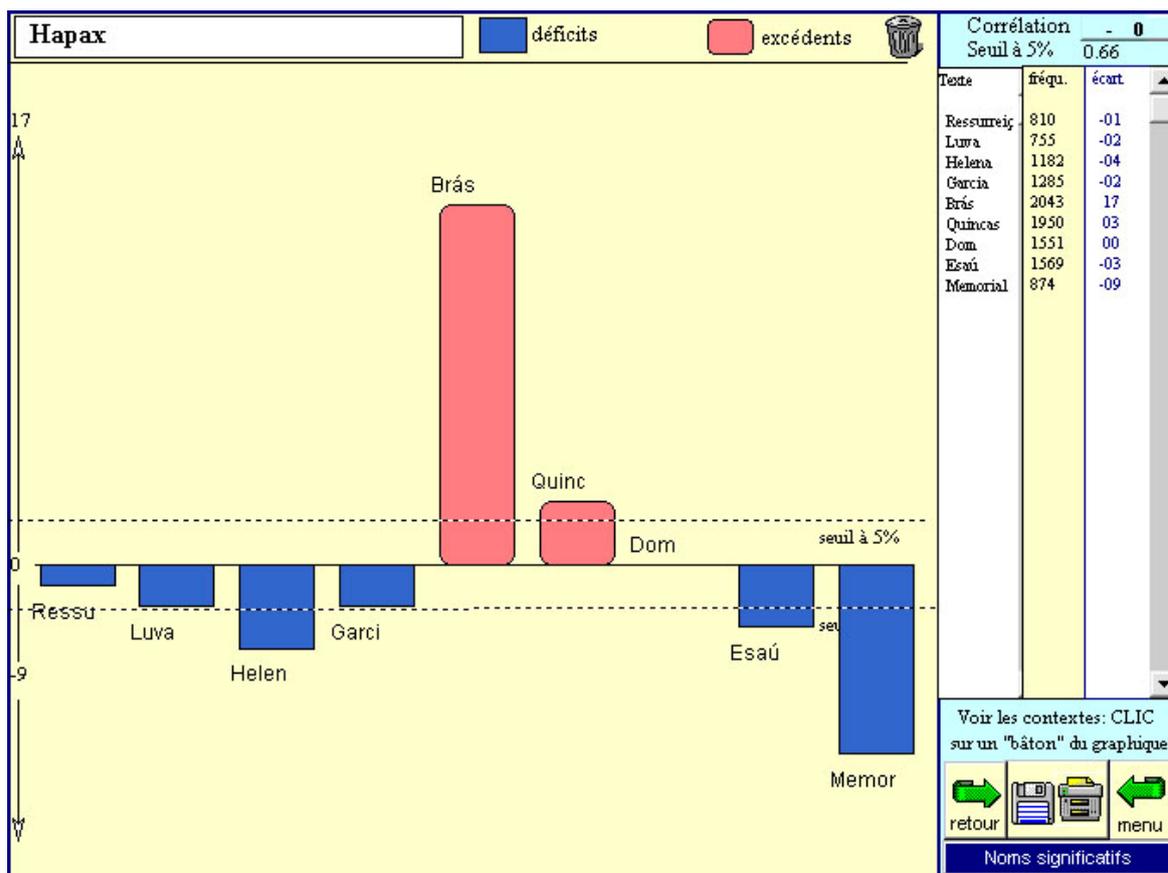


Gráfico 27 – Histogramas da riqueza lexical sobre hápax em todos os romances

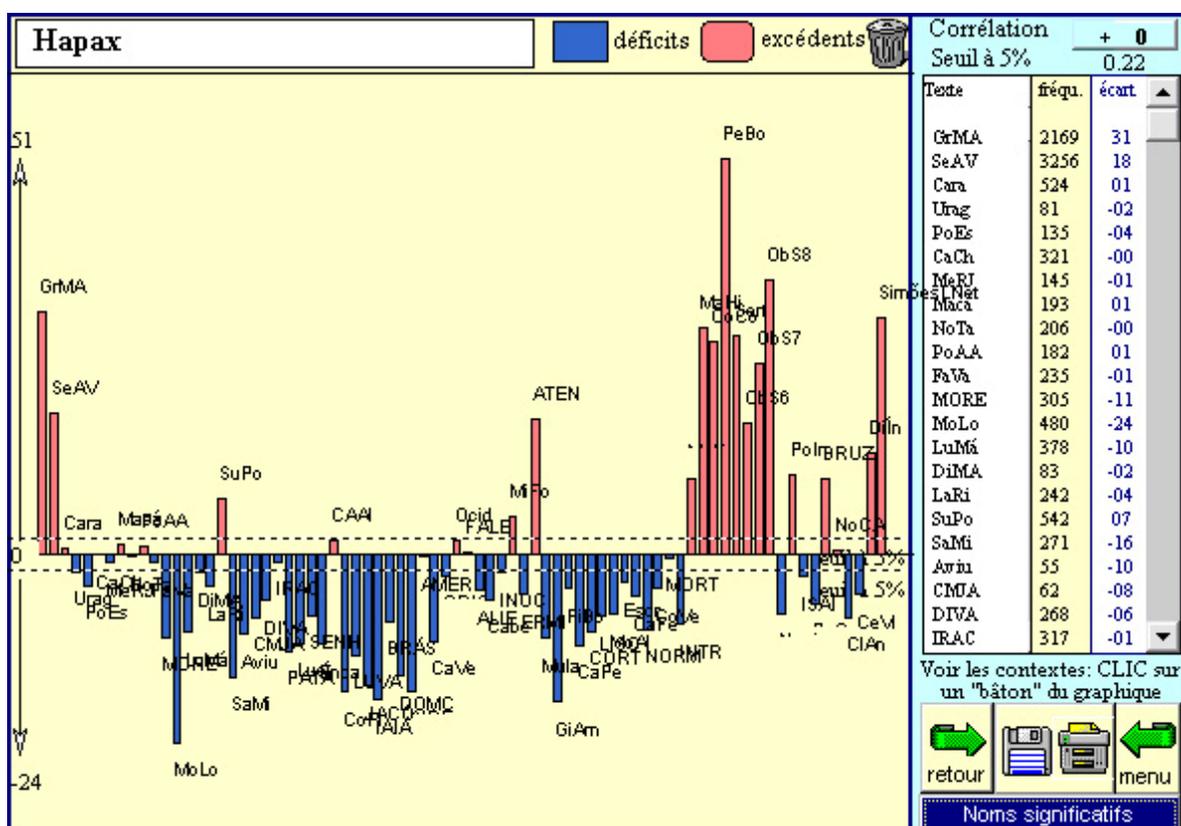


Gráfico 28 – Histogramas da riqueza lexical sobre hápax na base Portext

Enquanto no Gráfico 20, de riqueza sobre o vocabulário, *O Ateneu* e *Contrastes e confrontos* estão solitários em seus bastões vermelhos, aqui nos hápax eles perdem essa exclusividade (Gráfico 28). Muitos são os textos com excedente na frequência 1, e Machado está entre eles – porém, como poeta, com *Ocidentais* e *Falenas*. Como são muitos os gêneros e autores, e a época não é homogênea, não nos arriscaremos a fazer qualquer interpretação dessas diferenças específicas – o que, além disso, foge ao objeto de nosso trabalho.

Finalizando este capítulo, concluímos, a partir dos indicadores acima, que não é possível considerar o vocabulário de Machado rico. A riqueza lexical, contudo, não está necessariamente ligada à qualidade literária do conto, visto que textos importantes, como *Casa velha*, encontram-se entre os mais pobres em vocabulário, e outros clássicos, como *Missa do Galo*, nem apareçam nas listas de maior riqueza e hápax (o que também acontece com *Dom Casmurro*, nos romances). Isso não desmerece em nada a qualidade do autor.

Então, onde está a grande qualidade desse que ainda hoje é nosso maior contista? Se não temos uma originalidade significativa no aspecto lexical em um autor da qualidade de Machado, é preciso investigar onde ela se encontra. O estudo de Etienne Brunet sobre o vocabulário de Proust pode nos ajudar nessa busca:

Pour être complète, la liste devra incorporer les hapax, qu'on rencontre sous la plume de Proust et qui représentent seulement la moitié de l'effective attendu (380: une simple règle de trois suffit pour ce calcul). C'est dire la discrétion de Proust dans la fabrication lexicale: il se contente des ressources que la langue offre dans ce domaine et sa marque est plus dans la combinaison des éléments du lexique que dans leur invention. C'est dire que l'originalité de Proust est moins dans le lexique que dans la syntaxe, et moins dans la syntaxe que dans la pensée¹⁹⁵.

Se a originalidade não está no vocabulário, vamos procurá-la em outro aspecto. No próximo capítulo, investigamos a construção das frases e a distribuição do texto, a partir da análise da pontuação.

¹⁹⁵ BRUNET, Etienne. *Le vocabulaire de Proust : étude quantitative*. Genève: Slatkine, 1983, v. 1, p. 27.

CAPÍTULO 6 A PONTUAÇÃO E A EXTENSÃO DE FRASE

6.1 Pontuação

Num estudo que envolve estatística lexical, a primeira coisa que nos vem à mente são as palavras de alta frequência. Porém, a estatística textual pode ser útil para compreender melhor a distribuição e a ordenação do material linguístico ao longo do *corpus*, o que chamaremos de composição.

Etienne Brunet afirma que a pontuação pode constituir um campo privilegiado de uma análise textual quantitativa, pois ela permite escapar ao círculo estreito da palavra ao abrir uma perspectiva sobre a frase e o ritmo do discurso¹⁹⁶. Como o tratamento automático de textos prejudica a análise do sentido, já que separa o material que o compõe de seu contexto, a pontuação, segundo Brunet, acaba restando como um dos únicos acessos ao nível supralexical do *corpus* em análise. Ou ainda, como afirmam Leech e Short, “*written prose has an implicit, ‘unspoken’ intonation, of which punctuation marks are written indicators*”¹⁹⁷.

Em trabalho anterior, analisamos quatro contos de Machado de Assis utilizando recursos bem mais simples de estatística textual¹⁹⁸. Embora nossa amostra fosse muito

¹⁹⁶ “La ponctuation pourrait constituer un domaine privilégié de la linguistique quantitative. Elle permet en effet d’échapper au cercle étroit du mot et d’ouvrir une perspective sur la phrase et le rythme du discours. Tant que la reconnaissance du sens – qui conditionne dans une large part celle de la syntaxe – n’aura pas trouvé une solution acceptable dans le traitement automatique des textes, la ponctuation restera l’un des seuls accès au supralexical dont on puisse tirer parti.” BRUNET, Etienne. La ponctuation et le rythme du discours (d’après les données du Trésor de la langue française). *CUMFID*, Université de Nice, n. 13, juil. 1981, p. 1.

¹⁹⁷ LEECH, Geoffrey N.; SHORT, Michael H. *Style in fiction*. New York: Longman, 1990, p. 215.

¹⁹⁸ O ponto de partida desta tese são os dados que levantei em minha dissertação de mestrado sobre as características que, a meu ver, mais marcaram a transição estilística de Machado para seu estilo maduro, demarcada pela publicação de *Papéis avulsos*. Na ocasião fiz um estudo de caso de quatro contos do autor com o objetivo de identificar os principais pontos de intersecção e ruptura entre o estilo juvenil dos primeiros contos e o estilo apurado da maturidade artística de Machado. Para efetuar-lo, parti do estudo ontológico da obra literária de Roman Ingarden; passei pela teoria literária de Warren e Wellek, pela estilística de Geoffrey Leech e Michael Short, até chegar aos conceitos usados por Gérard Genette para a análise da obra de Marcel Proust. Selecionei elementos que considere complementares – em razão de alguns não terem aplicabilidade prática ou serem muito vagos para explicar determinadas fenômenos – e assim compus um modelo híbrido de análise. Como resultado, pude constatar que a mudança estilística mais sensível do conto machadiano se deu no nível da composição (no sentido ingardiano de ordenação da sequência). Os dois programas de estatística textual usados foram o *Wordlist* e o *Mconcord*. Essas ferramentas não contavam com muitas funções (apenas listagem do vocabulário por ordem de frequência, cálculo de média de extensão de palavras, frases e parágrafos e concordâncias) e só tinham capacidade para analisar um texto por vez. Ainda assim, proporcionaram um avanço em termos de precisão e rapidez na contagem e hierarquização das palavras e frases usadas, informações que estariam muito mais vulneráveis a erro se feitas à mão. Ver: FREITAS, Deise J. T. de. *A revolução do estilo de Machado de Assis*:

pequena, ela apontou as mudanças na extensão de frases e parágrafos como mais relevantes do que as alterações de vocabulário, ou seja, as principais diferenças no conto machadiano aconteceriam em termos da ordenação da sequência, distribuição na composição.

Com as palavras tratadas de forma isolada pelas listas de frequências, a exploração da pontuação pode nos ajudar a analisar, por exemplo, a extensão de frase (ou período). Em *corpus* de grande extensão, usados para fins comparativos – como em nosso caso –, isso pode revelar mudanças estilísticas importantes.

Outra vantagem da análise da pontuação é que se trata de uma categoria com poucas variações: não são muitos os sinais e, além disso, eles não sofrem nenhum tipo de flexão – dificuldade que toma muito tempo de pesquisa, como ocorre no processo de separação nas listas de frequência com os verbos e os substantivos¹⁹⁹.

Entretanto, podemos com o estudo da pontuação observar mudanças no estilo de escrever de um autor? O primeiro problema que se coloca ante esse objetivo são as ambiguidades que podem decorrer da análise, por exemplo, dos pontos. No caso do Hyperbase, o próprio programa separa os pontos das reticências, mas não dos pontos que servem para abreviar palavras. Assim sendo, fizemos manualmente a separação a partir das palavras abreviadas que mais aparecem nas listas de altas frequências, pois estas poderiam mascarar o número de pontos. Como Machado não usa muitas abreviações, foi possível, pela análise das listas de concordância, obter os números relativos às repetições de *d.*, *sr.* ou *dr.* e suas respectivas flexões. Uma vez contabilizadas, estas foram subtraídas do número total de pontos, restando apenas os pontos finais.

Além do ponto final, as interrogações e as exclamações também fazem parte da pontuação forte, ou seja, dos sinais que demarcam o limite de uma frase²⁰⁰. Nas Tabelas 3 e 4, a seguir, veremos como se comporta a pontuação forte no conjunto dos sete volumes de contos publicados em livro por Machado de Assis.

uma análise de quatro contos. 1998. Dissertação (Mestrado em Teoria Literária) – Curso de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis, 1998.

¹⁹⁹ “Au surplus ces éléments sont très peu nombreux, une dizaine au maximum, dont le système apparaît bien pauvre si on le compare aux mille ressources dont dispose le langage parlé pour marquer les intonations, les pauses, les accents, la mélodie, la mimique. Et cet appauvrissement représente un avantage méthodologique. [...] L’abstraction et la simplification du codage écrit, par quoi l’on transpose l’accompagnement métalinguistique du discours parlé, permettent en outre, en réduisant les variétés étudiées, d’augmenter les effectifs. Et la statistique se complaît dans le désert des grands nombres.” BRUNET, Etienne. *La punctuation et le rythme du discours...*, p. 2.

²⁰⁰ “[...] les séparateurs de phrase se confondent avec les ponctuations fortes [...]” BRUNET, Etienne. *Le vocabulaire de Proust...*, v. 1, p. 122.

Tabela 3 – Pontuação forte nos volumes de contos de Machado de Assis

Texto/Sinal	.	?	!	Total
CFLU	3.767	633	358	4.758
HMN	2.377	310	245	2.932
PA	2.528	217	235	2.980
HSD	2.711	252	132	3.095
VH	2.534	232	131	2.897
PR	1.363	157	80	1.600
RCV	1.792	126	58	1.976
Total	17.072	1.927	1.239	20.238

Tabela 4 – Percentual de pontuação forte pelo número de ocorrências dos volumes

Texto	N	Pontuação Forte	%
CFLU	69.439	4.758	6,8520
HMN	50.106	2.932	5,8515
PA	65.716	2.980	4,5447
HSD	63.233	3.095	4,8945
VH	55.570	2.897	5,2132
PR	28.186	1.600	5,6765
RCV	36.595	1.976	5,3996

Os números brutos mostram uma variação para baixo da pontuação forte, mas tal variação é pouco relevante em termos percentuais, pouco mais de 2 pontos de diferença entre o maior e o menor, e desvia pouco da média (em torno de 5,6%). Embora seja uma opção de relativização, os percentuais são muitas vezes enganadores; por isso, adotamos um modelo estatístico²⁰¹ que nos possibilita o cálculo dos valores teóricos (relativos à esperança matemática) para serem comparados aos valores reais (observados), a fim de obtermos os desvios²⁰².

Vamos primeiro ao Quadro 12, com a representação dos desvios entre o valor real e o valor teórico de cada categoria, dados que servirão posteriormente ao teste do qui-quadrado.

²⁰¹ Neste caso, a hipótese é nula, a qual toma por situação ideal que a distribuição das categorias se dê de modo aleatório, sem que o uso de uma implique o uso de outra. Em comparação a esse modelo é que se pode verificar, por meio de testes estatísticos, se as variáveis estão associadas ou se são independentes. Se a probabilidade de a hipótese nula se concretizar for fraca, há um fato linguístico ou estilístico a ser investigado. Ver MULLER, op. cit., 1968, p. 43-46 e MACIEL, op. cit., p. 40-41.

²⁰² Fórmula de cálculo do valor teórico: $E = (\text{total da linha}) \times (\text{total da coluna}) / (\text{total geral})$. Ver BARBETTA, op. cit., p. 249.

Volume	Efetivo Real	Efetivo Teórico	Desvio Absoluto
CFLU	4.758	3.810	+948
HMN	2.932	2.749	+183
PA	2.980	3.606	-626
HSD	3.095	3.470	-375
VH	2.897	3.049	-152
PR	1.600	1.547	+53
RCV	1.976	2.008	-32

Quadro 12 – Pontuação forte em cada um dos sete volumes publicados com valores reais (observados), teóricos e desvios (números relativos)

O Gráfico 29 apresenta a representação visual desses desvios.

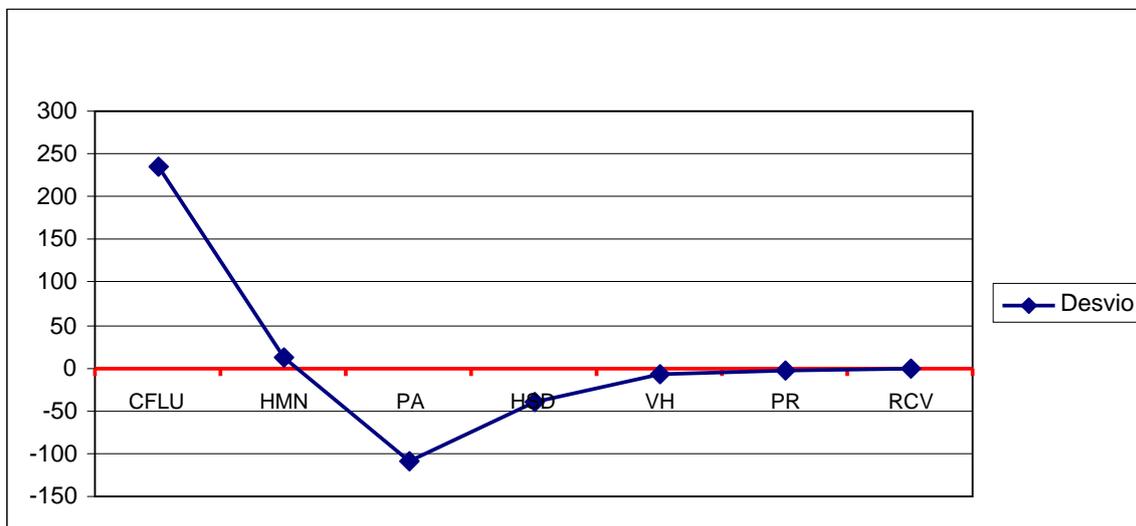


Gráfico 29 – Desvio da pontuação forte nos sete volumes publicados

A distância entre CFLU e PA não parecia tão grande em percentual, mas, ao calcular o número teórico, ou o número esperado a partir da estatística do todo, vemos que há grandes variações tanto para mais quanto para menos, que cessam em VH e ficam em equilíbrio dali em diante.

Um modo ainda mais eficaz de avaliar essas distâncias nos é dado pelo teste do qui-quadrado²⁰³ (X^2), que, conforme Barbetta, é “uma espécie de medida de distância entre as

²⁰³ “O teste de associação qui-quadrado é o teste estatístico mais antigo e um dos mais usados em pesquisa social. É um método que permite testar a significância da associação entre duas variáveis qualitativas.” Ver BARBETTA, op. cit., p. 246.

$$\text{Fórmula do qui-quadrado: } X^2 = \sum \frac{(o-c)^2}{c},$$

em que: Σ = soma; o = valor observado, ou real; c = valor calculado, ou teórico.

freqüências observadas”. O cálculo indica se a probabilidade de a hipótese nula se concretizar é forte ou não. Para sua análise, é preciso levar em conta os graus de liberdade (gl), que em nosso caso são 6 gl. Se a soma do X^2 atingir o valor de 12,59 para uma margem de erro de 5%²⁰⁴, confirma-se a hipótese nula. Como o qui-quadrado ultrapassou muito esse valor, verificamos que a probabilidade diminuiu. Se a probabilidade $p \leq \alpha$ (α = margem de 0,05 ou 5%), o teste rejeita a hipótese nula (H_0), logo as variáveis estão *associadas*²⁰⁵. A aceitação da hipótese nula aponta para uma distribuição aleatória das variáveis; sua rejeição significa que as variáveis estão ligadas por alguma razão, seja estilística ou temática, para que esses números estejam dessa forma distribuídos.

Tabela 5 – Teste de Associação Qui-quadrado (X^2) por volume (variáveis: cronologia e pontuação forte)

Volume	Número teórico	X^2
CFLU	3.810	235,8696
HMN	2.749	12,1486
PA	3.606	108,5920
HSD	3.470	40,4247
VH	3.049	7,5821
PR	1.547	1,8490
RCV	2.008	0,5073
Total	20.238	406,9733

Conforme a Tabela 5, como a soma do qui-quadrado resultou em um valor alto (que o afasta muito da hipótese nula), há o que investigar em termos estilísticos ou temáticos. Outro dado a ser observado é a queda brutal que acontece ao longo do tempo, que o contraste entre CFLU e RCV deixa evidente.

Vejamos nas Tabelas 6 e 7 como a pontuação se comporta nos quatro grandes grupos, divididos também por ordem cronológica, tomados dois a dois (Flumeia e AVG1 até 1880; e 5VOL e AVG2 de 1881 em diante).

Tabela 6 – Pontuação forte nos volumes de contos agrupados de Machado de Assis

Grupo/Sinal	.	?	!	Total
FLUMEIA	6.052	943	603	7.598
AVG1	28.113	4.247	2.856	35.216
5VOL	10.927	984	636	12.547
AVG2	10.754	1.191	664	12.609

²⁰⁴ Ver tabela de distribuição do X^2 em BARBETTA, op. cit., p. 326.

²⁰⁵ BARBETTA, op. cit., p. 252.

Tabela 7 – Pontuação forte nos volumes de contos agrupados de Machado de Assis

Grupo	N	Pontuação forte	%
FLUMEIA	119.545	7.598	6,3557
AVG1	547.611	35.216	5,8830
5VOL	249.300	12.547	5,0328
AVG2	230.788	12.609	5,4634

Podemos perceber que a variação também não parece muito significativa. Vamos aos números relativos, na Tabela 8.

Tabela 8 – Pontuação forte em todos os contos de Machado de Assis, por grupos: valores reais, teóricos e desvios (números relativos)

Grupo	Efetivo real	Efetivo teórico	Desvio absoluto
Flumeia	7.598	7.083	+37
AVG1	35.216	32.444	+237
5VOL	12.547	14.770	-335
AVG2	12.609	13.693	-83

Aqui o que fica mais evidente é que, na fase inicial do autor, a pontuação forte excedia a expectativa matemática, enquanto de 1881 em diante ela decresce sensivelmente. Nos conjuntos, tomados por época, fica mais evidente a diminuição da pontuação forte, o que aponta para a adoção de frases mais longas por parte do autor.

Tabela 9 – Teste de Associação Qui-quadrado (X^2) por grupo (variáveis: cronologia e pontuação forte)

Grupo	Número Teórico	X^2
Flumeia	7.083	37,5052
AVG1	32.444	236,8484
5VOL	14.770	334,6097
AVG2	13.673	82,8492
Total	67.970	691,8124

A soma do qui-quadrado (Tabela 9) resultou em 691, com 3 gl; para 5%, o valor é de 7,81. Logo $p < \alpha$. O teste rejeita a hipótese nula (H_0) e confirma o resultado por volume: as variáveis estão *associadas*²⁰⁶. O teste do qui-quadrado confirma que a distribuição da pontuação forte não é aleatória. Uma explicação possível estaria relacionada à diminuição do uso do discurso direto.

²⁰⁶ Ver BARBETTA, op. cit., p. 252.

Nos primeiros contos de Machado, como *O segredo de Augusta* (contido em CFLU), predominam os diálogos, que, muitas vezes, ocupam uma longa sequência de páginas. Com o fortalecimento do narrador e com uma prosa mais narrativa, que incorpora a fala das personagens ao discurso do narrador, seja por discurso direto ou discurso indireto livre, Machado acaba por reduzir expressivamente o uso do diálogo (direto). Como afirma Maciel:

La structure du texte conditionne la longueur moyenne de la phrase et, par voie de conséquence, intervient dans la distribution des catégories grammaticales en plus d'avoir une relation directe avec la répartition des signes de ponctuation.

[...] Là où prédomine le dialogue, la phrase sera plus courte; la phrase sera plus longue lorsqu'il y aura davantage de passages non-dialogués, c'est-à-dire essentiellement descriptifs ou narratifs avec des répercussions sur l'usage des signes de ponctuation²⁰⁷.

Se a estrutura do texto condiciona a extensão de frase, pode-se afirmar que esta está relacionada com as opções estilísticas do autor. Analisemos uma estimativa²⁰⁸ da extensão média de frase por volume publicado, obtida a partir da pontuação forte.

Texto	Pontuação forte	Ocorrências	Extensão*
CFLU	4.758	69.439	12,89
HMN	2.932	50.106	15,26
PA	2.980	65.716	18,93
HSD	3.095	63.233	17,26
VH	2.897	55.570	16,37
PR	1.600	28.186	15,17
RCV	1.976	36.595	16,09

* Número médio de palavras = número de ocorrências / pontuação forte

Quadro 13 – Extensão média de frase nos volumes de contos de Machado de Assis

Pode-se observar no Quadro 13 um aumento da extensão média que atinge seu ápice em *Papéis avulsos* para logo depois começar um decréscimo gradual até atingir um equilíbrio nos últimos três volumes. Vejamos como se comporta a extensão de frase nos grupos macro.

²⁰⁷ MACIEL, Carlos. *Richesse et évolution du vocabulaire d'Erico Veríssimo...*, p. 55.

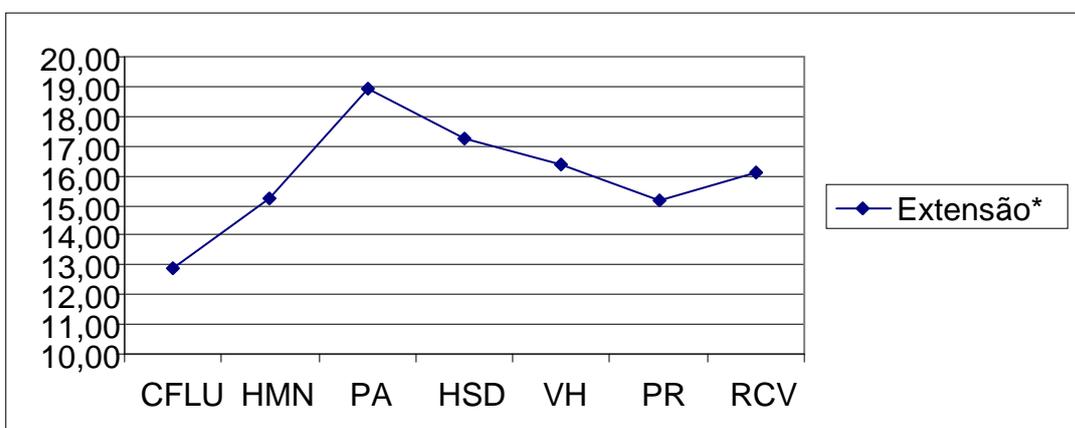
²⁰⁸ O cálculo foi feito subtraindo-se o total de pontuação do total de ocorrências. A partir desse valor líquido fizemos a divisão pela pontuação forte.

Texto	Extensão*
FLUMEIA	13,81
AVG1	13,81
5VOL	17,05
AVG2	15,72
R1	15,99
R2	17,63

* Número médio de palavras por frase = número de ocorrências / pontuação forte

Quadro 14 – Extensão média de frase nos volumes de contos agrupados e romances de Machado de Assis, e no Portext

No Quadro 14 temos um dado interessante: as extensões médias de Flumeia e de AVG1, que compreendem a produção de contos até 1880, são idênticas (13,81); já os 5VOL, comparados a AVG2, têm frases mais longas. Nos romances – R1 e R2 –, a título de comparação, observamos um movimento semelhante, com um aumento da extensão média das frases nos textos a partir de 1881.



* Número médio de palavras por frase = número de ocorrências / pontuação forte

Gráfico 30 – Extensão média de frase nos volumes de contos de Machado de Assis

Podemos constatar a partir desses dados o processo de alongamento das frases à medida que o tempo passa (Gráfico 30). Antes de tentarmos complementar esses dados com mais dados relativos à pontuação, vamos verificar como esse alongamento da frase está relacionado ao tamanho do texto. Será que esse crescimento da frase é proporcional ao da extensão do texto?

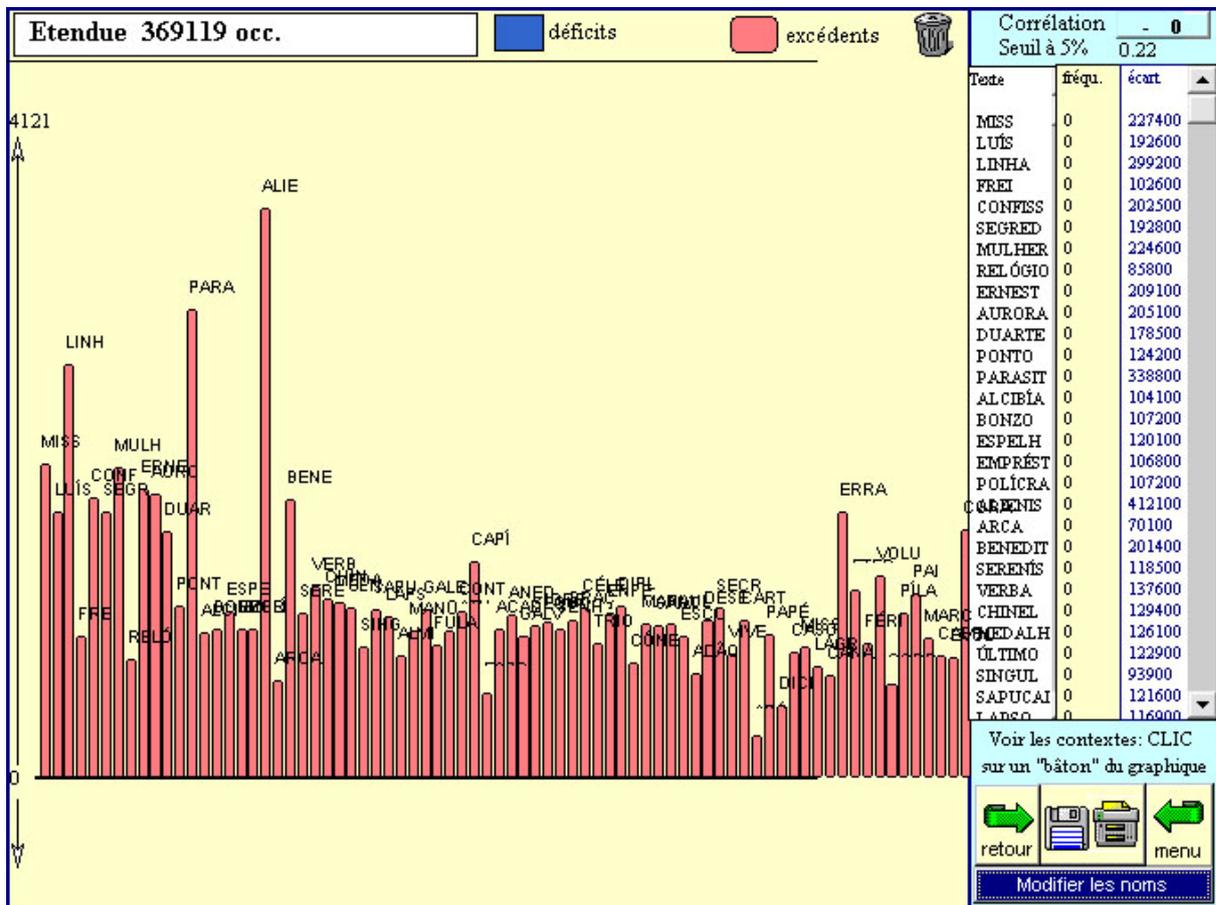


Gráfico 31 – Extensão dos textos, calculada sobre as ocorrências (N) em 7VOL

O Gráfico 31 mostra com nitidez a diminuição da extensão dos contos de Machado à medida que o tempo passa. Mesmo no início, a quantidade de textos longos é limitada e conhecemos as controvérsias sobre o gênero de *O alienista*, para não tomá-lo como parâmetro. O que mais chama a atenção é a regularidade do autor quanto à extensão dos contos após *D. Benedita*, ainda em PA, em 1882. Depois disso, Machado produziria contos até 1906; isso denota que, assim que alcançou um equilíbrio quanto à extensão, o autor a manteve até o último conto, salvo duas ou três exceções. A redução da extensão do texto confirma a concisão machadiana, sempre realçada pela crítica.

Pela leitura dos contos do autor seguindo a cronologia de sua produção, podemos notar que a extensão média de frase é inversamente proporcional à extensão dos textos. Os primeiros contos são mais longos e têm forte ênfase na cena e no diálogo, provável influência do teatro na formação de Machado.

Buscaremos nos dados a seguir sobre o ponto e, posteriormente, sobre a vírgula observar se há uma confirmação do alongamento da extensão média das frases, seguindo o critério cronológico.

6.1.1 O ponto

Agora nos concentramos no ponto, a fim de verificar se o movimento dele segue o do conjunto dos sinais de pontuação forte. Para obter uma ideia mais clara do movimento e visualizar melhor as mudanças ao longo da produção do autor, optamos por centrar a atenção sobre os grupos macro que compõem o *corpus*.

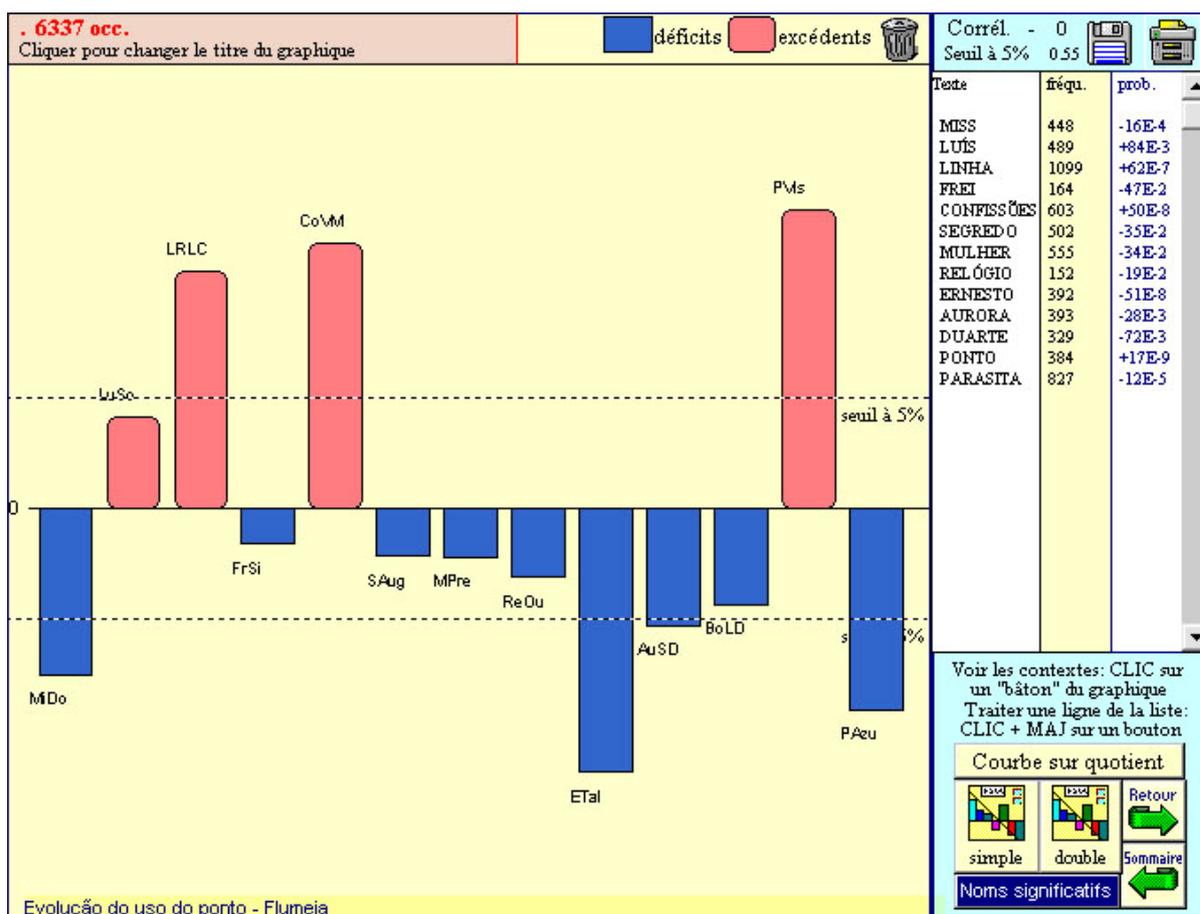


Gráfico 32 – Evolução do uso do ponto em Flumeia

A evolução do uso do ponto cronologicamente²⁰⁹ descreve uma curva ascendente no primeiro volume (CFLU) (Gráfico 32). Dos sete contos que o compõem, apenas um é de fato deficitário MiDo (*Miss Dollar*), que é o mais narrativo do volume, fugindo aos excessivos diálogos de contos como o SAug (*O segredo de Augusta*), estando abaixo da média que compreende a margem de erro. No segundo volume (HMN), que começa no meio do gráfico (ReOu), podemos observar certo desequilíbrio, pois, dos seis, um tem excedente e dois têm déficits importantes, o que já aponta para uma mudança na estrutura de frase.

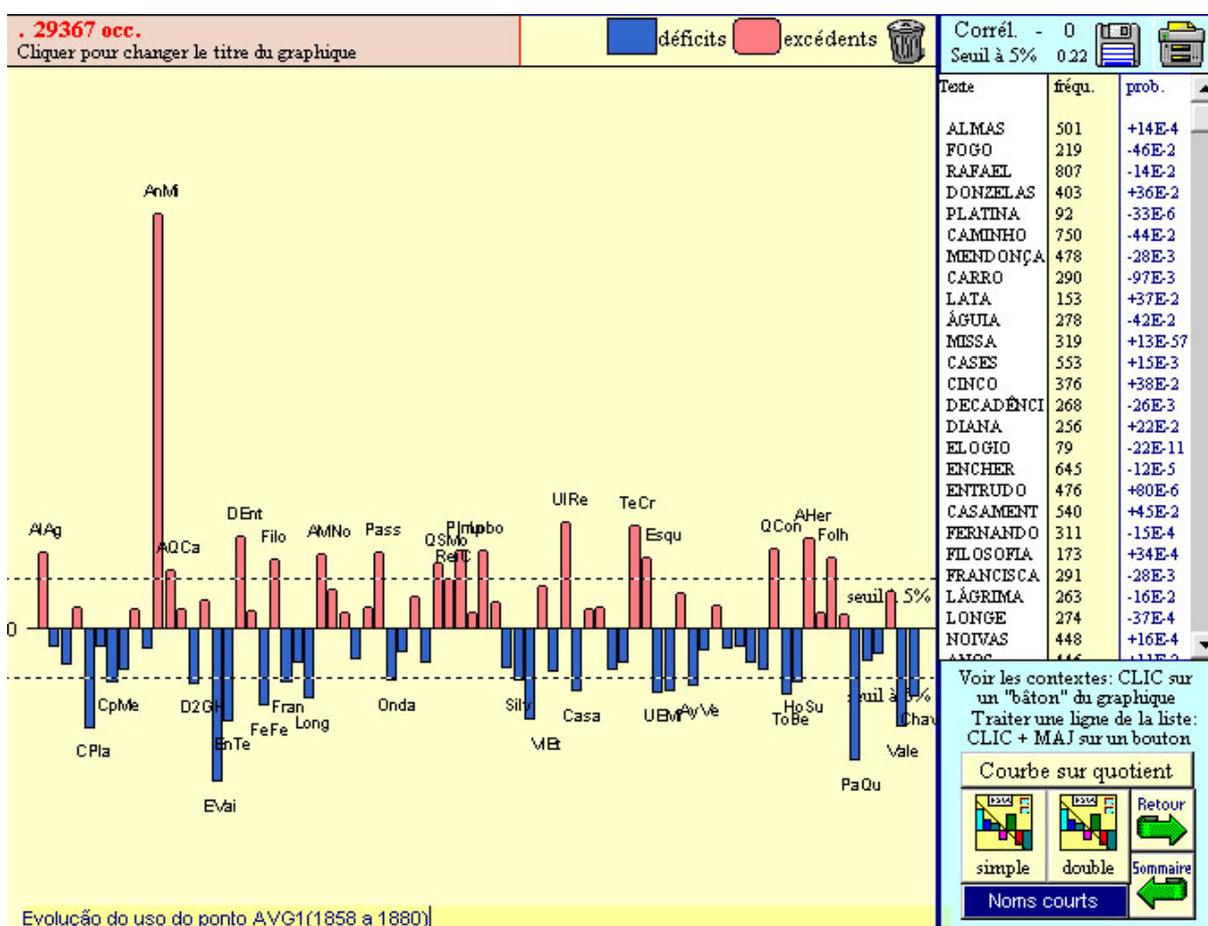


Gráfico 33 – Evolução do uso do ponto em AVG1

Pode-se observar a semelhança nos movimentos dos Gráficos 33 e 32 quanto à maior concentração dos excedentes na primeira metade. Porém, a relação entre déficits e excedentes

²⁰⁹ As bases formadas pelos volumes publicados estão em ordem de publicação, em ordem cronológica. Nesse caso, em que a base é constituída por CFLU, 1870, e HMN, 1873, como o primeiro volume contém sete contos, os primeiros sete que aparecem no gráfico pertencem a esse volume e assim sucessivamente. No entanto, a ordem de nosso arquivo de texto (ancorado nas *Obras completas* de Machado de Assis, da Nova Aguilar) que gerou a base não é mantida pelo programa, que a altera automaticamente, sem, no entanto, perder o viés cronológico.

é mais equilibrada, com ambos quase se espelhando. A única exceção é AnMi (*Antes da missa*). O que justifica o uso excessivo do ponto é o fato de ser um conto todo em forma de diálogo entre duas amigas, à espera da hora da missa, como está no título.

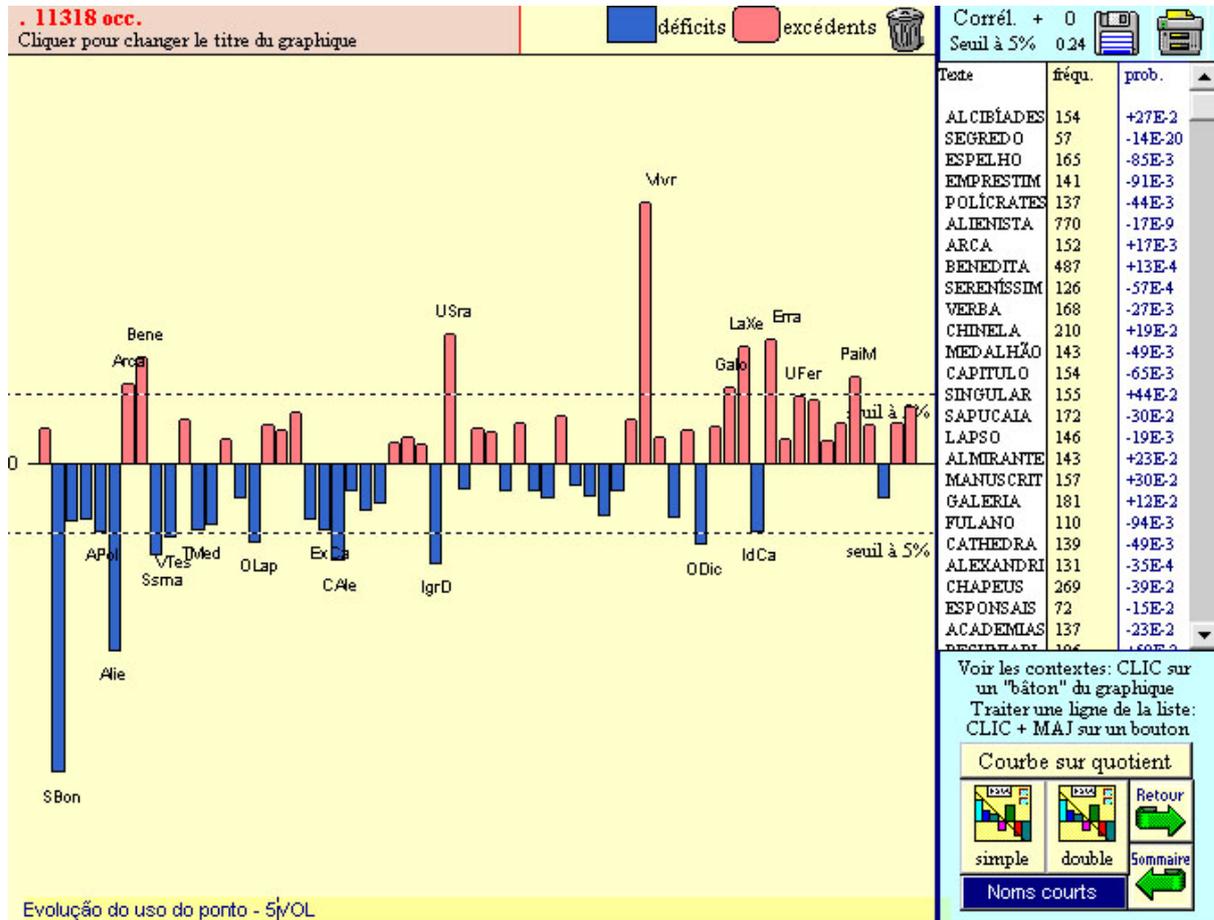


Gráfico 34 – Evolução do uso do ponto em 5VOL

Os cinco volumes dessa base são os publicados posteriormente a CFLU e a HMN, ou seja, a partir de *Papéis avulsos* (1882). No Gráfico 34, temos um equilíbrio entre déficits e excedentes quanto ao uso do ponto, semelhante ao gráfico anterior. Quanto à evolução cronológica, o movimento se assemelha ao da base Flumeia, só que em sentido contrário, tendo um déficit inicial mais significativo, que é abrandado depois da metade do quadro. Quando comparados esses volumes entre si, que já fazem parte da maturidade estilística do escritor, podemos observar um equilíbrio entre o uso do ponto e as outras formas de marcar as pausas da narrativa.

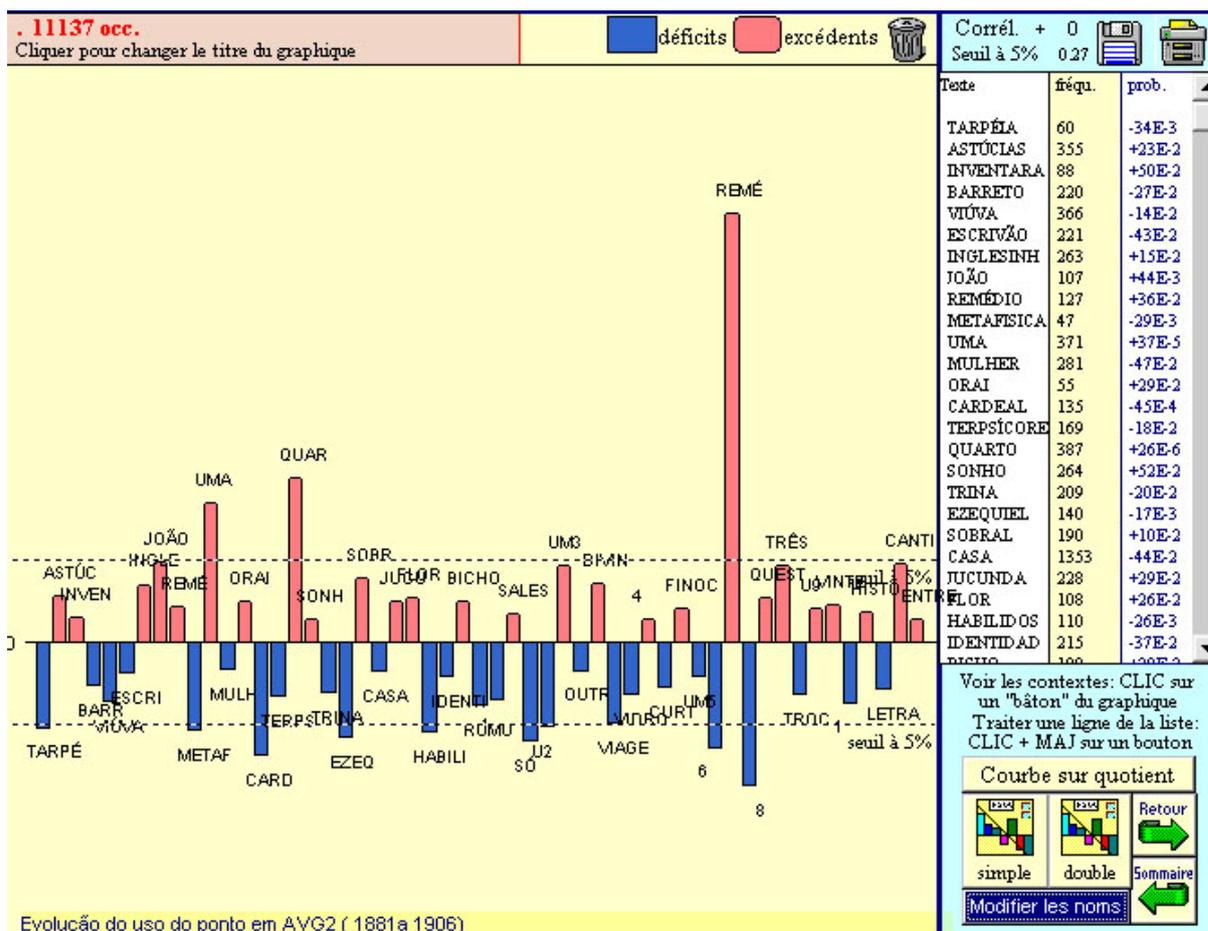


Gráfico 35 – Evolução do uso do ponto em AVG2

Continuemos em ordem cronológica observando os contos avulsos (não-publicados em livro pelo autor) da mesma época dos cinco volumes do gráfico anterior.

Novamente há um equilíbrio entre déficits e excedentes. Cronologicamente, no Gráfico 35 o movimento é mais sutil do que o anterior, com poucos contos ultrapassando a média, seja para cima ou para baixo. A exceção é o conto *O melhor remédio*, de 1884, outro texto todo em forma de diálogo, o que justifica sua posição.

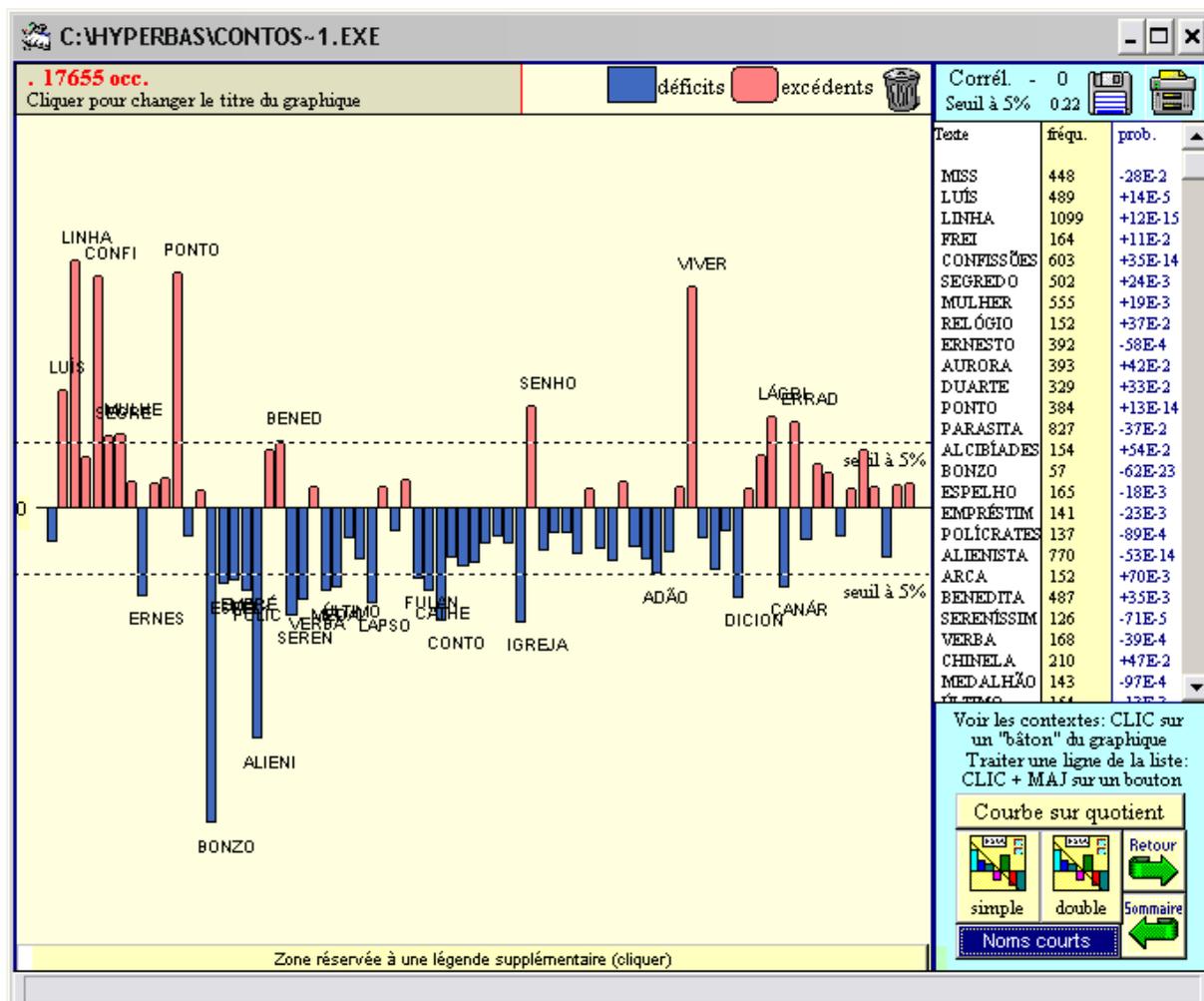


Gráfico 36 – Evolução do uso do ponto em 7VOL

No Gráfico 36, que agrupa os sete volumes que unem Flumeia e 5VOL²¹⁰, podemos observar que se mantém um equilíbrio quantitativo, embora os excedentes de Flumeia ganhem em intensidade. Em termos de evolução cronológica, há uma semelhança maior com o Gráfico 38, onde a maior concentração de excedentes aparece na primeira parte do quadro, para depois haver uma baixa expressiva no uso do ponto até atingir um equilíbrio em que prevalece a média. Tomados os sete volumes juntos, fica bem mais clara a diminuição gradativa do uso do ponto, que aponta para um alongamento das frases usadas pelo autor, embora não haja nenhuma indicação de mudança radical nesse aspecto.

A título de comparação vamos observar o uso do ponto nos romances.

²¹⁰ Infelizmente, não podemos fazer o mesmo quanto a AVG1 e AVG2, ou mesmo quanto ao total dos contos, pois o Hyperbase tem 76 como o número máximo de textos para a mesma base.

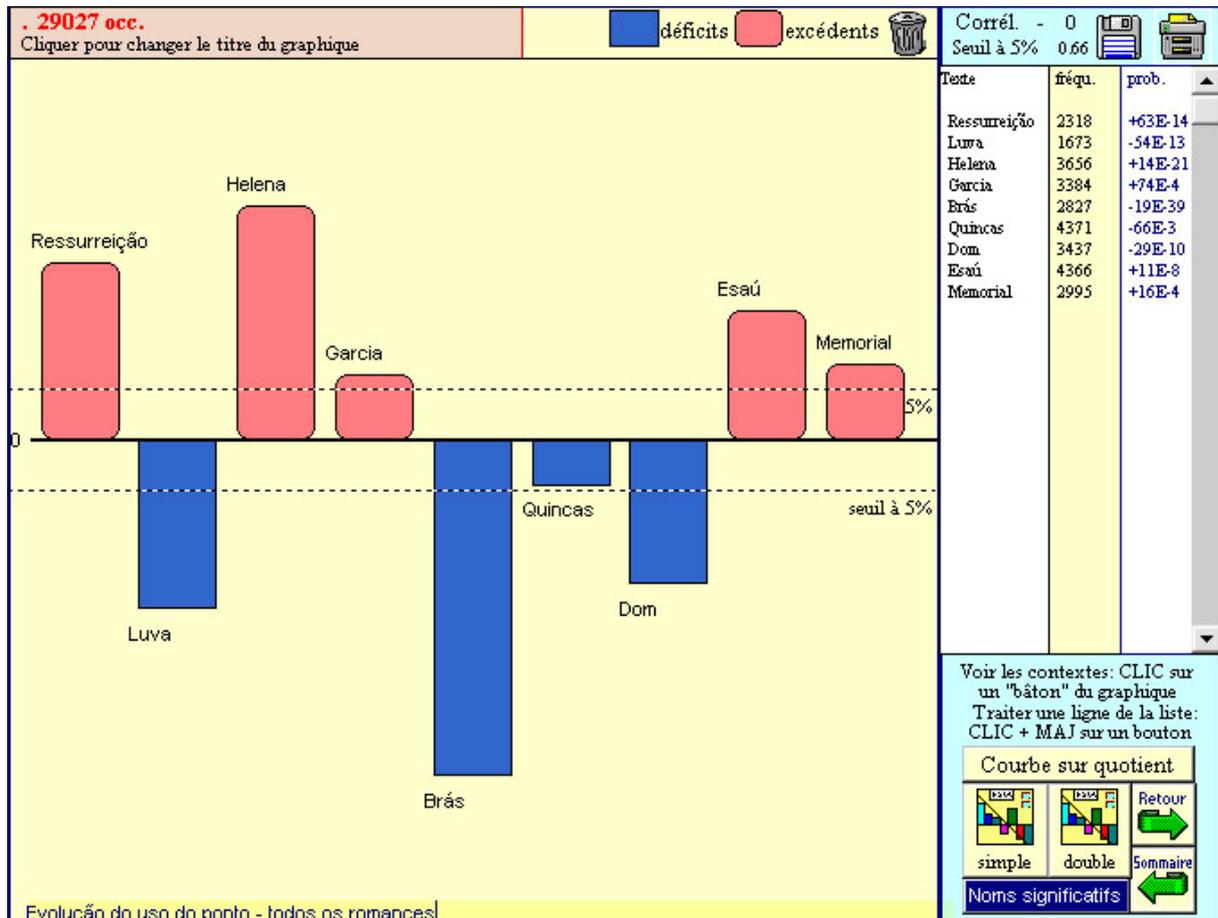


Gráfico 37 – Evolução do uso do ponto nos romances

Embora a distribuição do ponto seja mais radical nos déficits e excedentes do que nos contos, o movimento cronológico é muito semelhante, com muita força no início, contrastando com uma baixa nos três principais romances, para retomar o crescimento nas últimas obras. O Gráfico 37 nos induz a, mais uma vez, tomar o alongamento das frases em *Brás Cubas*, *Quincas Borba* e *Dom Casmurro* como uma explicação possível para esse movimento.

Para reforçarmos a nossa hipótese do alongamento da frase, não basta observar apenas o ponto. Como a vírgula aparece como a forma em maior progressão de uso no conjunto dos sete volumes publicados, por ser uma das principais ferramentas para o alongamento das frases, é ela que observaremos a seguir.

6.1.2 A vírgula

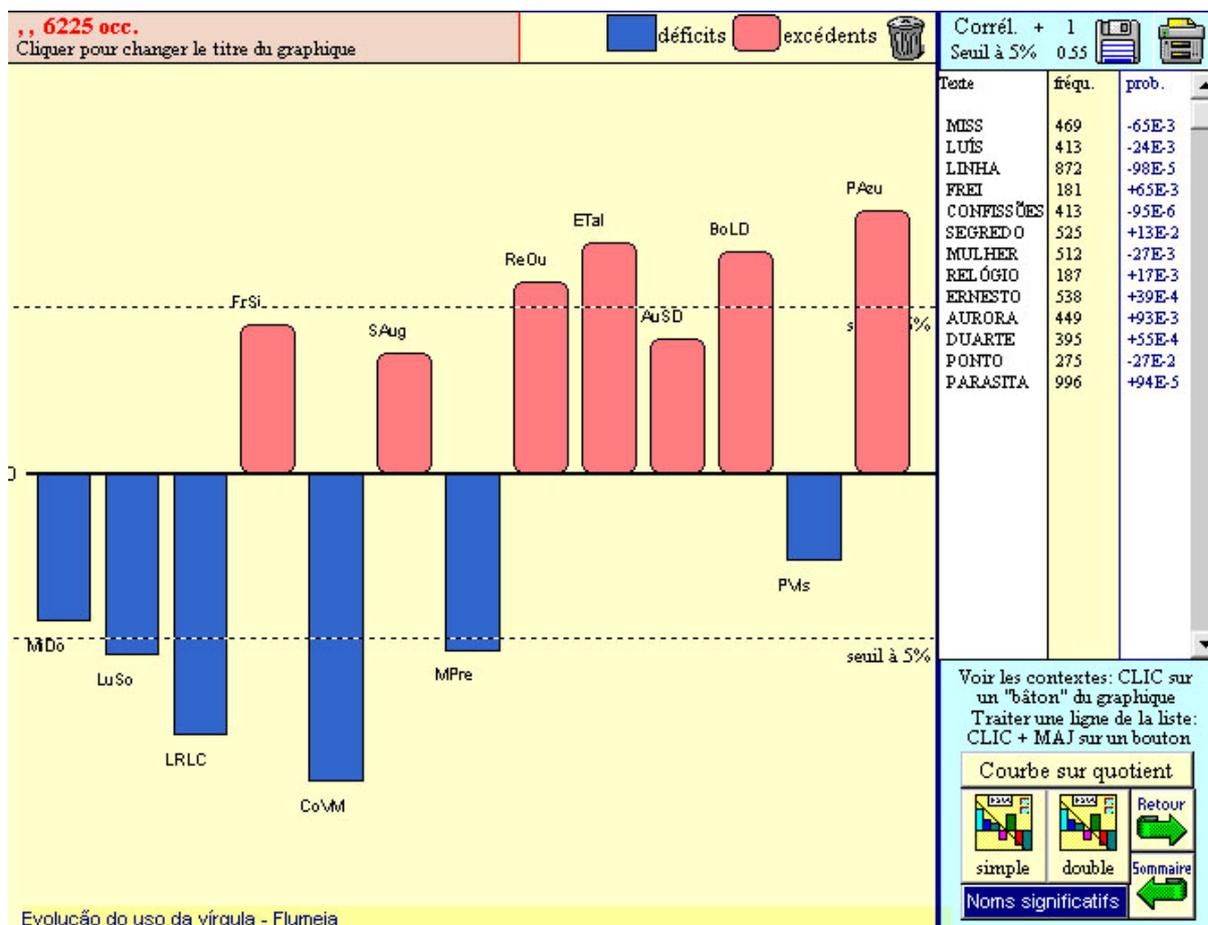


Gráfico 38 – Evolução do uso da vírgula em Flumeia

Em relação ao ponto, pudemos observar que tanto as diferenças quantitativas quanto as cronológicas se acirraram. O crescimento do uso da vírgula, já em Flumeia (Gráfico 38), se mostra forte e não aponta para um recrudescimento ao longo do tempo. Ao contrário, mantém a curva ascendente, embora quantitativamente os déficits e os excédentes estejam equilibrados. A frase mais longa, o texto mais narrativo vai se fixando como opção de Machado e sua repetição ao longo da linha do tempo vai consolidando essa forma como um padrão estilístico, uma marca do autor.

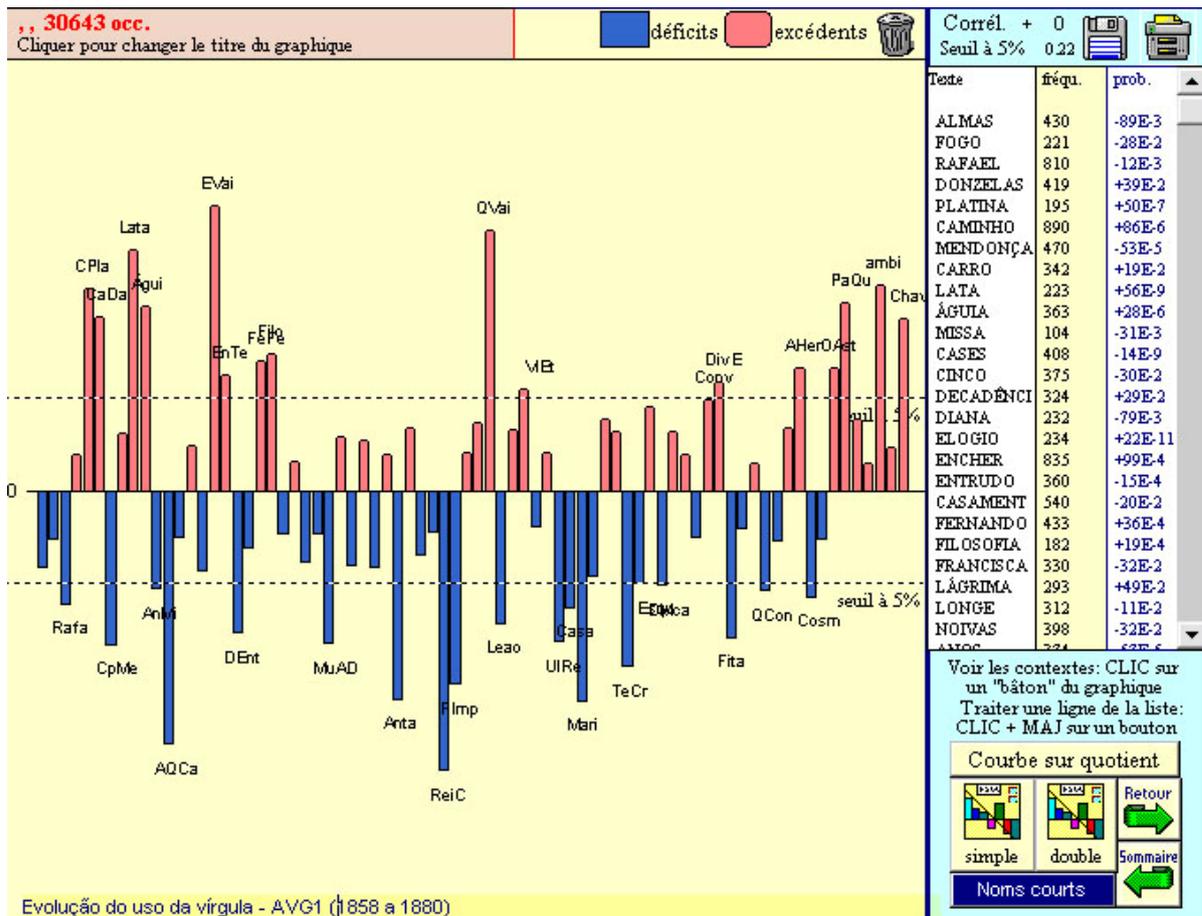


Gráfico 39 – Evolução do uso da vírgula em AVG1

As diferenças em AVG1 não são tão nítidas quanto em Flumeia, mas cronologicamente, enquanto os déficits se encaminham em direção à média, os excedentes continuam significativos até o fim do período (Gráfico 39), ou seja, pode-se observar uma tendência, em termos de cronologia, ao alongamento da frase machadiana. Em termos quantitativos, mantém-se um equilíbrio entre ambos.

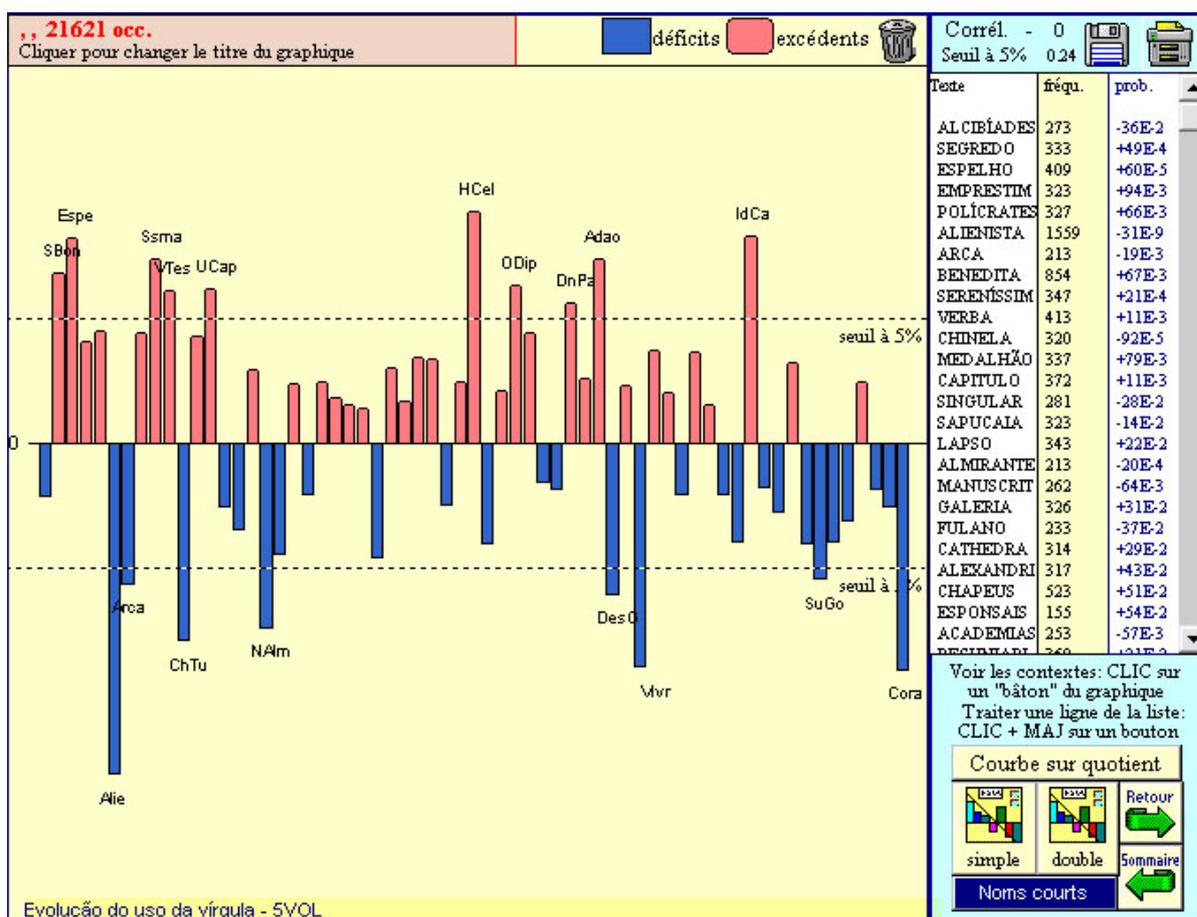


Gráfico 40 – Evolução do uso da vírgula em 5VOL

No grupo 5VOL (Gráfico 40), já há diferenças expressivas quanto ao uso da vírgula, que predomina na média em quase todo o quadro, com uma queda maior nos últimos contos. Nesse grupo, se comparado a seu gráfico relativo ao uso do ponto (Gráfico 40), podemos observar que a vírgula aparece com mais força em um maior número de casos. Embora seja mais evidente no início da linha do tempo (*Papéis avulsos*, 1882), há uma continuidade.

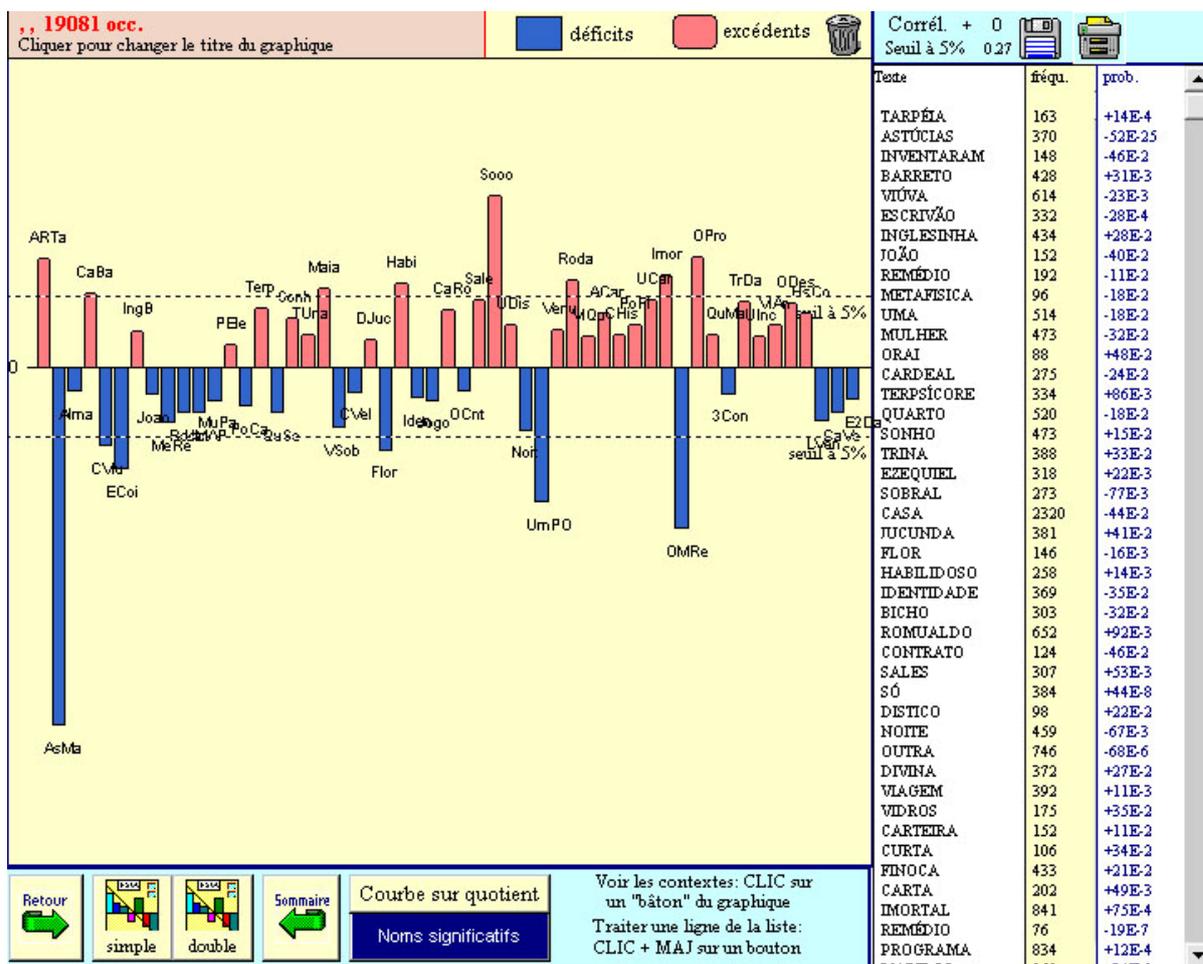


Gráfico 41 – Evolução do uso da vírgula em AVG2

Em termos gerais, em relação à quantidade de déficits e excedentes, AVG2 é muito semelhante a seu contemporâneo 5VOL (Gráfico 41). No entanto, é importante observar que cronologicamente há um fortalecimento da vírgula, a mesma tendência de formação de padrão que vimos em gráficos anteriores. Não há tantos excedentes, mas abaixo da média há uma redução expressiva dos déficits em relação a Flumeia e AVG1, que vão até 1880.

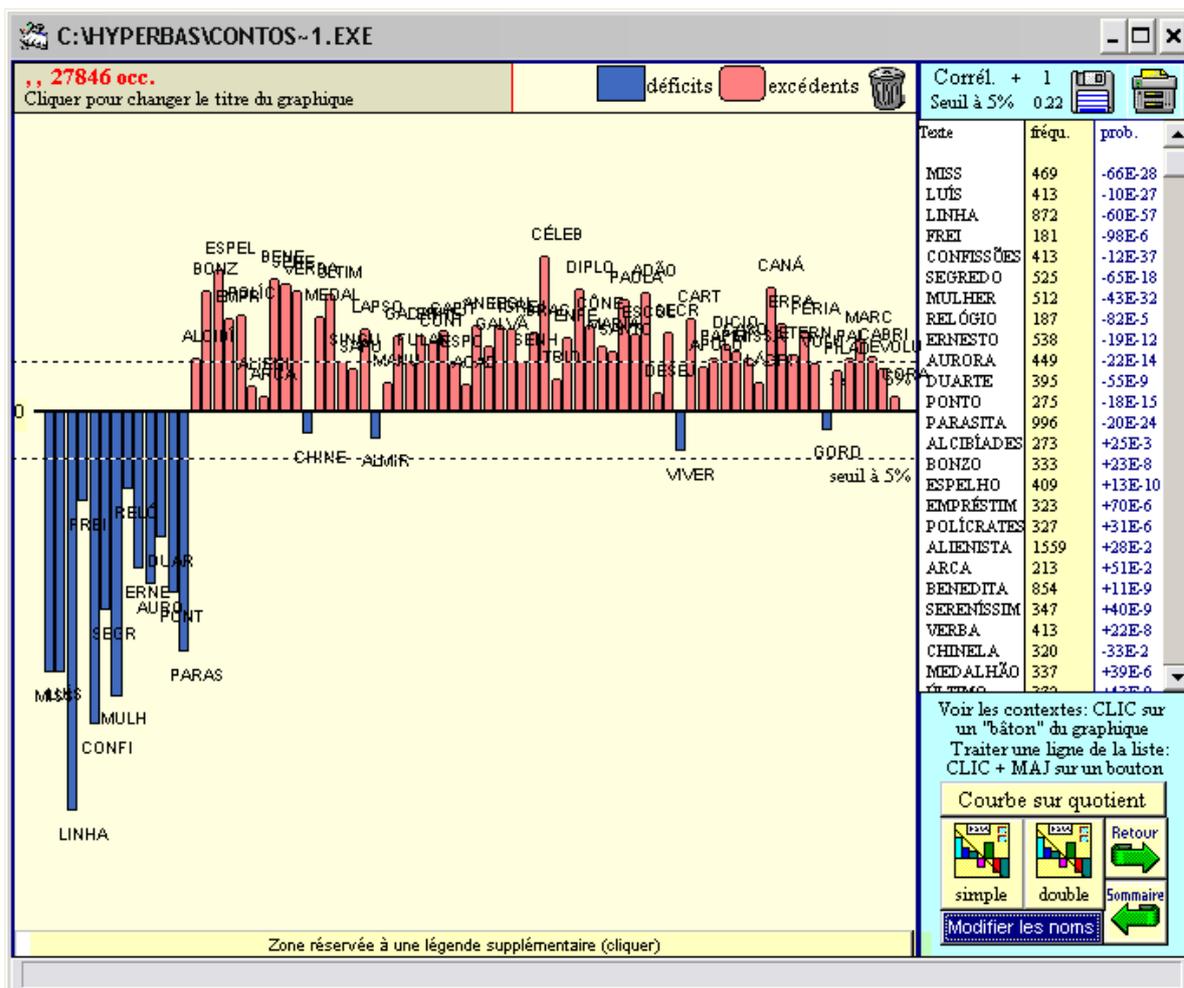


Gráfico 42 – Evolução do uso da vírgula em 7VOL

O Gráfico 42 praticamente fala por si. Ele é o único dado que obtivemos em nossa pesquisa que aponta para uma ruptura: treze são os contos que formam a base Flumeia. Treze são as barras que nos levam à data fatídica de *Papéis avulsos*, que, na ordem imposta pelo Hyperbase, começa em *Uma visita de Alcibíades*. O ano? 1882.

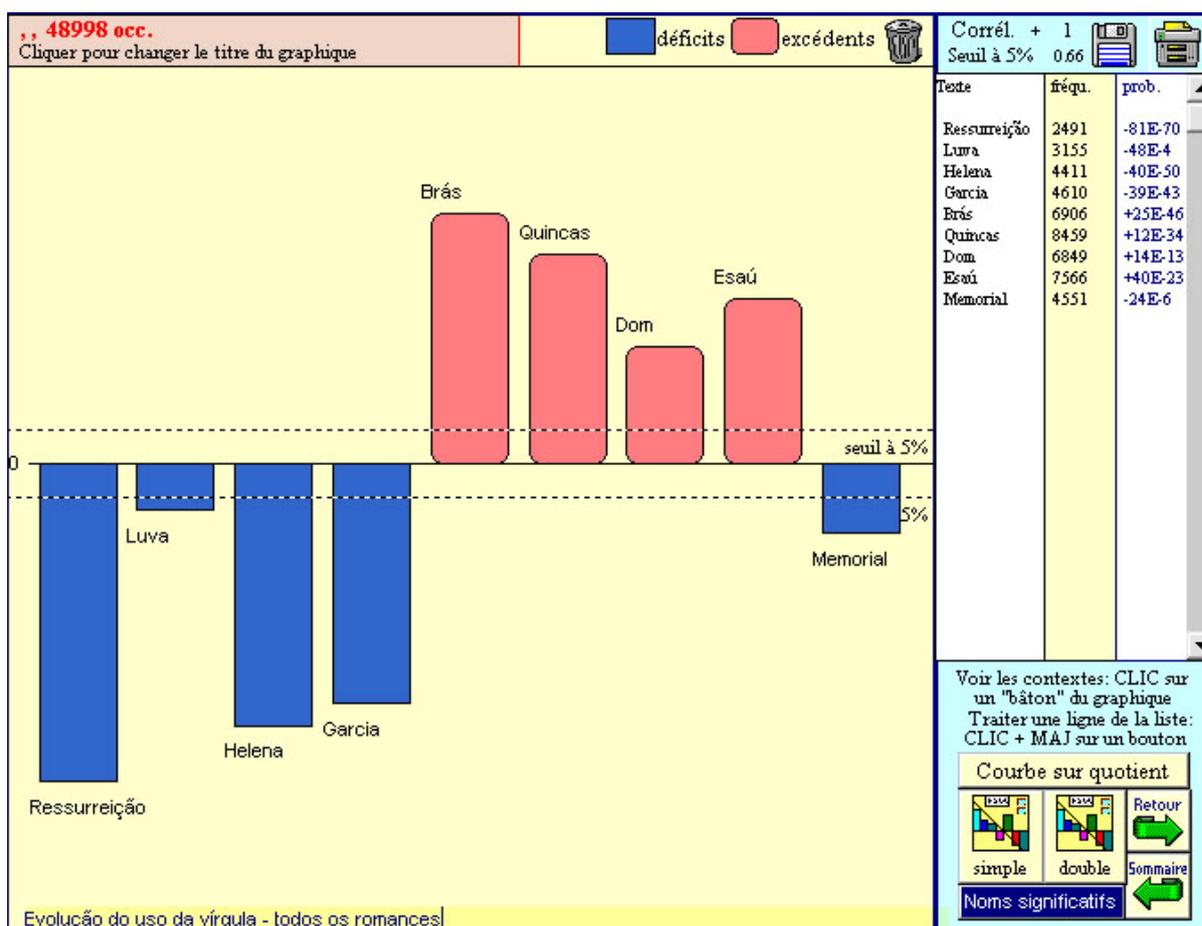


Gráfico 43 – Evolução do uso da vírgula nos romances

A vírgula definitivamente está do lado da crítica, como reforça o Gráfico 43. Mas Aires, o velho diplomata, é conservador.

Desde o início deste capítulo, os dados nos remetem a um alongamento das frases no conto machadiano. Se o uso do ponto não foi explícito o suficiente sobre essa mudança de estruturação do texto, a vírgula não deixou dúvidas, principalmente quando analisada nos grandes conjuntos. Vejamos como essa mudança estrutural repercute na escrita de um autor, a partir do que diz Maciel:

Si la phrase est plus longue, l'auteur a utilisé davantage de mots relationnels, qui s'imposent lorsque la phrase s'allonge, et aussi éventuellement davantage de virgules. Si la phrase est courte, l'auteur fait moins souvent appel aux mots-outils; il utilise plus fréquemment des structures fondamentales de la langue [...] et sans doute y trouvera-t-on plus de point.

La structure du texte conditionne la longueur moyenne de la phrase et, par voie de conséquence, intervient dans la distribution des catégories grammaticales en plus d'avoir une relation directe avec la répartition des signes de ponctuation²¹¹.

²¹¹ MACIEL, op. cit., p. 55.

A relação aqui explicitada entre pontuação, extensão de frase e distribuição das classes gramaticais é bastante útil para o estudo da contística machadiana. A constatação de um alongamento da frase, como vimos aqui, se associada à informação da alta incidência de palavras funcionais numa curva ascendente cronologicamente, nos remete a uma reestruturação da escrita de Machado de Assis, já que essas distribuições, como afirma Maciel, são condicionadas pelo modo como o texto se estrutura.

Tal reestruturação, no entanto, se configura gradualmente, num movimento ascendente, para, nas últimas obras, retomar algumas características das primeiras, prevalecendo a nostalgia de quem já está em tempo de escrever memórias.

CAPÍTULO 7 DISTÂNCIA LEXICAL

Pode-se medir a distância lexical entre dois ou mais textos? Que parâmetros usar? Quais os critérios para escolher as variáveis a serem medidas? Como em muitas das ciências humanas ou biológicas, precisamos medir, contar e calcular probabilidades para que possamos concluir desde os riscos de uma doença até a probabilidade de um candidato vencer uma eleição. Com o advento da computação e de programas como o Hiperbase, temos a oportunidade de calcular a distância entre o vocabulário de dois ou mais textos, o conceito a que nos referimos no título deste capítulo.

Vários são os métodos já testados a fim de se estabelecer a distância lexical²¹². Em geral, o estudo da distância tem sido feito sobre os valores de V (palavras diferentes ou formas) sem levar em conta os efetivos ou ocorrências (N). Leva-se em conta apenas a presença ou a ausência de determinada forma, sem se considerar seu efetivo, diferentemente do que ocorre com o conceito de distância intertextual, cujo cálculo é feito sobre o conjunto dos textos (N), e não sobre seu vocabulário (V). Ou, ainda, da conexão lexical, complementar da distância, que seria a intersecção do vocabulário de dois textos²¹³.

Infelizmente persistem problemas técnicos que inviabilizam a utilização de recursos do Hiperbase (ao menos a versão por nós utilizada). Dada a grande extensão do *corpus*, muitas vezes o programa falha e não consegue completar o cálculo dos dados que gostaríamos de extrair. No caso dos contos de Machado, como são muitos e extensos, foi inviável tentar o cálculo sobre o efetivo ou sobre o total de ocorrências, o que nos levaria à distância intertextual, citada acima. Acabamos por usar o cálculo sobre as formas (V).

Como dissemos anteriormente²¹⁴, a análise em árvore é mais completa em relação à análise fatorial, pois esta só abarca uma parte da variância, quando somos obrigados a escolher entre as opções de combinação dos fatores (1 e 2, 1 e 3, 2 e 3), além de permitir uma leitura mais fácil dos dados e das distâncias. Daí nossa opção pela análise em árvore, desenvolvida por Xuan Luong em formato radial no *corpus* de Machado e radial e retangular na base Portext, este último para melhor visualização devido ao grande número de textos desta base.

²¹² Ver BRUNET, Etienne. Peut-on mesurer la distance entre deux textes? *Corpus*, n. 2, déc. 2003. Disponível em: <<http://corpus.revues.org/index30.html>>. Acesso em: 27 jun. 2007.

²¹³ LABBÉ, Cyrill; LABBÉ, Dominique. La distance intertextuelle. *Corpus* [online], n. 2, déc. 2003. Disponível em: <<http://corpus.revues.org/index31.html>>. Acesso em: 27 jun. 2007.

²¹⁴ Ver Capítulo 4, item 4.2.

Como nos capítulos anteriores, seguiremos a apresentação dos gráficos pela ordem cronológica das bases, por gênero.

7.1 O Romance

Começaremos pelos romances, pois, no capítulo anterior, os dados de pontuação apontaram, em termos de estrutura de frase, para uma ruptura que confirma o que é canônico na crítica machadiana: as duas fases.

Vamos aos quadros.

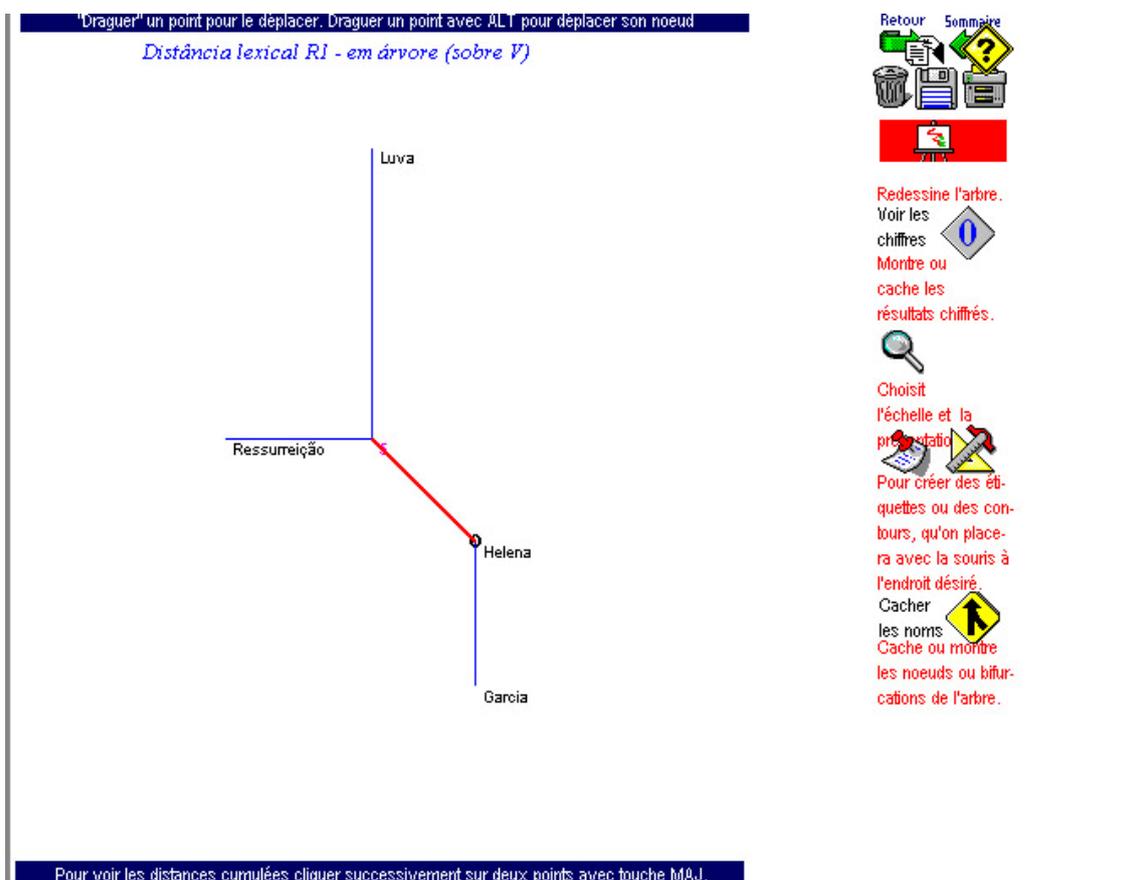


Gráfico 44 – Análise em árvore da distância lexical na base R1

Esse gráfico mostra a distância lexical entre os romances da chamada primeira fase ou, como preferimos, da juventude do autor. Nota-se que o fator cronológico tem determinado peso, já que *Ressurreição*, de 1872, agrupa-se a *A mão e a luva*, de 1874, enquanto *Helena*, de 1876, aproxima-se de *Iaiá Garcia*, de 1878. Pode-se observar ainda que os dois primeiros

romances estão mais distantes um do outro em termos de vocabulário do que os dois últimos, o que nos faz acreditar que o autor estava em processo de estabilização de seu universo lexical.

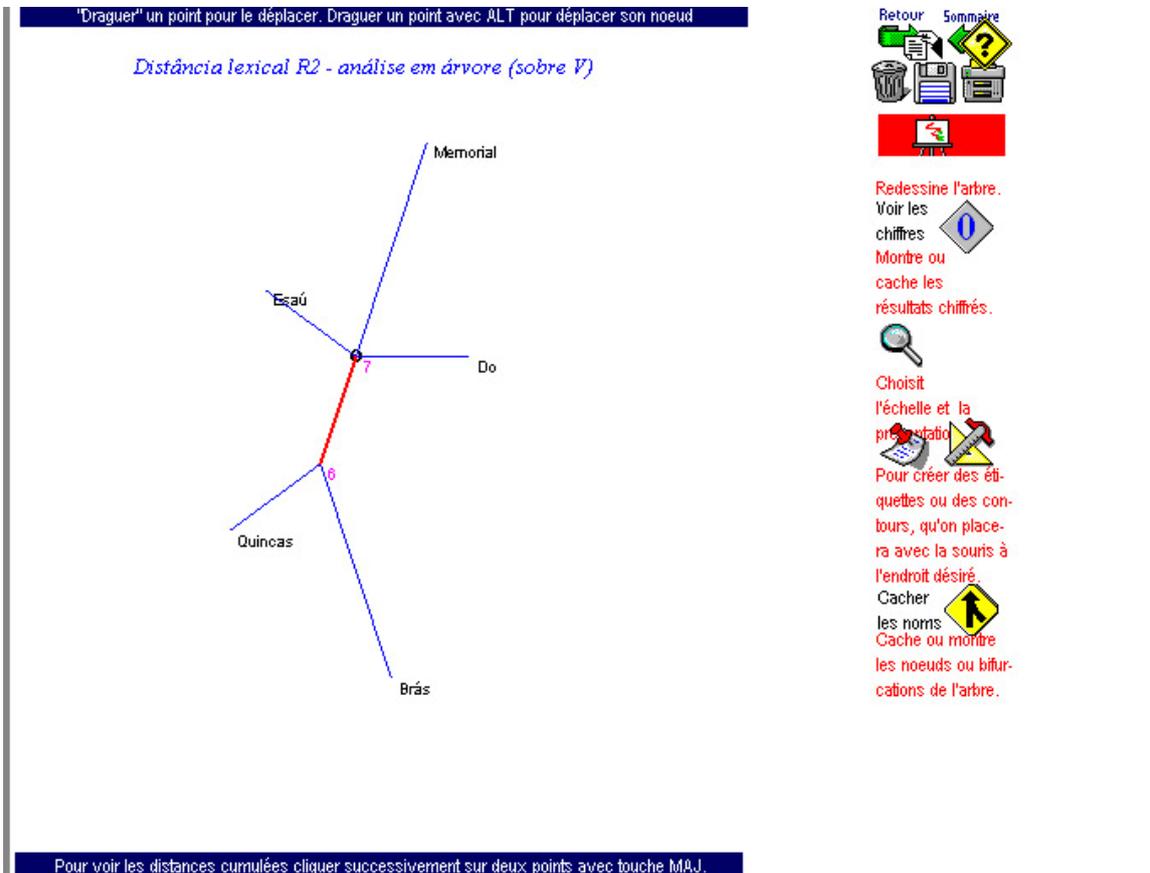


Gráfico 45 – Análise em árvore da distância lexical na base R2

Aqui temos o gráfico dos romances da fase madura. Pode-se observar que tanto o *Memorial* quanto *Brás Cubas* se afastam dos demais em termos de vocabulário – por motivos distintos, afinal os perfis dos narradores Brás e Aires são bastante diferentes, logo seu universo de palavras não pode se assemelhar muito. Mesmo assim é interessante notar, já que o fator gênero costuma prevalecer nas análises de distância lexical, o fato de que ambos são semelhantes no sentido de serem estruturados como autobiografias.

Quanto à separação em dois grupos, parece coerente com o fato de que Quincas Borba e Brás Cubas estejam presos ao mesmo tronco, pois representam a fase mais experimental do romancista Machado, enquanto os outros três são mais conservadores em seu vocabulário.

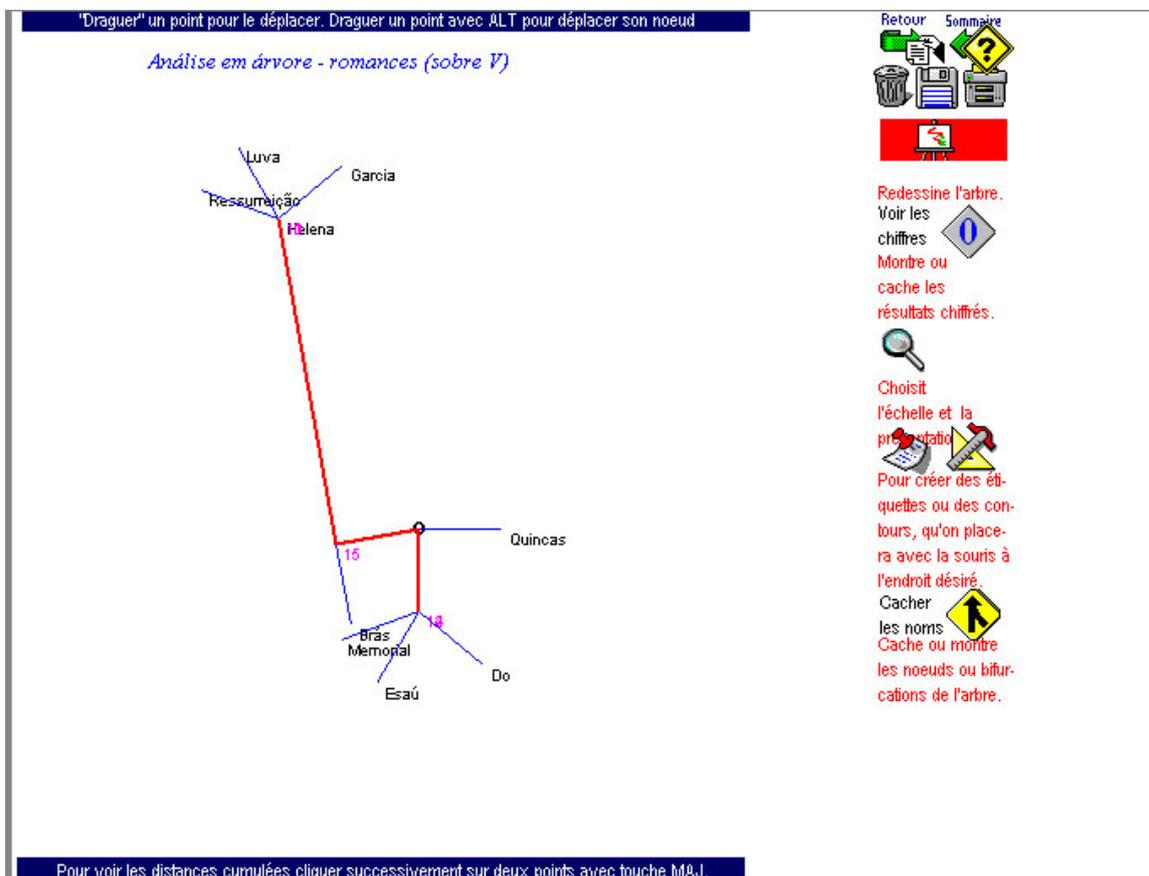


Gráfico 46 – Análise em árvore da distância lexical em todos os romances

Podemos ver como é relativa a distância de acordo com a forma que agrupamos as bases. Novamente a cronologia tem grande importância, separando os romances em dois grupos macro, seguindo a máxima da crítica que separa a obra de Machado em duas fases. Na parte de cima, temos as obras até 1878 e, na de baixo, as de 1880 em diante. Porém, há dois desmembramentos na fase madura, em que *Brás Cubas* e *Quincas Borba* aparecem em locais independentes. Ainda é importante ver que os dados dessa base, que agrupa todos os romances, confirma as posições de um romance em relação a outro, o que vimos nos gráficos parciais por fase.

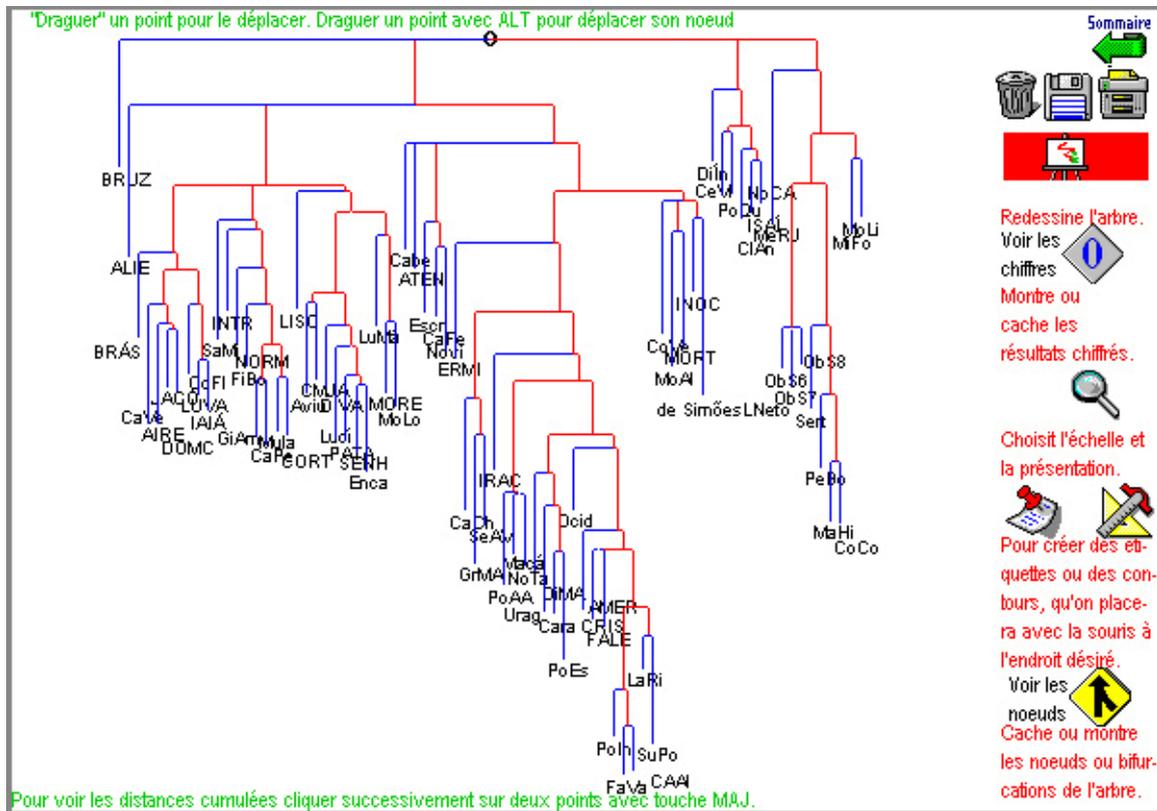


Gráfico 48 – Análise em árvore (retangular) da distância lexical na base Portext

Aqui temos uma melhor visualização dos dados e podemos observar uma concentração de obras de prosa de Machado no primeiro agrupamento na parte de cima, à esquerda do quadro. *O alienista* e *Brás Cubas* se destacam, isolando-se dos galhos de onde pendem, enquanto *Casa velha* e *Memorial de Aires* e *Dom Casmurro* se concentram num mesmo galho. Logo ao lado, os romances da juventude se agrupam em outro galho: *Contos fluminenses*, *Iaiá* e *A mão e a luva*, próximos de outros romances da época, como os de Alencar, logo à direita.

O gênero nesse tipo de grupo parece ser o mais determinante, já que *Ocidentais*, *Americanas*, *Crisálidas* e *Falenas*, estas três presas em um mesmo galho, aparecem na parte de baixo do gráfico, acompanhando outras obras de poesia de outros autores.

7.3 O Conto

Voltando à análise dos dados por gênero, chegamos finalmente aos gráficos de distância que tratam dos volumes de contos publicados por Machado de Assis.

Seguindo a ordem cronológica da contística machadiana, começaremos pela base Flumeia, que contempla os dois primeiros volumes publicados pelo autor.

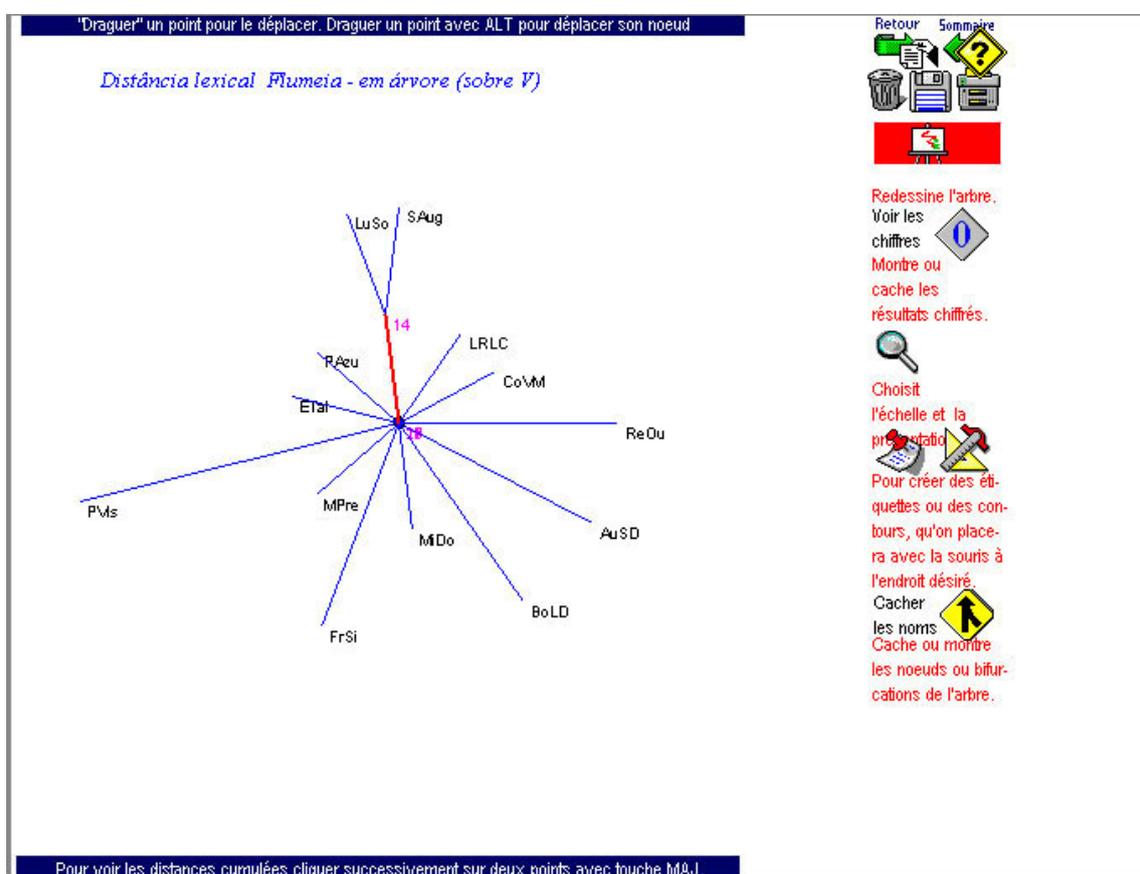


Gráfico 49 – Análise em árvore da distância lexical na base Flumeia

A distância lexical dos dois primeiros volumes é maior entre “Luís Soares” e “O segredo de Augusta” dos demais. Ambos são do volume *Contos fluminenses*. “Ponto de vista” é outro conto cujo vocabulário se afasta dos demais, embora saia do mesmo nó da maioria, um tipo de afastamento diferente dos dois primeiros, que são folhas de um galho específico.

Como nosso objetivo aqui não é o de analisar o motivo específico de afastamento de cada conto, mas o de buscar padrões que se mantenham ao longo da produção do autor, passaremos aos próximos gráficos.

'Draguer' un point pour le déplacer. Draguer un point avec ALT pour déplacer son noeud



Pour voir les distances cumulées cliquer successivement sur deux points avec touche MAJ.

Gráfico 50 – Análise em árvore da distância lexical na base AVG1

AVG1 é a base de contos não publicados contemporânea a Flumeia, logo aquela serve de referência de comparação para que possamos buscar padrões e quebras de padrões que nos apontem continuidade ou ruptura em termos de estilo.

Nota-se nesse gráfico uma semelhança com o anterior (onde as folhas estão mais afastadas porque o número de contos é menor). Aqui temos dois troncos que se afastam do restante. O mais afastado contém os contos “Dona Mônica”, de 1876, e “O país das quimeras”, de 1862. No outro tronco estão “Um cão de lata ao rabo” e “Elogio da Vaidade”, ambos de 1878.

Mantém-se uma forte unidade do vocabulário tanto nos contos não publicados quanto nos publicados, por época.

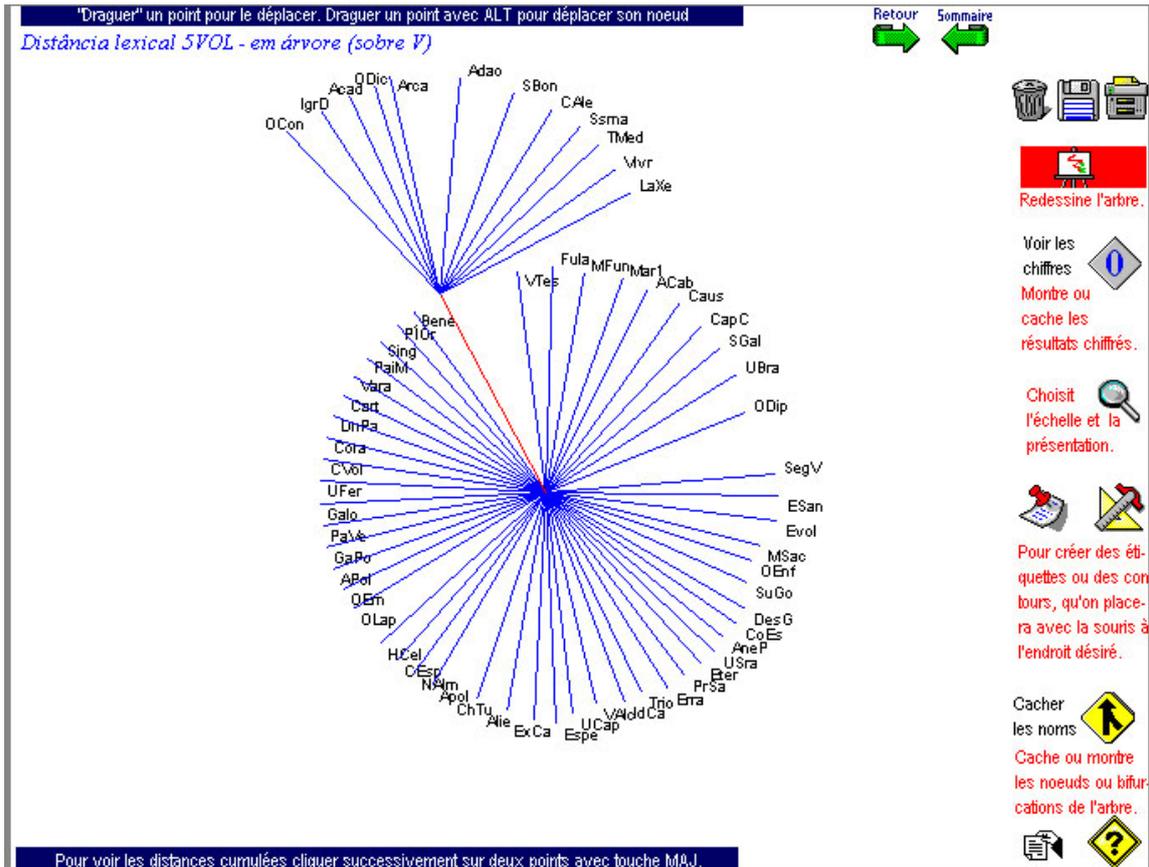


Gráfico 51 – Análise em árvore da distância lexical na base 5VOL

Nos cinco volumes publicados da década de 80 em diante, apenas 12 dos 63 contos que compõem a base se afastam dos demais num tronco específico. Esses 12 são de quatro volumes diferentes: “Na Arca”, “O Segredo do Bonzo”, “A Sereníssima república”, “Teoria do Medalhão, de *Papéis avulsos* (1882); e “A Igreja do Diabo”, “As Academias de São” e “Conto Alexandrino”, de *Histórias sem data* (1884); “O Cônego”, “Adão e Eva” e “Viver”, de *Várias histórias* (1896); e “O Dicionário” e “Lágrimas de Xerxes”, de *Páginas recolhidas* (1899).

Nesses casos é bem clara a influência da temática dos contos (muito específica) na diferenciação de seu vocabulário em relação aos demais.

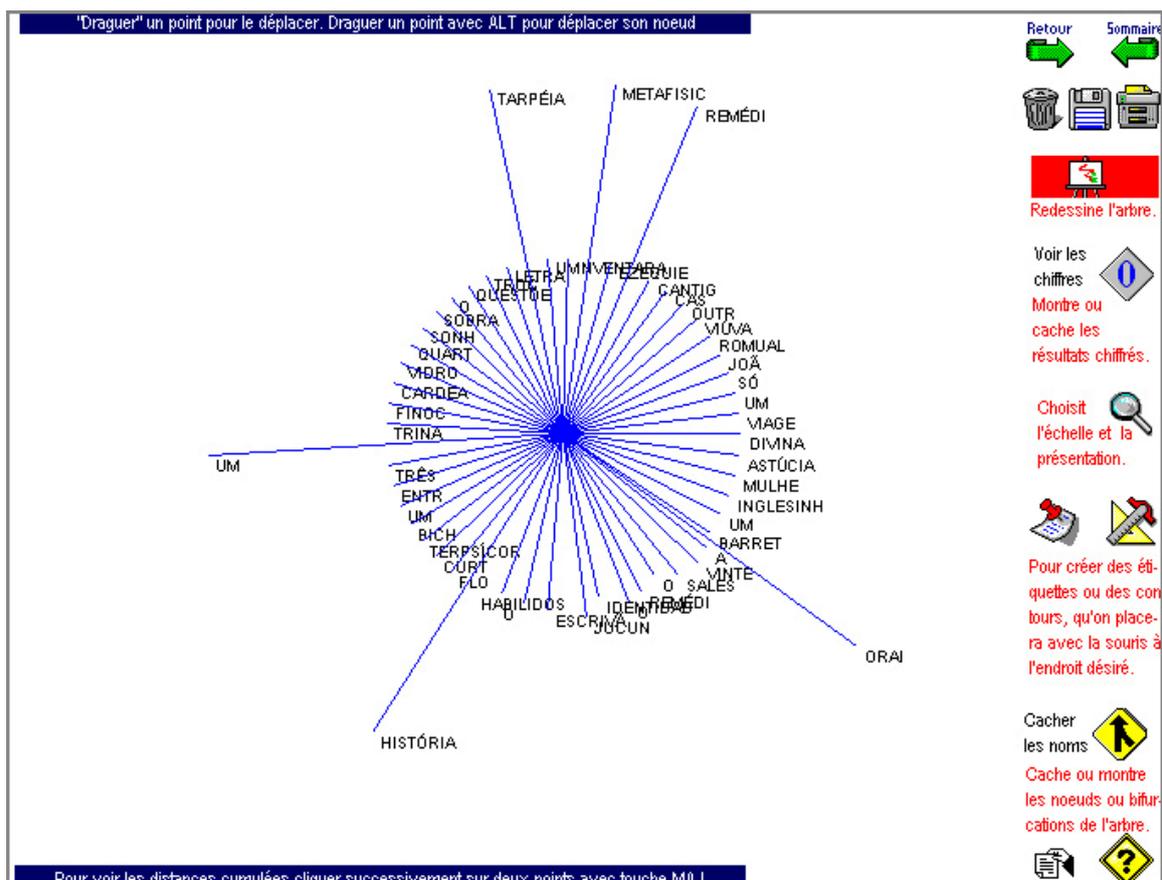


Gráfico 52 – Análise em árvore da distância lexical na base AVG2

AVG2, sendo a base composta dos contos avulsos contemporâneos aos da base anterior (5VOL), mostra-se a mais homogênea de todas até aqui em termos de distância lexical. Do nó principal não sai nenhum tronco com folhas separadas; apenas seis de 54 contos se distanciam dos demais.

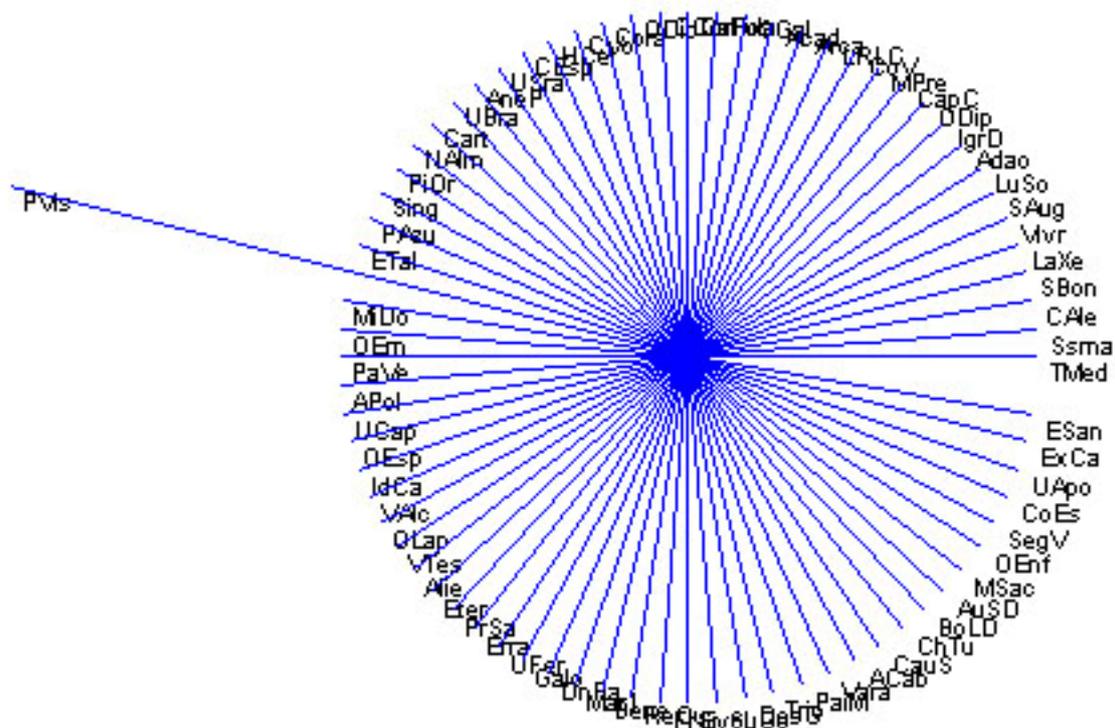


Gráfico 53 – Análise em árvore da distância lexical em todos os volumes

O ponto principal de nossa tese é a análise de padrões de continuidade e/ou ruptura no estilo do contista Machado de Assis. A imagem acima mostra que, ao reunirmos os 76 contos publicados em livro pelo autor em diferentes épocas, prevalece um padrão de vocabulário. De um só nó saem os 76 contos de forma praticamente equidistante, formando a figura de um círculo. Apenas o conto “Ponto de Vista” se afasta dos demais. Mas como vem da mesma origem, a unidade lexical no universo do conto machadiano nos aponta que, sob esse aspecto, não podemos falar em ruptura.

CONSIDERAÇÕES FINAIS

Buscamos, ao longo deste trabalho, investigar o material linguístico utilizado por Machado de Assis em seus contos, a fim de encontrar indícios que confirmassem ou não uma ruptura estilística que cindiria sua obra em duas fases, como afirma parte da crítica.

Como este trabalho é de natureza mais descritiva do que especificamente analítica, passemos a nossas conclusões, centradas nos sete volumes publicados pelo autor.

Quanto ao léxico, verificamos que Machado não apresenta uma riqueza expressiva nem quanto às formas (V), nem quanto às palavras de frequência 1. A constituição de seu léxico é equilibrada cronologicamente, e a distribuição dos crescimentos e decréscimos é gradual e em geral constante.

Quanto às palavras de alta frequência, os contos machadianos apresentam grande regularidade tanto nos substantivos quanto nos verbos, conservando naquela lista cerca de uma dezena de substantivos que vão do primeiro ao último volume publicado. O mesmo acontece com uma lista de aproximadamente 70 verbos.

O crescimento do léxico é gradual, atravessando a década de 1880, para depois equilibrar-se até o fim.

A divisão das categorias gramaticais, feita a partir das altas frequências, aponta um excedente de palavras funcionais que é ascendente ao longo do tempo, enquanto os nomes próprios decrescem significativamente. Os verbos e os substantivos têm uma queda discreta. Essa alta incidência de palavras funcionais aponta para um alongamento das frases, dissociado do aumento da extensão dos textos, que sofrem um decréscimo após *Histórias da meia-noite* e ficam equilibrados até o último volume. O encurtamento dos textos, associado ao alongamento das frases, confirma uma mudança estrutural no conto machadiano que, mediante a análise da pontuação, podemos confirmar, mesmo que parcialmente.

Enfim, o único traço de ruptura encontrado nos dados analisados é relativo ao uso da vírgula, que sofre uma alta tanto no conto quanto no romance, no período que a crítica aponta como marco da transição entre as fases: 1881 para o romance e 1882 para os contos. Nos demais aspectos analisados, constatamos uma regularidade tanto nos crescimentos lexicais quanto nos decréscimos, sempre seguindo a linha do tempo.

No entanto, se tomarmos os dados relativos à distância lexical apresentados no último capítulo podemos verificar uma forte unidade entre os contos, independentemente de época, se foi publicado ou não, ou mesmo para que tipo de publicação tenha sido produzido. O gráfico dos sete volumes publicados em livro não deixa dúvidas de que há um padrão que une a obra do Machado contista.

Quanto ao romance, há, de fato, um afastamento em termos de distância lexical, a partir de *Brás Cubas*, mas essa talvez seja uma questão relativa não apenas ao autor; pode estar relacionada ao gênero. Muitos são os dados que essa nova metodologia nos permite obter, mas também muitos são os riscos que corremos na interpretação, na aventura de testar uma abordagem pioneira. Brunet alerta o pesquisador sobre esses riscos: “*Même lorsqu’une distance paraît établie solidement entre deux textes, on ne sait pas toujours à quoi la rattacher. À l’auteur ? À l’époque ? Au sujet traité ? Au genre littéraire ?*”²¹⁶

Porém, não há avanço em nenhum campo do conhecimento sem que se enfrente o que está estabelecido, sem que se experimentem novas abordagens.

Enfim, voltando aos dados, sua maioria aponta para uma transformação gradual no conto machadiano – e não uma ruptura –, no modo como o autor estrutura seu texto, que acompanha seu amadurecimento e o exercício incansável da escrita. Quanto ao lugar da originalidade do estilo machadiano, nossos dados apontam para a distribuição, a ordenação, enfim, para a composição.

Um novo método, não só para a literatura

Ler e estudar uma obra como a de Machado é um prazer e um privilégio. Falar do estilo de um autor tão estudado é, no entanto, uma aventura. Fazê-lo utilizando ferramentas que não são usuais nos estudos literários brasileiros torna o desafio maior ainda. A opção por um caminho interdisciplinar, que envolveu vários estudos paralelos, além da investigação literária que já é complexa o suficiente, enriqueceu nosso percurso.

Partimos para esta pesquisa com um único livro sobre o assunto disponível na biblioteca da UFSC. A data de publicação: 1968. Sem bibliografia adicional, nem nas bibliotecas, nem nas livrarias, nem em português. As assinaturas de nossa biblioteca das

²¹⁶ BRUNET, Étienne. Peut-on mesurer la distance entre deux textes? *Corpus*, n. 2, déc. 2003.

revistas eletrônicas internacionais que tratam da estatística textual expiraram em 2004. O preço do pioneirismo é alto, mas a aventura da descoberta, como dizem Leech e Short, compensa.

Este trabalho significa, juntamente com outros desenvolvidos em nosso núcleo de pesquisa, a adoção de mais um método para a investigação da literatura, que tem por foco a anatomia do texto. No caso dos contos de Machado, apesar da escassez de trabalhos que mencionamos no início desta pesquisa, “a fortuna crítica de sua contística”, como afirma Luís Augusto Fischer, “encontrou aí o impensável. Repertórios de temas; repertórios de constantes filosóficas, repertórios de alusões”. Já “quanto às constantes estruturais dos contos, porém, a tradição é sensivelmente mais pobre do que quanto às constantes temáticas”²¹⁷.

Como nosso interesse em relação a Machado sempre esteve mais direcionado a essas constantes estruturais, o instrumental que a estatística textual e os programas de análise de texto oferecem ao pesquisador é válido. O instrumental técnico e metodológico, imperativamente multidisciplinar e interdisciplinar, representa um incremento da leitura, em sentido amplo, do texto literário, bem como de textos de qualquer natureza, pois as informações obtidas por meio dele podem posteriormente ser submetidas à interpretação, ao tipo de teoria que o pesquisador considere mais adequada para seus objetivos. Como bem sintetiza Muller,

Les repertoires linguistiques, dont quelques-uns viennent d'être cités et les logiciels qui leur sont adaptés permettent d'arracher aux textes des secrets qu'aucune lecture ne pouvait déceler, et d'ajouter aux connaissances traditionnelles du linguiste et du littéraire des données nouvelles.
Ceux qui étudient ou enseignent notre langue et sa littérature savent-ils?²¹⁸

Essa metodologia, como mencionamos no início deste trabalho, em países como França ou Estados Unidos já constitui um ramo dos estudos literários e tem uma tradição que se estende há quase meio século, contando-se apenas os estudos que usam o computador. Na Europa, é usada para estudos em várias áreas, como Sociologia, Psicologia, Direito, Jornalismo e Publicidade.

²¹⁷ FISCHER, op. cit., p. 149.

²¹⁸ MULLER, Charles. Des participes, de leurs accords..., p. 288.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARBETTA, Pedro Alberto. *Estatística aplicada às Ciências Sociais*. 5. ed. rev. Florianópolis: Ed. da UFSC, 2005.
- BÉHAR, Henri. Hubert Phalèse's, Method. *Literary and Linguistic Computing*, Oxford: Oxford University Press, v. 10, n. 2, p. 129-134, 1995.
- BERNARD, Michel. *Introduction aux études littéraires assistées par ordinateur*. Paris: Presses Universitaires de France, 1999.
- BLOOM, Harold. *O cânone ocidental*. Tradução de Marcos Santarrita. Rio de Janeiro: Objetiva, 1995.
- BOSI, Alfredo et al. *Machado de Assis*. São Paulo: Ática, 1982.
- BRAYNER, Sonia. *Labirinto do espaço romanesco: tradição e renovação da literatura brasileira (1880-1920)*. Rio de Janeiro: Civilização Brasileira; Brasília: INL, 1979.
- BRUNET, Etienne. La punctuation et le rythme du discours (d'après les données du Trésor de la langue française). *CUMFID*, Université de Nice, Nice, n. 13, juil. 1981.
- BRUNET, Etienne. *Le vocabulaire de Proust : étude quantitative*. Genève: Slatkine, 1983.
- BRUNET, Etienne. *Hyperbase: manuel de référence (Version 5.4 par Windows)*, janv. 2002.
- BRUNET, Etienne. Peut-on mesurer la distance entre deux textes? *Corpus*, n. 2, déc. 2003. Disponível em: <<http://corpus.revues.org/index30.html>>. Acesso em : 27 jun. 2007.
- CALVINO, Ítalo. *Por que ler os clássicos*. Tradução de Nilson Moulin. São Paulo: Companhia das Letras, 1994.
- CANDIDO, Antonio. *Vários escritos*. São Paulo: Duas Cidades, 1970.
- CANDIDO, Antonio. *Formação da literatura brasileira*. 7. ed. Rio de Janeiro: Itatiaia, 1993.
- CANDIDO, Antonio. *Iniciação à literatura brasileira*. São Paulo: Humanitas FFLCH/USP, 1999.
- CARPEAUX, Otto Maria. *Ensaio reunidos (1946-1971)*. Rio de Janeiro: UniverCidade; Topbooks, 2005.
- MELLET, Sylvie et alii. *Mots chiffrés et déchiffrés : mélanges offerts à Etienne Brunet*. Paris: Honoré Champion, 1998.
- CORNS, Thomas N. Literary Theory and computer-based criticism: current problems and future prospects. In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET

INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. *Actes...*, Nice, 1986. p. 223-227.

COUTINHO, Afrânio. *Machado de Assis na literatura brasileira*. 2. ed. Rio de Janeiro: São José, 1966.

CUNHA DE SOUZA, Cilene. *Um método quantitativo para a análise lexical*. Rio de Janeiro: Tempo Brasileiro; Brasília: MEC, 1979.

CUNHA, Patrícia Lessa Flores da. *Machado de Assis: um escritor na capital dos trópicos*. Porto Alegre: IEL; São Leopoldo: Unisinos, 1998.

DIXON, Paul B. *Os contos de Machado de Assis: mais do que sonha a filosofia*. Porto Alegre: Movimento, 1992. 116 p. (Coleção Machadiana, 6).

DUNCAN JR., John C. *A frequency dictionary of Portuguese words*. Michigan-London: University Microfilms International, 1971.

EMORINE, Jacques. *Lexique et analyse lexicale de l'Auto da Compadecida*. Étude statistique, Thèse de Doctorat, Université de Toulouse, Toulouse, 1968.

FAORO, Raimundo. *Machado de Assis: a pirâmide e o trapézio*. 4. ed. ver. São Paulo: Globo, 2001.

FISCHER, Luís Augusto. Contos de Machado: da ética à estética. In: SECCHIN, Antonio Carlos et al. *Machado de Assis: uma revisão*. Rio de Janeiro: In-Fólio, 1998. p. 147-165.

FOLHA ONLINE. *Bird reafirma que Brasil tem maior desigualdade da América Latina*. Dinheiro. 24 out. 2003. Disponível em: <<http://www1.folha.uol.com.br/folha/dinheiro/ult91u75627.shtml>>. Acesso em: 18 abr. 2006.

FREITAS, Deise J. T. *A revolução do estilo de Machado de Assis: uma análise de quatro contos*. 1998. Dissertação (Mestrado em Teoria Literária) – Curso de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis, 1998.

GLEDSON, John. A História do Brasil em Papéis Avulsos de Machado de Assis. In: CHALOUB, Sidney; PEREIRA, Leonardo Affonso de M. (Org.). *A história contada: capítulos de história social da literatura no Brasil*. Rio de Janeiro: Nova Fronteira, 1998. p. 15-34.

GLEDSON, John. *Por um novo Machado de Assis: ensaios*. São Paulo: Companhia das Letras, 2006.

GUIRAUD, Pierre. *A estilística*. Tradução de Miguel Maillet. São Paulo: Mestre Jou, 1970.

GUIRAUD, Pierre. *Essais de stylistique*. Paris: Klincksieck, 1985.

HOCKEY, Susan. *Electronic texts in the humanities*. London; New York: Oxford University Press, 2004.

INGARDEN, Roman. *A obra de arte literária*. 2. ed. Tradução de Albin E. Beau Maria da Conceição Puga e João F. Barrento. Lisboa: Fundação Calouste Gulbenkian, 1965.

- KENNY, Anthony. *The computation of style*. Oxford: Pergamon Press, 1982.
- LABBÉ, Cyrill; LABBÉ, Dominique. La distance intertextuelle. *Corpus* [online], n. 2, déc. 2003. Disponível em: <<http://corpus.revues.org/index31.html>>. Acesso em: 27 jun. 2007.
- LEBART, Ludovic; SALEM, Andre. *Statistique textuelle*. Paris: Dunod, 1994.
- LEECH, Geoffrey N.; SHORT, Michael H. *Style in fiction*. New York: Longman, 1990.
- LUKÁCS, Georg. *Ensaaios de literatura*. Rio de Janeiro: Civilização Brasileira, 1968.
- LUONG, Xuan; NOVI, Michel. Représentations arborées de données textuelles. In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. *Actes...*, Nice, 1986. p. 577-586.
- MACHADO DE ASSIS, Joaquim Maria. *Obras completas*. Rio de Janeiro: Nova Aguilar, 1997.
- MACIEL, Carlos Alberto Antunes. Da base Portext ao CD-ROM de textos jurídicos. In: SARDINHA, Tony Berber (Org.). *A língua portuguesa no computador*. Campinas, SP: Mercado das Letras; São Paulo: Fapesp, 2005, p. 185-195.
- MACIEL, Carlos Alberto Antunes. *Richesse et evolution du vocabulaire d'Érico Veríssimo (1905-1975 – Porto Alegre, Brésil)*. Paris: Champion; Genève: Slaktine, 1986.
- MASSA, Jean Michel. A juventude de Machado de Assis. In: BOSI, Alfredo et al. *Machado de Assis*. São Paulo: Ática, 1982.
- MOISÉS, Massaud. *Machado de Assis: ficção e utopia*. São Paulo: Cultrix, 2001.
- MULLER, Charles. *Initiation à la statistique linguistique*. Paris: Larousse, 1968.
- MULLER, Charles. Des participes, de leurs accords, et des inquiétudes qu'ils occasionnent. In: MELLET, Sylvie et alii. *Mots chiffrées et déchiffrées : mélanges offerts à Etienne Brunet*. Paris: Honoré Champion, 1998.
- PEREIRA, Lúcia Miguel. *Machado de Assis: estudo crítico e biográfico*. 6. ed. Belo Horizonte: Itatiaia; São Paulo: Edusp, 1988.
- PEREIRA, Lúcia Miguel. *História da literatura brasileira: prosa de ficção: de 1870 a 1920*. Belo Horizonte: Itatiaia; São Paulo: Edusp, 1988.
- PHALÈSE, Hubert de. *Comptes a rebours : l'oeuvre de Huysmans à travers les nouvelles technologies*. Paris: Nizet, 1991. (Collection Cap'agreg).
- RASTIER, François. *Arts et sciences du texte*. Paris: PUF, 2001.
- ROCKWELL, Geoffrey. What is text analysis, really? *Literary and Linguistic Computing*, v. 18, n. 2, p. 209-219, 2003.
- ROMERO, Sílvio. *Machado de Assis*. 2. ed. Rio de Janeiro: José Olympio, 1936.

SÁ REGO, Enylton José de. *O calundu e a panacéia: Machado de Assis, a sátira menipéia e a tradição luciânica*. Rio de Janeiro: Forense Universitária, 1989.

SANTIAGO, Silviano. *Uma literatura nos trópicos: ensaios sobre dependência cultural*. 2. ed. Rio de Janeiro: Rocco, 2000.

SCHWARZ, Roberto. *Ao vencedor as batatas: forma literária e processo social nos inícios do romance brasileiro*. São Paulo: Duas Cidades; Ed. 34, 2000.

SIEMENS, Raymond G. A new computer-assisted literary criticism? *Computers and the Humanities*, n. 36, p. 259-267, 2002.

VERÍSSIMO, José. *História da Literatura Brasileira*. Rio de Janeiro: Francisco Alves, 1916. Disponível em: <<http://geocities.yahoo.com.br/ciberespao/ebooksgratuitos.htm>>. Acesso em: 20 mar. 2009.

BIBLIOGRAFIA

- ALMEIDA, José Maurício Gomes de. Da humana comédia ou no teatro em Itaguaí. In: SECCHIN, Antônio Carlos; ALMEIDA, José Maurício Gomes de; SOUZA, Ronaldo de Melo e (Org.). **Machado de Assis: uma revisão**. Rio de Janeiro: In-Fólio, 1998. p. 167-177.
- ANTI, Peter Takyi. **The early short stories of Machado de Assis: 1858-1878**. Dissertation Abstracts International, Ann Arbor, 1991. Degree: U. of Liverpool.
- ARISTÓTELES. **Poética**. Tradução de Eudoro de Souza. São Paulo: Nova Cultural, 1987. (Coleção Os Pensadores).
- BACHELARD, Gaston. A “novidade” das ciências contemporâneas. In: LECOURT, Dominique. **Gaston Bachelard: epistemologia**. Tradução de Nathanael C. Caixeiro. Rio de Janeiro: Zahar, 1977. p. 13-17.
- BACHELARD, Gaston. As categorias principais da epistemologia. In: LECOURT, Dominique. **Gaston Bachelard: epistemologia**. Tradução de Nathanael C. Caixeiro. Rio de Janeiro: Zahar, 1977. p. 101-146.
- BAKHTIN, Mikhail. **Marxismo e filosofia da linguagem**. 6. ed. Tradução de Michel Lahud e Yara Frateschi Vieira. São Paulo: Hucitec, 1992.
- BAKHTIN, Mikhail. **Questões de literatura e estética: a teoria do romance**. 3. ed. Tradução de Aurora Fornoni Bernardini et al. São Paulo: Hucitec/Unesp, 1993.
- BAPTISTA, Abel Barros. Singular experiência: para uma revisão dos contos de Machado de Assis. **Scripta**, Belo Horizonte, v. 3, n. 6, p. 13-19, jan./jun. 2000.
- BARBETTA, Pedro Alberto. **Estatística aplicada às Ciências Sociais**. 5. ed. rev. Florianópolis: Ed. da UFSC, 2005.
- BARBIERI, Ivo. “O alienista”: a razão que enlouquece. **Revista Brasileira**, Rio de Janeiro, v. 4, n. 14, jan./mar. 1998.
- BARBIERI, Ivo. O cônego ou invenção da linguagem. **Revista Tempo Brasileiro**, Rio de Janeiro, n. 133-134, p. 23-34, 1998.
- BARR, George K. Graphical analysis of the sentence length distribution curve and non-rational components. **Literary and Linguistic Computing**, v. 16, n. 4, p. 375-388, 2001.
- BAUDELAIRE, Charles. **A modernidade de Baudelaire**. Seleção de Teixeira Coelho. Tradução de Suely Cassal. Rio de Janeiro: Paz e Terra, 1988.
- BÉHAR, Henri. Hubert Phalèse’s: method. **Literary and Linguistic Computing**, Oxford: Oxford University Press, v. 10, n. 2, p. 129-134, 1995.
- BELLEI, Sérgio Luiz Prado. Machado de Assis: uma poética da nacionalidade. **Travessia - Machado de Assis, 150 anos**, Curso de Pós-Graduação em Literatura Brasileira, UFSC, Florianópolis, n. 19, p. 24-38, 2. sem. 1989.

BENJAMIN, Walter. **Magia e técnica, arte e política**. Tradução de Sérgio Paulo Rouanet. Obras Escolhidas. São Paulo: Brasiliense, 1985.

BENJAMIN, Walter. **Charles Baudelaire: um lírico no auge do capitalismo**. 2. ed. Tradução de José Carlos Martins Barbosa e Hemerson Alves Baptista. São Paulo: Brasiliense, 1991. (Obras escolhidas, v. III).

BERMAN, Marshall. **Tudo o que é sólido desmancha no ar: a aventura da modernidade**. Tradução de Carlos Felipe Moisés e Ana Maria L. Ioriatti. São Paulo: Cia. das Letras, 1986.

BERNARD, Michel. **Introduction aux études littéraires assistées par ordinateur**. Paris: Presses Universitaires de France, 1999.

BLOOM, Harold. **O cânone ocidental**. Tradução de Marcos Santarrita. Rio de Janeiro: Objetiva, 1995.

BORBA, Osório. **A comédia literária**. Rio de Janeiro: Civilização Brasileira, 1959. (Vera Cruz, 12).

BORIM, Dario. O Machado e a batina: uma investigação sobre a Alma exterior de um católico irreverente. **Quadrant**, Montpellier, FR, n. 7, p. 95-111, 1990.

BOSI, Alfredo et al. **Machado de Assis**. São Paulo: Ática, 1982.

BOSI, Alfredo **História Concisa da Literatura Brasileira**. 3. ed. São Paulo: Cultrix, 1993.

BOSI, Alfredo **Machado de Assis, o enigma do olhar**. São Paulo: Ática, 2003.

BOSI, Alfredo. **Eurípides Simões de Paula, in memoriam**. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas da USP, 1983.

BRADBURY, Malcolm. **O mundo moderno: dez grandes escritores**. Tradução de Paulo Henriques Britto. São Paulo: Cia. das Letras, 1989.

BRAGA, Iara Solange. **Contos fluminenses e Histórias da meia-noite na formação de Machado de Assis**. Dissertação (Mestrado em Letras) – Universidade Federal de Santa Catarina, Florianópolis, 2000.

BRAYNER, Sônia. Edgar Allan Poe e Machado de Assis. **Minas Gerais, Suplemento Literário**, Belo Horizonte, n. 19, p. 1-2, jun. 1976.

BRAYNER, Sonia. **Labirinto do espaço romanescos: tradição e renovação da literatura brasileira (1880-1920)**. Rio de Janeiro: Civilização Brasileira; Brasília: INL, 1979.

BRAYNER, Sônia. Um passeio no Rio antigo: os contos de Machado de Assis. **Travessia**, Florianópolis, n. 16-18, p. 279-290, 1988-1989.

BREUILLARD, Jean. Pour une macrostylistique de la phrase. In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. **Actes...**, Nice, 1986. p. 135-141.

BRUNET, Etienne. La punctuation et le rythme du discours (d'après les données du Trésor de la langue française). **CUMFID**, Université de Nice, n. 13, p. 1-27, juil. 1981.

BRUNET, Etienne. **Le vocabulaire de Proust** : étude quantitative. Genève: Slatkine, 1983.

BRUNET, Etienne. **Hyperbase**: manuel de référence (Version 5.4 par Windows). CNRS, Nice, janv. 2002.

BRUNET, Etienne. Peut-on mesurer la distance entre deux textes? **Corpus**, n. 2, déc. 2003. Disponível em: <<http://corpus.revues.org/index30.html>>. Acesso em : 27 jun. 2007.

CALDWELL, Helen. **Machado de Assis**: the Brazilian master and his novels. Los Angeles/Berkeley: University of California Press, 1970.

CALDWELL, Helen. **O Otelo brasileiro de Machado de Assis**. São Paulo: Ateliê, 2002.

CALVINO, Ítalo. **Por que ler os clássicos**. Tradução de Nilson Moulin. São Paulo: Companhia das Letras, 1994.

CAMLONG, André. Esthétique et éthique dans les contes de Machado de Assis. **Travessia - Machado de Assis, 150 anos**. Revista do curso de Pós-Graduação em Literatura Brasileira, UFSC, Florianópolis, n. 19, p. 83-111, 2. sem. 1989.

CANDIDO, Antonio. **A educação pela noite e outros ensaios**. São Paulo: Ática, 1987.

CANDIDO, Antonio. **Recortes**. São Paulo: Companhia das Letras, 1993.

CANDIDO, Antonio. **Vários escritos**. São Paulo: Duas Cidades, 1970.

CANDIDO, Antonio. **Formação da literatura brasileira**. 7. ed. Rio de Janeiro: Itatiaia, 1993.

CANDIDO, Antonio. **Iniciação à literatura brasileira**. São Paulo: Humanitas; FFLCH/USP, 1999.

CARPEAUX, Otto Maria. **Ensaio reunidos (1946-1971)**. Rio de Janeiro: UniverCidade; Topbooks, 2005.

CARPEAUX, Otto Maria. **História da literatura ocidental**. Rio de Janeiro: Alhambra, 1978. v. VIII.

CASTRO, Luiz Gonzaga Garcia de. **Os temas como tecedura narrativa em alguns contos machadianos**. Bauru: Ed. da FASC, 1985.

CHKLOVSKI, V. A arte como procedimento. In: EIKHENBAUN, B. et al. **Teoria da literatura**: formalistas russos. Porto Alegre: Globo, 1973. p. 39-56.

CORNS, Thomas N. Literary theory and computer-based criticism: current problems and future prospects. In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. **Actes...**, Nice, 1986. p. 223-227.

COSTA, Mário. Pour une nouvelle esthétique. **Études de Leonardo/Olats**, déc. 2000.

COUCHOT, Edmond. **La technologie dans l'art**: de la photographie à la réalité virtuelle. Nîmes: Jacqueline Chambon, 1998.

COUTINHO, Afrânio. **Machado de Assis na literatura brasileira**. 2. ed. Rio de Janeiro: São José, 1966.

COUTINHO, Edilberto. O melhor conto. **Colóquio Letras**, Lisboa, n. 121-122, p. 58-59, jul./dez. 1991.

CRESSOT, Marcel. **O estilo e suas técnicas**. Tradução de Madalena Cruz Ferreira. Lisboa: Setenta, 1980.

CUNHA DE SOUZA, Cilene. **Um método quantitativo para a análise lexical**. Rio de Janeiro: Tempo Brasileiro; Brasília: MEC, 1979.

CUNHA, Celso; CINTRA, Luís F. Lindley. **Nova gramática do português contemporâneo**. Rio de Janeiro: Nova Fronteira, 2001.

CUNHA, Patrícia Lessa Flores da. **Machado de Assis**: um escritor na capital dos trópicos. Porto Alegre: IEL; São Leopoldo: Unisinos, 1998.

CURY, Maria Zilda Ferreira. Teoria do medalhão: uma pedagogia do poder. **Espelho – Revista Machadiana**, Porto Alegre: West Lafayette, n. 1, p. 39-49, 1995.

DANIEL, Mary L. Two failed fortune tellers: Machado de Assis's Cartomante and João Guimarães Rosa's Mme de Sajs. **Luso-Brazilian Review**, Madison, WI, v. 23, n. 2, p. 47-59, 1986.

DIMAS, Antônio. O espelho irônico de Machado: entrevista com John Gledson. **Cult – Revista Brasileira de Literatura**, São Paulo, n. 24, p. 42-50, jul. 1999.

DIXON, Paul B. **Os contos de Machado de Assis**: mais do que sonha a filosofia. Porto Alegre: Movimento, 1992. 116 p. (Coleção Machadiana, 6).

DUNCAN JR., John C. **A frequency dictionary of Portuguese words**. Michigan-London: University Microfilms International, 1971.

EAGLETON, Terry. **Teoria da literatura**: uma introdução. Tradução de Waltensir Dutra. São Paulo: Martins Fontes, 1994.

EIKHENBAUM, B. Como é feito o Capote de Gogol. In: EIKHENBAUM, B. et al. **Teoria da Literatura**: formalistas russos. Porto Alegre, Globo, 1971.

EMORINE, Jacques. **Lexique et analyse lexicale de l'Auto da Compadecida**. Étude statistique, thèse de doctorat, Université de Toulouse, Toulouse, 1968.

FAORO, Raymundo. **A pirâmide e o trapézio**. São Paulo: Companhia Editora Nacional, 1974.

FAORO, Raimundo. **Machado de Assis**: a pirâmide e o trapézio. 4. ed. ver. São Paulo: Globo, 2001.

FAORO, Raymundo. O espelho e a lâmpada: uma introdução a Machado de Assis. **Revista Brasileira de Cultura**, Rio de Janeiro, v. 2, n. 5, p. 153-172, 1970.

FERREIRA, Eliane Fernanda Cunha. **Para traduzir o século XIX**: Machado de Assis. São Paulo: Annablume; Rio de Janeiro: ABL, 2004.

FERREIRA, João Martins. Contribuições da estatística, matemática e informática em análises lingüísticas e semióticas. In: SARDINHA, Tony Berber (Org.). **A língua portuguesa no computador**. Campinas (SP): Mercado das letras; São Paulo: Fapesp, 2005. p. 249-267.

FERREIRA, Sandra Aparecida. Quem conta um conto, entrelaça os pontos: elementos de coesão textual em um conto de Machado de Assis. **Alfa**, São Paulo, n. 37, p. 127-134, 1993.

FERRY, Luc. **Homo aestheticus**: a invenção do gosto na era democrática. Tradução de Eliana Maria de Melo e Souza. São Paulo: Ensaio, 1994.

FISCHER, Luís Augusto. Contos de Machado: da ética à estética. In: SECCHIN, Antonio Carlos et al. **Machado de Assis**: uma revisão. Rio de Janeiro: In-Fólio, 1998. p. 147-165.

FOLHA ONLINE. **Bird reafirma que Brasil tem maior desigualdade da América Latina**. Dinheiro. 24 out. 2003. Disponível em: <<http://www1.folha.uol.com.br/folha/dinheiro/ult91u75627.shtml>>. Acesso em: 18 abr. 2006.

FORNAZARO, Antônio E. O avesso e o poder em *O alienista*. **Minas Gerais, Suplemento Literário**, Belo Horizonte, p. 3, 31 jul. 1976.

FREITAS, Deise J. T. **A revolução do estilo de Machado de Assis**: uma análise de quatro contos. 1998. Dissertação (Mestrado em Teoria Literária) – Curso de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis, 1998.

GALANTE DE SOUSA, J. **Bibliografia de Machado de Assis**. Rio de Janeiro: Instituto Nacional do Livro, 1955.

GENETTE, Gérard. **Discurso da narrativa**. Tradução de Fernando Cabral Martins. Lisboa: Vega, 1976.

GENETTE, Gérard. **Introdução ao arquitexto**. Tradução de Cabral Martins. Lisboa: Vega, 1986.

GENETTE, Gérard. **Palimpsestes**. Paris: Éditions du Seuil, 1982.

GHELLER, Erinida Gema. **Unidade e autonomia dos Papéis avulsos**. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1978.

GLEDSON, John. A História do Brasil em *Papéis avulsos* de Machado de Assis. In: CHALHOUB, Sidney; PEREIRA, Leonardo Affonso de Miranda (Org.). **A história contada**: capítulos de história social da literatura no Brasil. Rio de Janeiro: Nova Fronteira, 1998. (Histórias do Brasil).

GLEDSON, John. O machete e o violoncelo. In: GLEDSON, John. **Por um novo Machado de Assis**: ensaios. São Paulo: Companhia das Letras, 2006. p. 35-69.

- GLEDSON, John. Uma lição de história: *Conto de escola*, de Machado de Assis. In: JOBIM, José Carlos (Org.). **A biblioteca de Machado de Assis**. Rio de Janeiro: Academia Brasileira de Letras; Topbooks, 2001. p. 359-372.
- GLEDSON, John. A História do Brasil em Papéis Avulsos de Machado de Assis. In: CHALOUB, Sidney; PEREIRA, Leonardo Affonso de M. (Org.). **A história contada**: capítulos de história social da literatura no Brasil. Rio de Janeiro: Nova Fronteira, 1998. p. 15-34.
- GLEDSON, John. **Machado de Assis**: ficção e história. Tradução de Sonia Coutinho. Rio de Janeiro: Paz e Terra, 1986.
- GLEDSON, John. **Machado de Assis**: impostura e realismo: uma reinterpretação de Dom Casmurro. Tradução de Fernando Py. São Paulo: Companhia das Letras, 1991.
- GLEDSON, John. **Por um novo Machado de Assis**: ensaios. São Paulo: Companhia das Letras, 2006.
- GOMES, Eugênio. A arte do conto em Machado de Assis. **Cadernos Brasileiros**, Rio de Janeiro, v. 6, n. 3, p. 27-36, 1964.
- GOMES, Eugênio. **Espelho contra espelho**. São Paulo: Ipê, 1949.
- GOMES, Eugênio. **Machado de Assis**: contos. Rio de Janeiro: Agir, 1963. v. 70. (Nossos Clássicos).
- GOMES, Roberto. O Alienista: loucura, poder e ciência. **Tempo Social**, São Paulo, v. 5, n. 1-2, p. 145-160, 1994.
- GRANGER, Gilles-Gaston. **Filosofia do estilo**. Tradução de Scarlett Zerbetto Marton. São Paulo: Perspectiva; Edusp, 1974.
- GUIRAUD, Pierre. **A estilística**. Tradução de Miguel Maillet. São Paulo: Mestre Jou, 1970.
- GUIRAUD, Pierre. **Essais de stylistique**. Paris: Klincksieck, 1985.
- HOCKEY, Susan. **Electronic texts in the humanities**. London; New York: Oxford University Press, 2004.
- HOOVER, David L. **Language and Style in *The Inheritors***. Maryland: University Press of America, 1999.
- HOOVER, David L. Statistical stylistics and authorship attribution: an empirical investigation. **Literary and Linguistic Computing**, v. 16, n. 4, p. 421-444, 2001.
- HOOVER, David L. Frequent word sequences and statistical stylistics. **Literary and Linguistic Computing**, v. 17, n. 2, p. 157-180, 2002.
- HOOVER, David L. Frequent collocations and authorial style. **Literary and Linguistic Computing**, v. 18, n. 3, p. 261-286, 2003.

HOOVER, David L. Multivariate analysis and the study of style variation. **Literary and Linguistic Computing**, v. 18, n. 4, p. 341-360, 2003.

HOOVER, David L. Perspective on vocabulary richness. **Computers and the Humanities**, v. 37, n. 2, p. 151-178, 2003.

HUMBLÉ, Philippe. Hemingway's "very short story": a stylistics frequency analysis. In: PAYNE, Jonathan. **Linguistic approaches to literature**: papers in literary stylistics. Birmingham (Eng.), University of Birmingham, 1995. p. 17-24. (English Language Research).

INGARDEN, Roman. **A obra de arte literária**. 2. ed. Tradução de Albin E. Beau Maria da Conceição Puga e João F. Barrento. Lisboa: Fundação Calouste Gulbenkian, 1965.

JAUBERT, Anna. Des genres comme précurseurs de style. **Loxias**, Nice, n. 8. Disponível em: <<http://revel.unice.fr/loxias/document.html?id=99>>. Acesso em: 8 abr. 2005.

JAUBERT, Anna. Corpus et champs disciplinaires : le rôle de point de vue. **Corpus**, p. 72-87, 2002.

JOBIM, José Luís (Org.). **A biblioteca de Machado de Assis**. Rio de Janeiro: Topbooks, 2001.

JOHNSON, Samuel. **Prefácio a Shakespeare**. Tradução de Enid Abreu Dobrámszky. São Paulo: Iluminuras, 1996.

KANT, Immanuel. **Crítica da faculdade do juízo**. Tradução de Valério Rohden e Antônio Marques. Rio de Janeiro: Forense Universitária, 1993.

KANT, Immanuel. **Crítica da razão pura**. 4. ed. Tradução de Valério Rohden e Udo Balduur Moosburger. São Paulo: Nova Cultural, 1991. (Coleção Os Pensadores).

KENNY, Anthony. **The computation of style**. Oxford: Pergamon Press, 1982.

LABBÉ, Cyrill; LABBÉ, Dominique. La distance intertextuelle. **Corpus** [online], n. 2, déc. 2003. Disponível em: <<http://corpus.revues.org/index31.html>>. Acesso em: 27 jun. 2007.

LEAL, Anabella Azevedo. Dois contos de Machado de Assis à luz das idéias de Mikhail Bakhtin. **Espelho – Revista Machadiana**, Porto Alegre: West Lafayette, n. 1, p. 7-17, 1995.

LEBART, Ludovic; SALEM, Andre. **Statistique textuelle**. Paris: Dunod, 1994.

LEECH, Geoffrey N.; SHORT, Michael H. **Style in fiction**. New York: Longman, 1990.

LEJEUNE, Cristophe. Représentations des réseaux de mots associés. **Jadt**, p.726-736, 2004.

LENOBLE, Michel. Statistique lexicale et critique littéraire : le mariage impossible? In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. **Actes...**, Nice, 1986. p. 567- 573.

LIMA, Luiz Costa (Org.). **Teoria da literatura em suas fontes**. Rio de Janeiro: Civilização Brasileira, 2002. v. I.

LUKÁCS, Georg. **Sociología de la literatura**. Tradução de Michael Faber-Kaiser. Barcelona: Península, 1971.

LUKÁCS, Georg. **Ensayos sobre el realismo**. Tradução de Juan Jose Sebrelli. Buenos Aires: Siglo Veinte, 1965.

LUKÁCS, Georg. **A teoria do romance**. Tradução de José Marcos M. de Macedo. São Paulo: Duas Cidades; Ed. 34, 2000.

LUKÁCS, Georg. **Ensaio de literatura**. Rio de Janeiro: Civilização Brasileira, 1968.

LUONG, Xuan. Le Consensus en Analyse Arborée. In: MELLET, Sylvie et alii. **Mots chiffrés et déchiffrés: mélanges offerts à Etienne Brunet**. Paris: Honoré Champion, 1998, p. 187-197.

LUONG, Xuan; MACIEL, Carlos. Fréquences e repartition des mots dans um corpus de littérature brésilienne. In: JADT 2002, 6ES. JOURNÉES INTERNATIONALES D'ANALYSE STATISTIQUE DES DONNÉES TEXTUELLES, 2002.

LUONG, Xuan; MELLET, Sylvie. Mesures de distance grammaticale entre les textes. **Corpus**, n. 2, déc. 2003. (Versão eletrônica).

LUONG, Xuan; NOVI, Michel. Représentations arborées de données textuelles. In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. **Actes...**, Nice, 1986. p. 577-586.

MACHADO DE ASSIS, Joaquim Maria. **Contos completos de Machado de Assis**. Disponível em: <<http://www2.uol.com.br/machadodeassis/>>. Acesso em: 15 jun. 2006.

MACHADO DE ASSIS, Joaquim Maria. **Obras completas**. Rio de Janeiro: Nova Aguilar, 1997. v. I, II, III. Versão eletrônica disponível em: <<http://www.cce.ufsc.br/~nupill>>. Acesso em: 15 jun. 2006.

MACIEL, Carlos Alberto Antunes. Da base Portext ao CD-ROM de textos jurídicos. In: SARDINHA, Tony Berber (Org.). **A língua portuguesa no computador**. Campinas, SP: Mercado das Letras; São Paulo: Fapesp, 2005. p. 185-195.

MACIEL, Carlos Alberto Antunes. Le vocabulaire des constitutions brésiliennes : quelques notes sur une langue de spécialité. In: MELLET, Sylvie et alii. **Mots chiffrés et déchiffrés: mélanges offerts à Etienne Brunet**. Paris: Honoré Champion, 1998. p. 199-214.

MACIEL, Carlos Alberto Antunes. **Richesse et evolution du vocabulaire d'Érico Veríssimo (1905-1975 – Porto Alegre, Brésil)**. Paris: Champion; Genève: Slaktine, 1986.

MALRIEU, Denise. Stylistique et Statistique textuelle: à partir de l'article de Charles Muller sur les pronoms de dialogue. **Texto**. Disponível em: <www.revue-texto.net/Inedits/Malrieu_Stylistique.pdf>. Acesso em: 15 jun. 2006.

MARETTI, Maria Lúcia L. Isto acaba: uma leitura do conto *D. Benedita: um retrato*, de Machado de Assis. **Remate de Males**, São Paulo, n. 14, p. 111-128, 1994.

MASSA, Jean Michel. A juventude de Machado de Assis. In: BOSI, Alfredo et al. **Machado de Assis**. São Paulo: Ática, 1982.

MAY, Charles E. **Short stories theories**. Ohio: Ohio University Press, 1976.

MAYAFFRE, Damon. Le corpus *réflexifs*: entre architextualité et hypertextualité. **Corpus**, p. 51-69, nov. 2002.

McKENNA, C. W. F.; ANTONIA, A. The statistical analysis of style: reflections on form, meaning, and ideology in the 'Nausicaa' episode of Ulysses. **Literary and Linguistic Computing**, v. 16, n. 4, p. 353-373, 2001.

MELLET, Sylvie et alii. **Mots chiffrés et déchiffrés** : mélanges offerts à Etienne Brunet. Paris: Honoré Champion, 1998.

MERQUIOR, José Guilherme. **Formalismo e tradição moderna**. Rio de Janeiro: Forense Universitária; São Paulo: Ed. da USP, 1974.

MILIC, Louis T. The apriori question in stylistics. In: COLLOQUE INTERNATIONAL DE MÉTHODES QUANTITATIVES ET INFORMATIQUES DANS L'ÉTUDE DES TEXTES. Université de Nice, 5-8 juin 1985, en hommage à Charles Muller / Slatkine – Champion. **Actes...**, Nice, 1986. p. 639-643.

MOISÉS, Massaud. **Dicionário de termos literários**. 7. ed. São Paulo: Cultrix, 1995.

MOISÉS, Massaud. **Machado de Assis: ficção e utopia**. São Paulo: Cultrix, 2001.

MOLINIÉ, Georges; CAHNÉ, Pierre (direction). Qu'est-ce que le style? In: COLLOQUE INTERNATIONAL, Paris. **Actes...** Presses Universitaires de France, Paris, 1994.

MULLER, Charles. **Initiation à la statistique linguistique**. Paris: Larousse, 1968.

MULLER, Charles. Des participes, de leurs accords, et des inquiétudes qu'ils occasionnent. In: MELLET, Sylvie et alii. **Mots chiffrés et déchiffrés** : mélanges offerts à Etienne Brunet. Paris: Honoré Champion, 1998.

ONÓFRIO, Salvatore d'. **Conto brasileiro**: quatro leituras: Machado de Assis, Graciliano Ramos, Guimarães Rosa, Osman Lins. Petrópolis: Vozes, 1979.

PASERO, Carlos Alberto. Machado de Assis cuentista. **Cuadernos hispanoamericanos**, Madrid, n. 598, p. 53-66, abr. 2000.

PEREIRA, Lúcia Miguel. **Machado de Assis: estudo crítico e biográfico**. 6. ed. Belo Horizonte: Itatiaia; São Paulo: Edusp, 1988.

PEREIRA, Lúcia Miguel. **História da literatura brasileira**: prosa de ficção: de 1870 a 1920. Belo Horizonte: Itatiaia; São Paulo: Edusp, 1988.

PERROT, Andréa Czarnobay. Literatura e loucura: a Casa Verde e outras questões n'O alienista de Machado de Assis. **Scripta**, Belo Horizonte, v. 3, n. 6, p. 53-60, jan./jun. 2000.

PHALÈSE, Hubert de. **Comptes a rebours** : l'oeuvre de Huysmans à travers les nouvelles technologies. Paris: Nizet, 1991. (Collection Cap'agreg).

PHALÈSE, Hubert de. **Beckett à la lettre : en attendant Godot, fin de partie**. Paris: Nizet, 1998.

POUND, Ezra. **Abc da literatura**. Tradução de Augusto de Campos e José Paulo Paes. São Paulo: Cultrix, 1990.

PROENÇA FILHO, Domício. **Estilos de época na literatura**. São Paulo: Ática, 2002.

PROPP, V. I. **Morfologia do conto maravilhoso**. Tradução de Jasna Paravich Sarhan. Rio de Janeiro: Forense Universitária, 1984.

QUENEAU, Raymond. **Exercícios de estilo**. Tradução de Luiz Resende. Rio de Janeiro: Imago, 1995. (Coleção Lazuli).

QUIJANO, Graciela R. de. **Metodologia de análise do conto: *Missa do Galo* de Machado de Assis**. 1981. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1981.

RAMA, Angel. Os processos de transculturação na narrativa latino-americana. In: AGUIAR, Flávio; VASCONCELOS, Sandra Gardini T. (Org.). **Ángel Rama: literatura e cultura na América Latina**. Tradução de Rachelza Corte dos Santos e Elza Gasparotto. São Paulo: Edusp, 2000.

RAMSAY, Stephen. Special Section: Reconceiving Text Analysis: Toward an Algorithmic Criticism. In **Literary and Linguistic Computing**, 2003, v. 18, n. 2, p. 167-174.

RASTIER, François. **Textes et sens**. Paris: Didier Érudition, 1996.

RASTIER, François. Stylistique et linguistique des styles. **Arts et Sciences du Texte**, Paris, PUF, p. 167-188, 2001.

RASTIER, François. **Arts et sciences du texte**. Paris: PUF, 2001.

REIS, Carlos; LOPES, Ana Cristina M.. **Dicionário de teoria da narrativa**. São Paulo: Ática, 1988. (Série Fundamentos).

RIEDEL, Dirce Côrtes. **Metáfora: o espelho de Machado de Assis**. Rio de Janeiro: Francisco Alves, 1974.

RIEDEL, Dirce Côrtes. **O tempo no romance machadiano**. Rio de Janeiro: São José, 1959.

ROCKWELL, Geoffrey. What is text analysis, really? **Literary and Linguistic Computing**, v. 18, n. 2, p. 209-219, 2003.

ROMERO, Sílvio. **Machado de Assis**. 2. ed. Rio de Janeiro: José Olympio, 1936.

ROSSI, Paolo. Os filósofos e as máquinas: 1400-1700. Tradução de Federico Carotti. São Paulo: Cia das Letras, 1989.

SÁ REGO, Enylton José de. **O calundu e a panacéia**: Machado de Assis, a sátira menipéia e a tradição luciânica. Rio de Janeiro: Forense Universitária, 1989.

SANTIAGO, Silviano. **Uma literatura nos trópicos**: ensaios sobre dependência cultural. 2. ed. Rio de Janeiro: Rocco, 2000.

SANTOS, Alckmar Luiz dos. **Acerca de uma textualidade informatizada**. Disponível em: <<http://www.cce.ufsc.br/nupill/teoria.html>>. Acesso em: 22 jun. 2006.

SANTOS, Alckmar Luiz dos. **Artefato, artemáquina**. Disponível em: <<http://www.cce.ufsc.br/nupill/teoria.html>>. Acesso em: 22 jun. 2006.

SANTOS, Alckmar Luiz dos. **Criação poética (?) e eletrônica (?)**. Disponível em: <<http://www.cce.ufsc.br/nupill/teoria.html>>. Acesso em: 14 maio 2006.

SANTOS, Alckmar Luiz dos. **Literatura e(m) computador**. Disponível em: <<http://www.cce.ufsc.br/nupill/teoria.html>>. Acesso em: 14 maio 2006.

SANTOS, Alckmar Luiz dos. **Textualidade literária e hipertexto informatizado**. Disponível em: <<http://www.cce.ufsc.br/nupill/teoria.html>>. Acesso em: 14 maio 2006.

SANTOS, João Camilo dos. Algumas reflexões sobre O alienista de Machado de Assis. **Colóquio**, Letras, Lisboa, n. 121-122, p. 41-46, jul./dez. 1991.

SARDINHA, Tony Berber (Org.). **A língua portuguesa no computador**. Campinas (SP): Mercado das letras; São Paulo: Fapesp, 2005. (Coleção As faces da linguística aplicada).

SARDINHA, Tony Berber. Análise multidimensional. **Delta**, São Paulo, v. 16, n. 1, 2000. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext@pid=S0102-44502000000100005>. Acesso em: 24 abr. 2006.

SCANU, Ada Myriam. Hyperbase: um logiciel pour l'analyse textuelle. **Revue des Littératures de l'Union Européenne**. Disponível em: <http://www.rilune.org/dese/tesinepdf/Scanu/Scanu_Litt%E9ratureetinformatique.pdf>. Acesso em: 24 abr. 2006.

SCHILLER, Friedrich. **A educação estética do homem**. Tradução de Roberto Schwarz e Márcio Suzuki. São Paulo: Iluminuras, 1990.

SCHOPENHAUER, Arthur. **Sobre Livros e Leitura**. Tradução de Walter Carlos Costa e Philippe Humblé. Porto Alegre: Paraula, 1993.

SCHÜLLER, Donaldo. **A prosa fraturada**. Porto Alegre: EDUFRGS, 1983.

SCHÜLLER, Donaldo. **Plenitude perdida**: uma análise das seqüências narrativas de Dom Casmurro. Porto Alegre: Movimento, 1978.

SCHWARZ, Roberto. **Ao vencedor as batatas**: forma literária e processo social nos inícios do romance brasileiro. São Paulo: Duas Cidades; Ed. 34, 2000.

SCHWARZ, Roberto. **Duas meninas**. São Paulo: Companhia das Letras, 1997.

SCHWARZ, Roberto. **Um mestre na periferia do capitalismo**: Machado de Assis. São Paulo: Duas Cidades, 1990.

SHORT, Mick; SEMINO, Elena; CULPEPER, Jonathan. Using a corpus for stylistics research: Speech and thought presentation. In: SHORT, Mick; THOMAS, Jenny. **Using corpora for language research**: studies in the honour of Geoffrey Leech. Londres; Nova Iorque, Longman, 1996. p. 110-131.

SIEMENS, Raymond G. A new computer-assisted literary criticism? **Computers and the Humanities**, n. 36, p. 259-267, 2002.

SILVA, Ana Maria Vieira. Contos machadianos: radiografias psicossociais. **Scripta**, Belo Horizonte, v. 3, n. 6, p. 44-52, jan./jun. 2000.

SILVA, Vera Maria Tietzmann. Missa do galo: um processo de iniciação. **Espelho – Revista Machadiana**, Porto Alegre: West Lafayette, n. 1, p. 105-122, 1995.

SINCLAIR, Stéfan. Computer-assisted reading: reconceiving text analysis. **Literary and Linguistic Computing**, v. 16, n. 4, p. 175-184, 2001.

SPITZER, Leo. **Études de Style**. Traduction de Eliane Kaufholz, Alain Coulon et Michel Foucault. Paris: Gallimard, 1970.

SPITZER, Leo. **Lingüística e historia literaria**. 2. ed. Madrid: Biblioteca Románica Hispánica; Gredos, 1968.

STEILEIN, Sandra Maria. O capitão Mendonça: um conto fantástico de Machado de Assis. **Travessia**, Florianópolis, n. 25, p. 32-39, 1992.

STEWART, Larry L. Charles Brockden Brown: quantitative analysis and literary interpretation. **Literary and Linguistic Computing**, v. 18, n. 2, p. 129-138, 2003.

SULLIVAN, Patrícia. O segredo da estrutura em *Uns braços*. **Minas Gerais, Suplemento Literário**, p. 1, 31 maio 1969.

TADIÉ, Jean-Yves. **La critique littéraire au XX^e siècle**. Paris: Belfond, 1987. (Collection Agora).

TODOROV, Tzvetan. **Os gêneros do discurso**. Tradução de Elias Angotti Kossovitch. São Paulo: Martins Fontes, 1980.

VAL, Ana Cristina Pimenta da Costa. Machado de Assis e o fantástico espelho. **Scripta**, Belo Horizonte, v. 3, n. 6, p. 20-29, jan./jun. 2000.

VERÍSSIMO, José. **História da literatura brasileira**. Rio de Janeiro: Francisco Alves, 1916. Disponível em: <<http://geocities.yahoo.com.br/ciberespao/ebooksgratuitos.htm>>. Acesso em: 20 mar. 2009.

VIRGILLO, Carmelo. Love and the *Causa secreta* in the tales of Machado de Assis. **Hispania**, Los Angeles, CA, n. 49, p. 778-786, 1966.

WELLEK, René; WARREN, Austin. **Teoria da literatura**. 2. ed. Tradução de José Palla e Carmo. Rio de Janeiro: Publicações Europa-América, 1971.

WESCHENFELDER, Eládio Vilmar. **A paródia nos contos de Machado de Assis**. Pelotas: UFP, 2000.

APÊNDICE

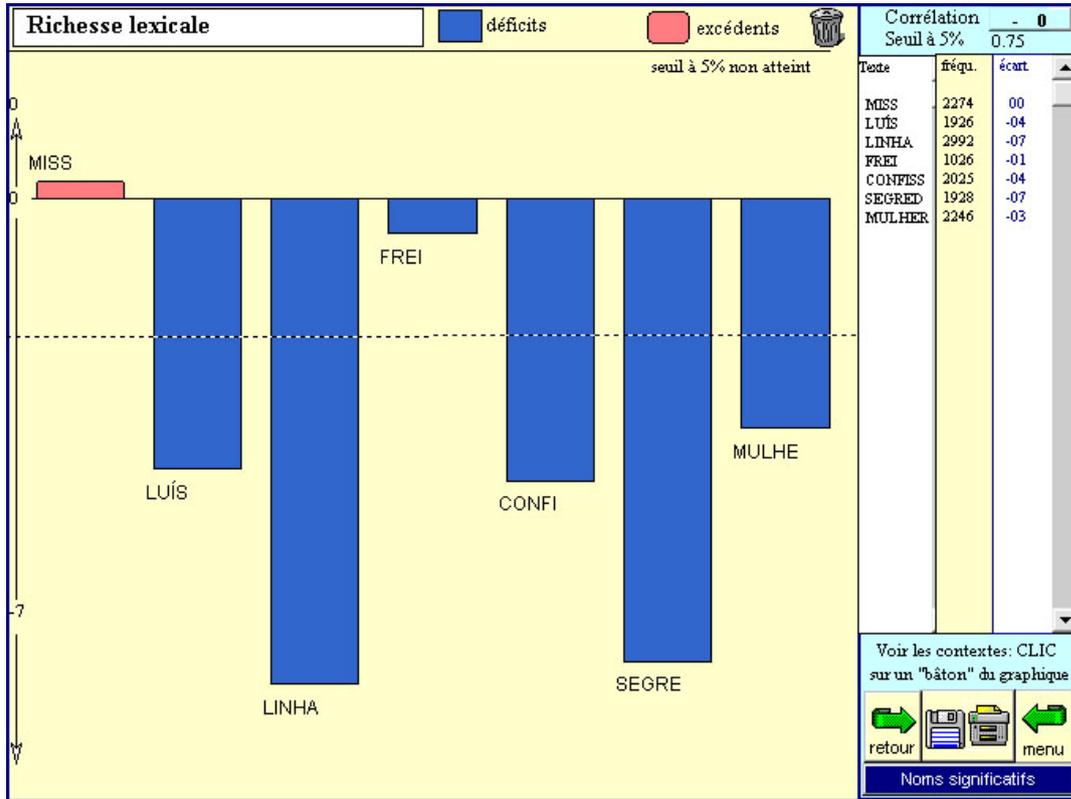


Gráfico 1 – Riqueza lexical em CFLU

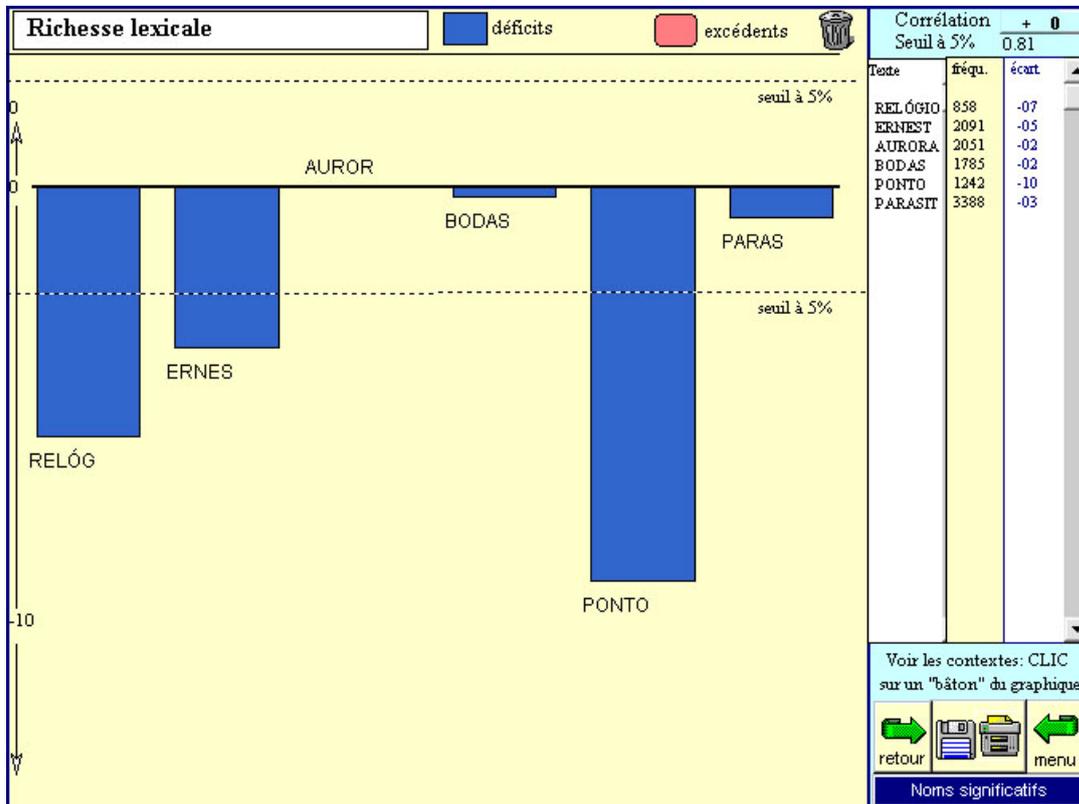


Gráfico 2 – Riqueza lexical em HMN

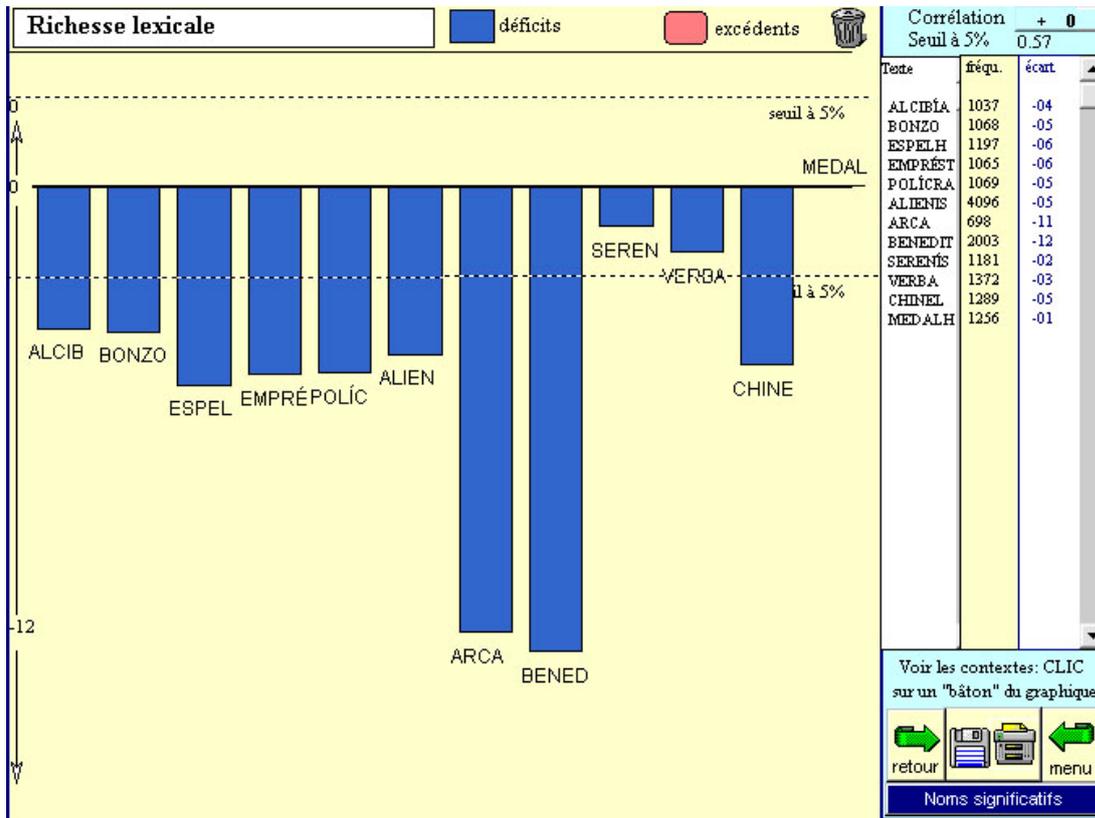


Gráfico 3 – Riqueza lexical em PA

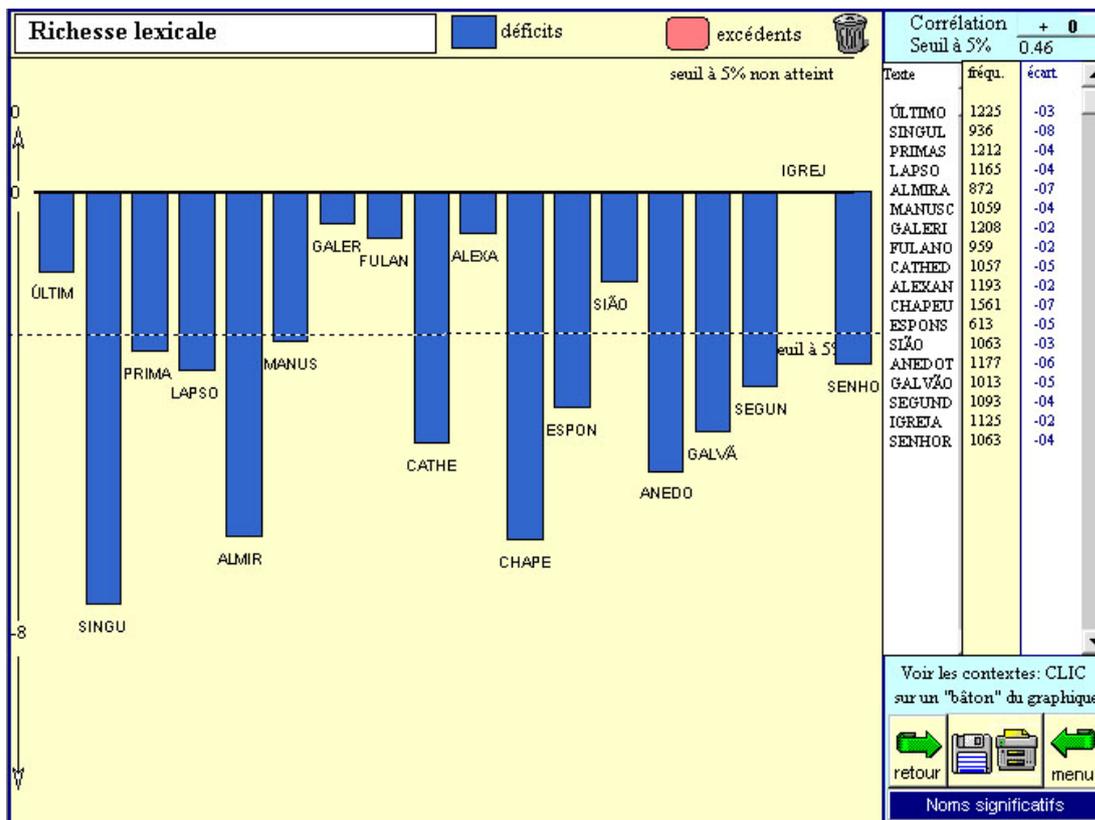


Gráfico 4 – Riqueza lexical em HSD

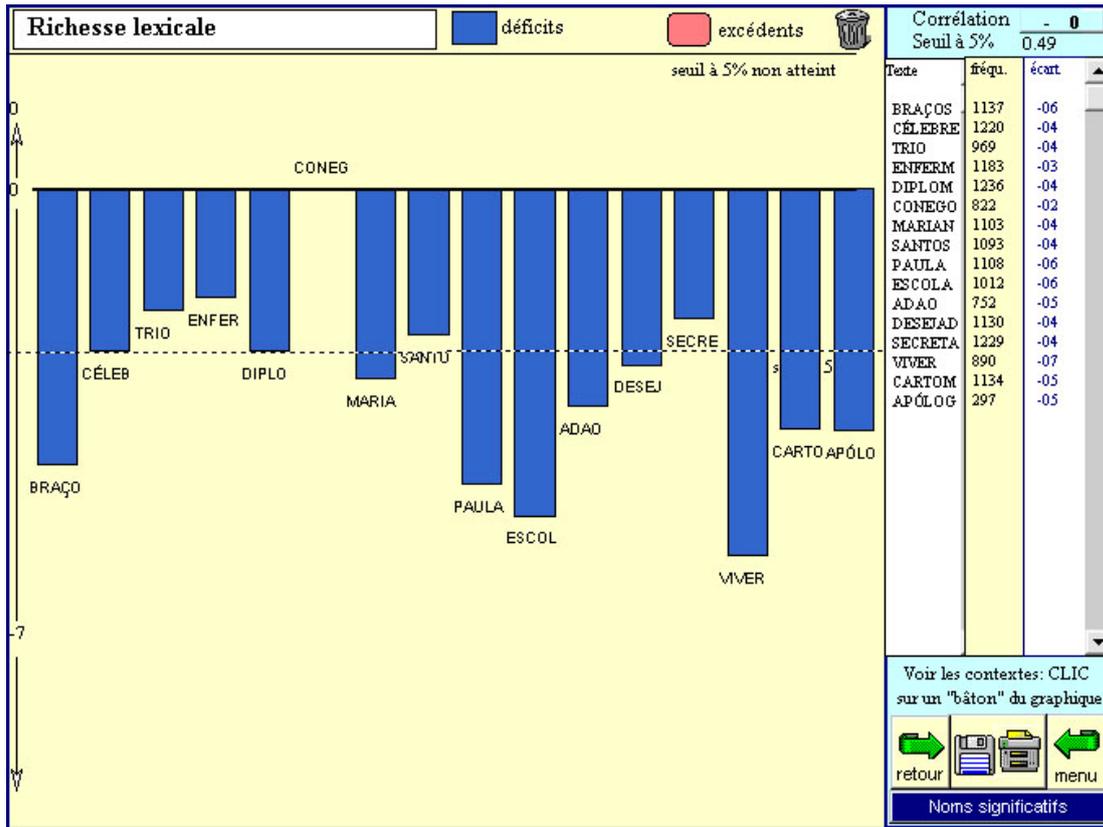


Gráfico 5 – Riqueza lexical em VH

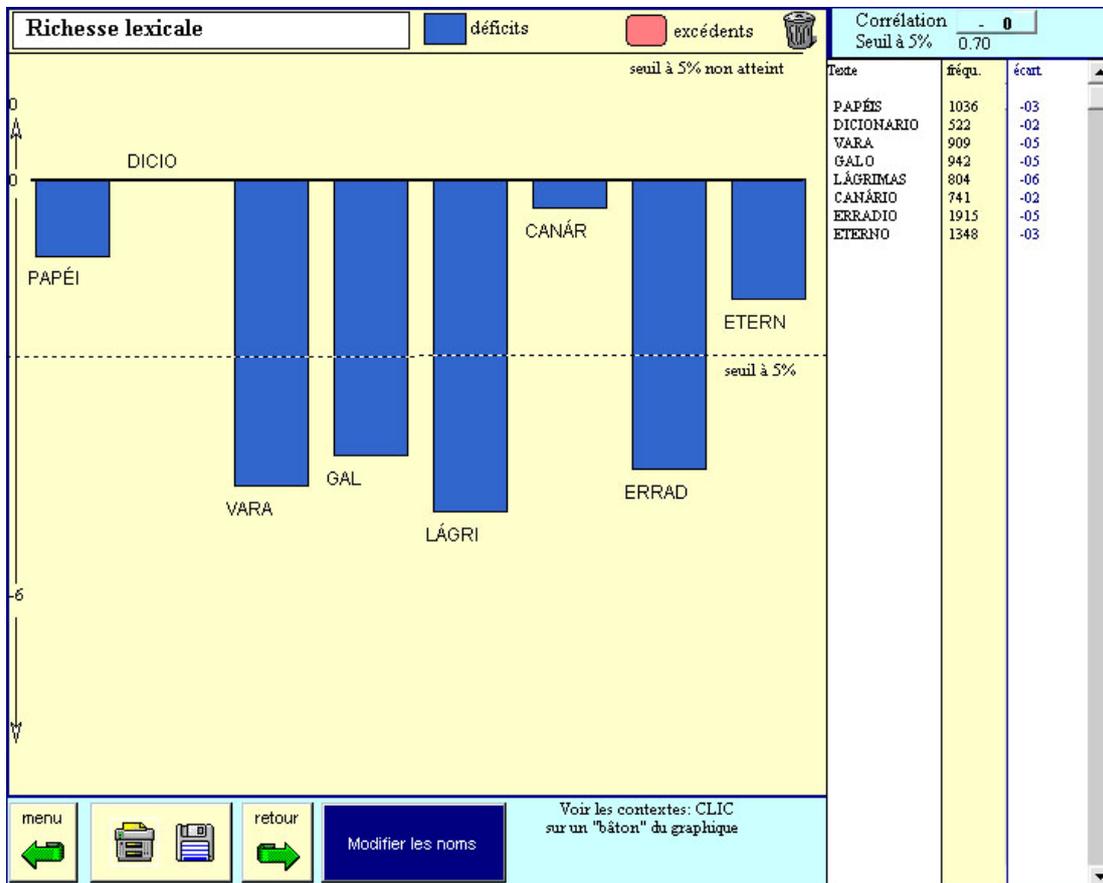


Gráfico 6 – Riqueza lexical em PR

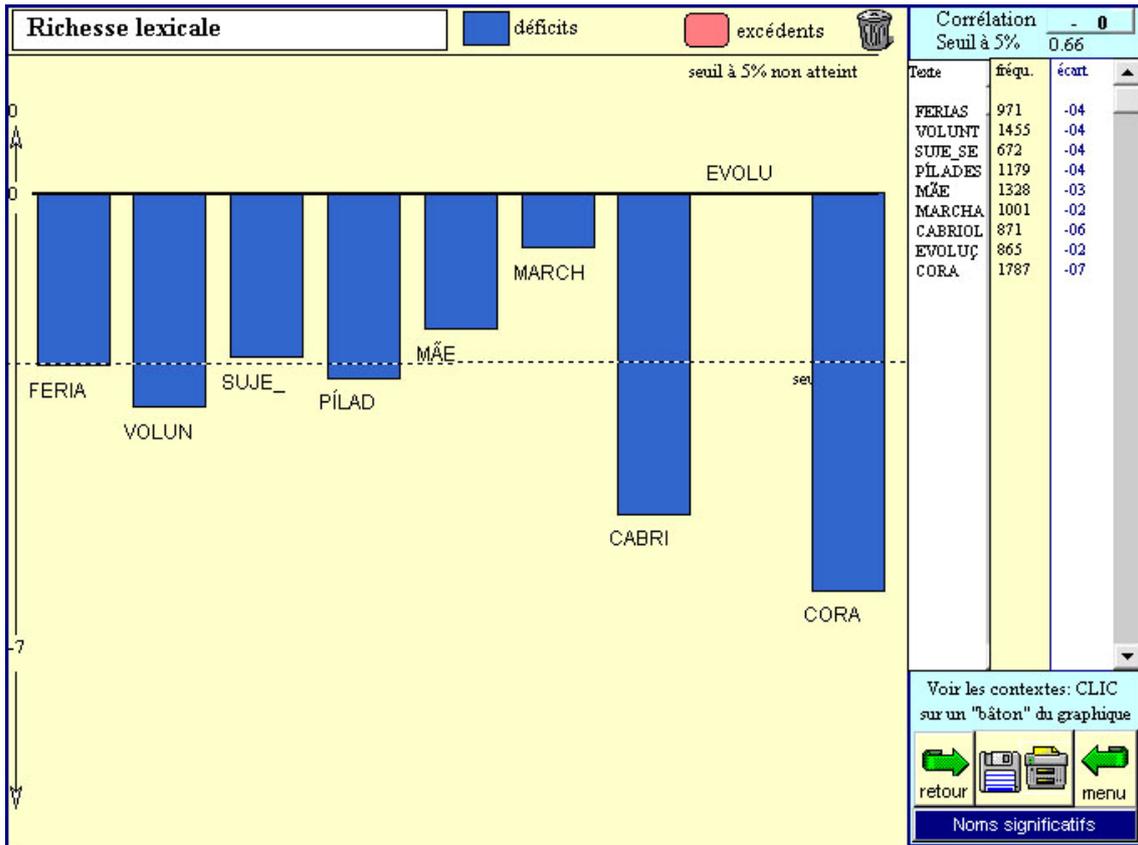


Gráfico 7 – Riqueza lexical em RCV

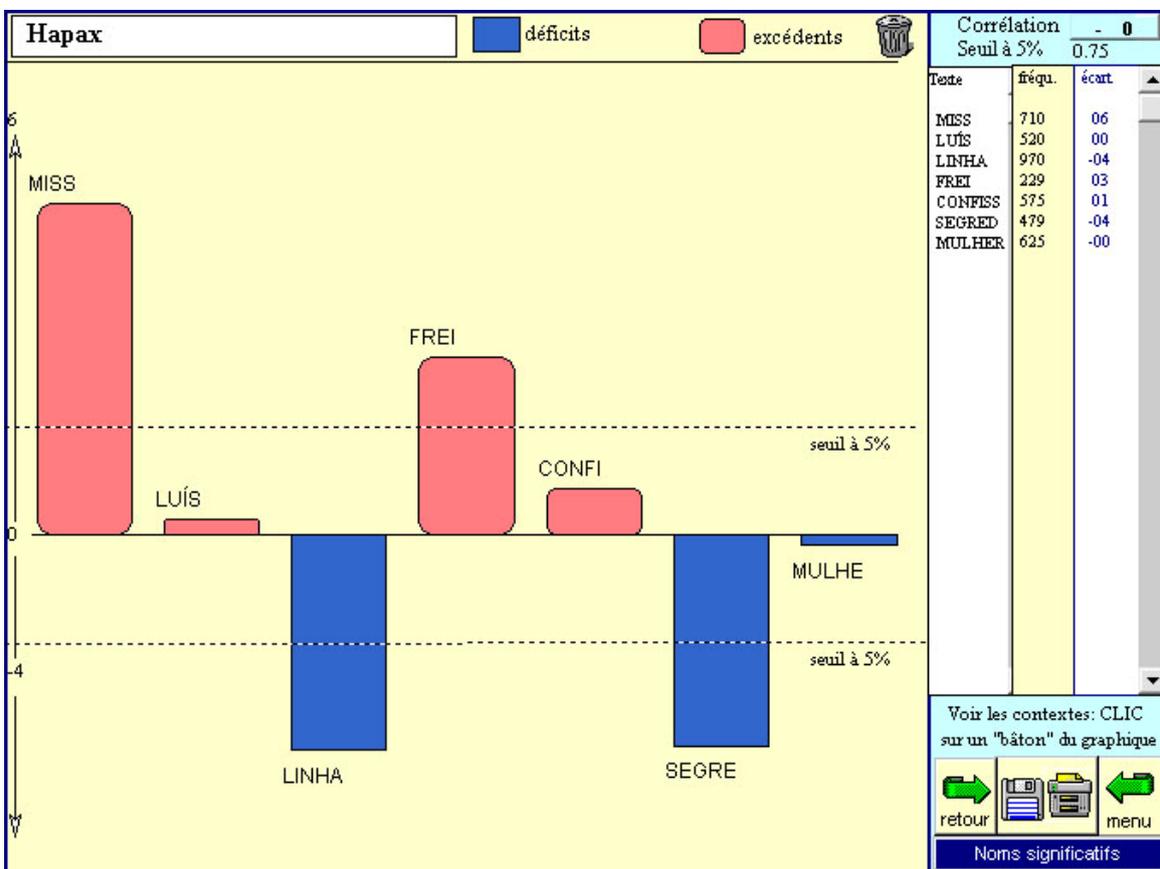


Gráfico 8 – Hápax em CFLU

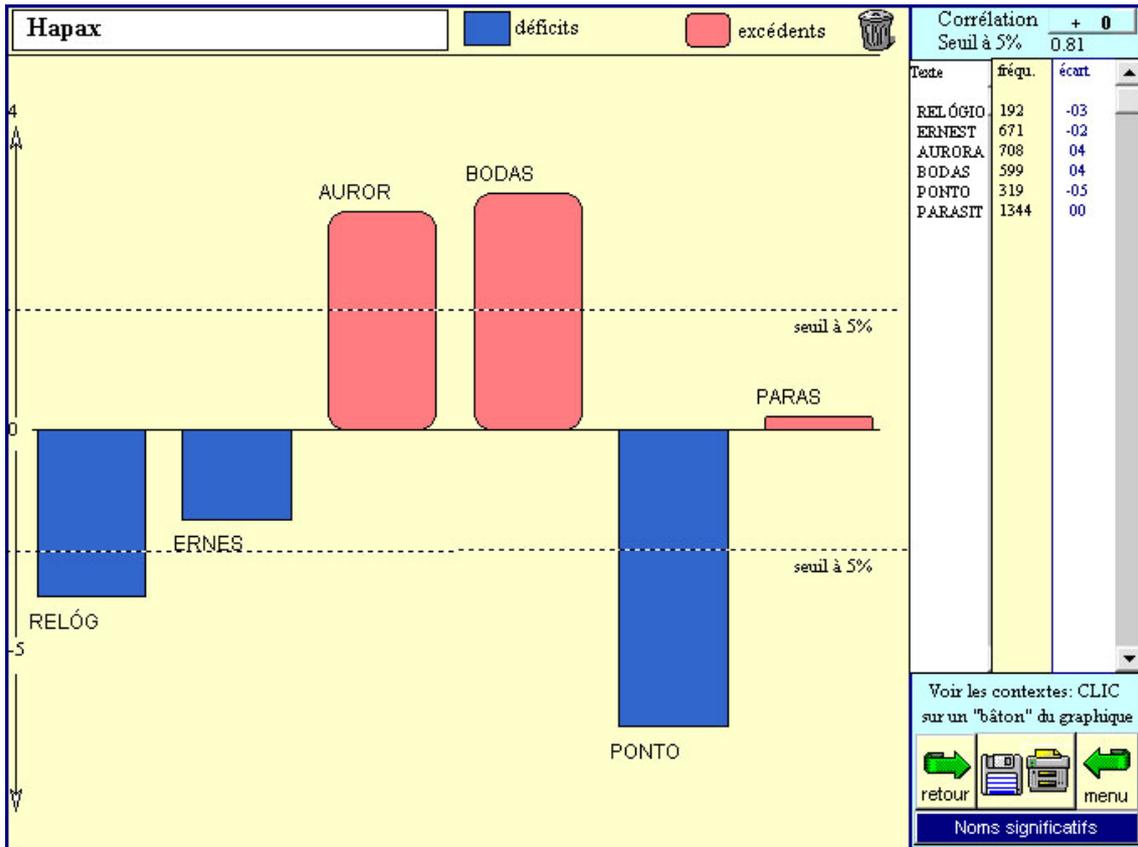


Gráfico 9 – Hápax em HMN

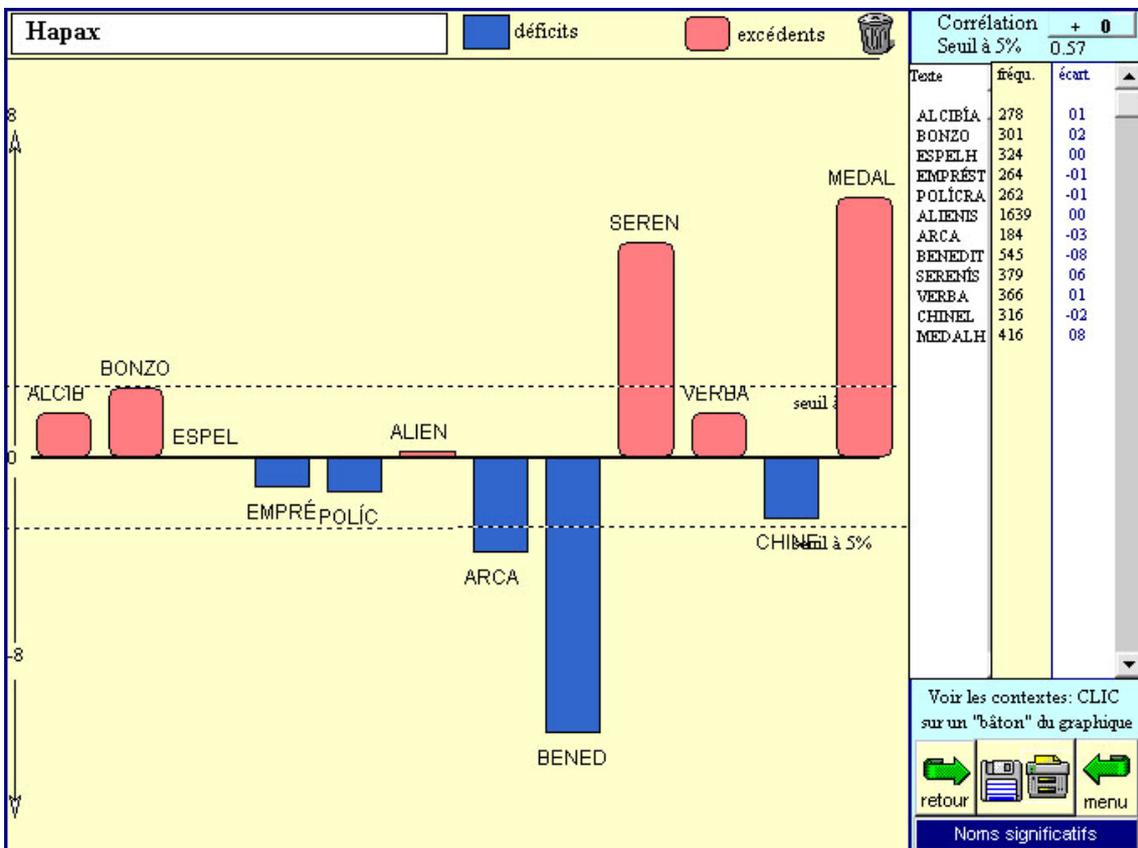


Gráfico 10 – Hápax em PA

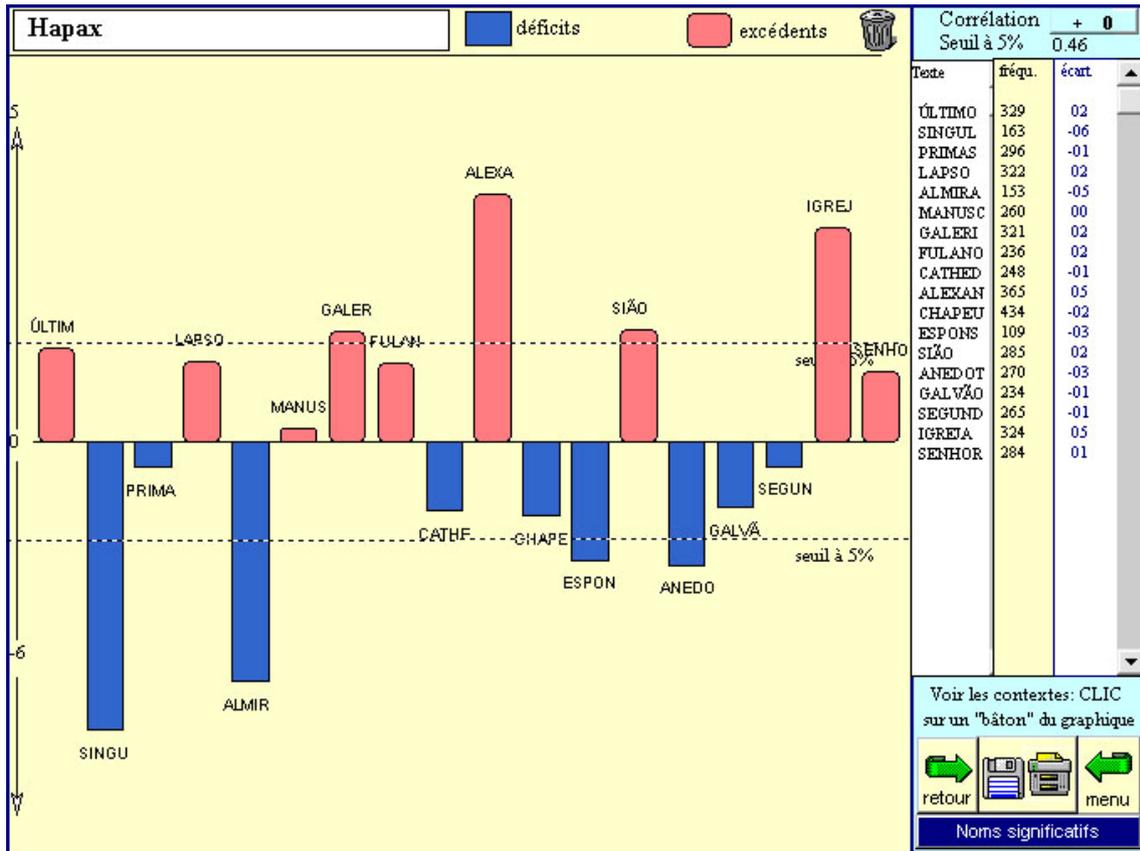


Gráfico 11 – Hápax em HSD

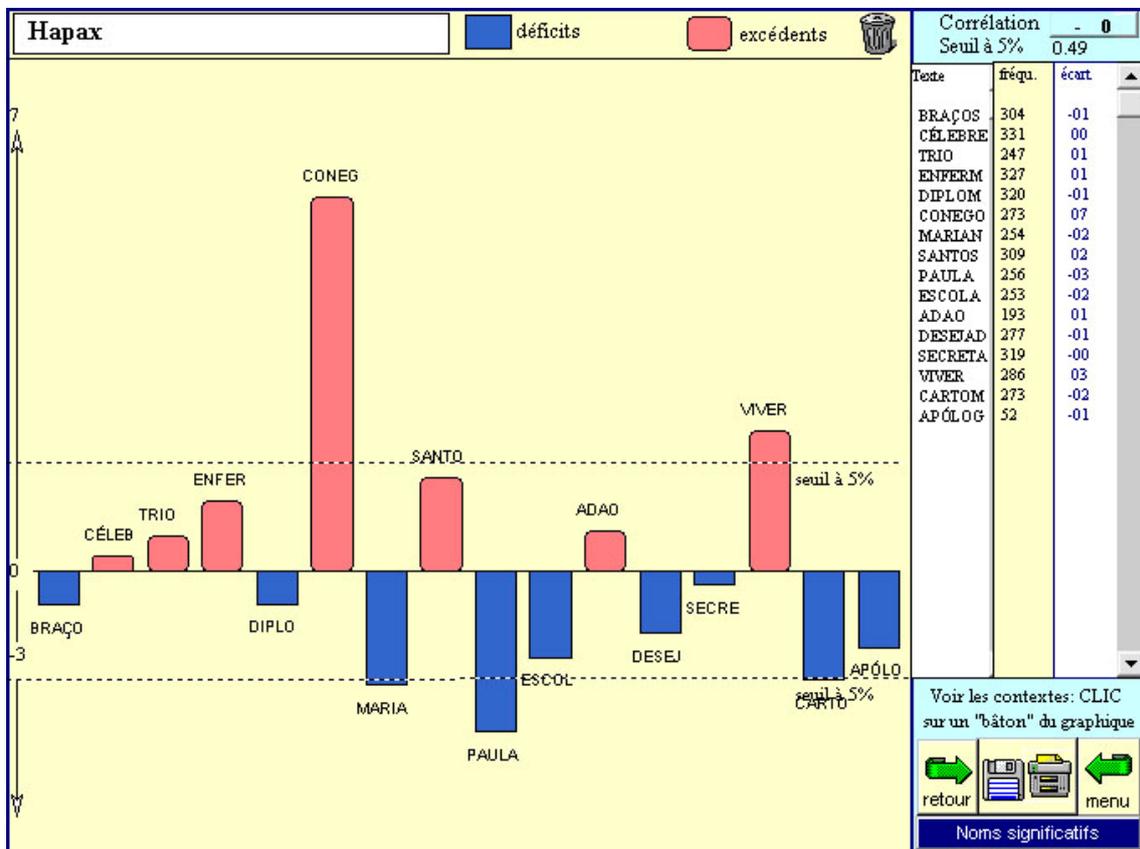


Gráfico 12 – Hápax em VH

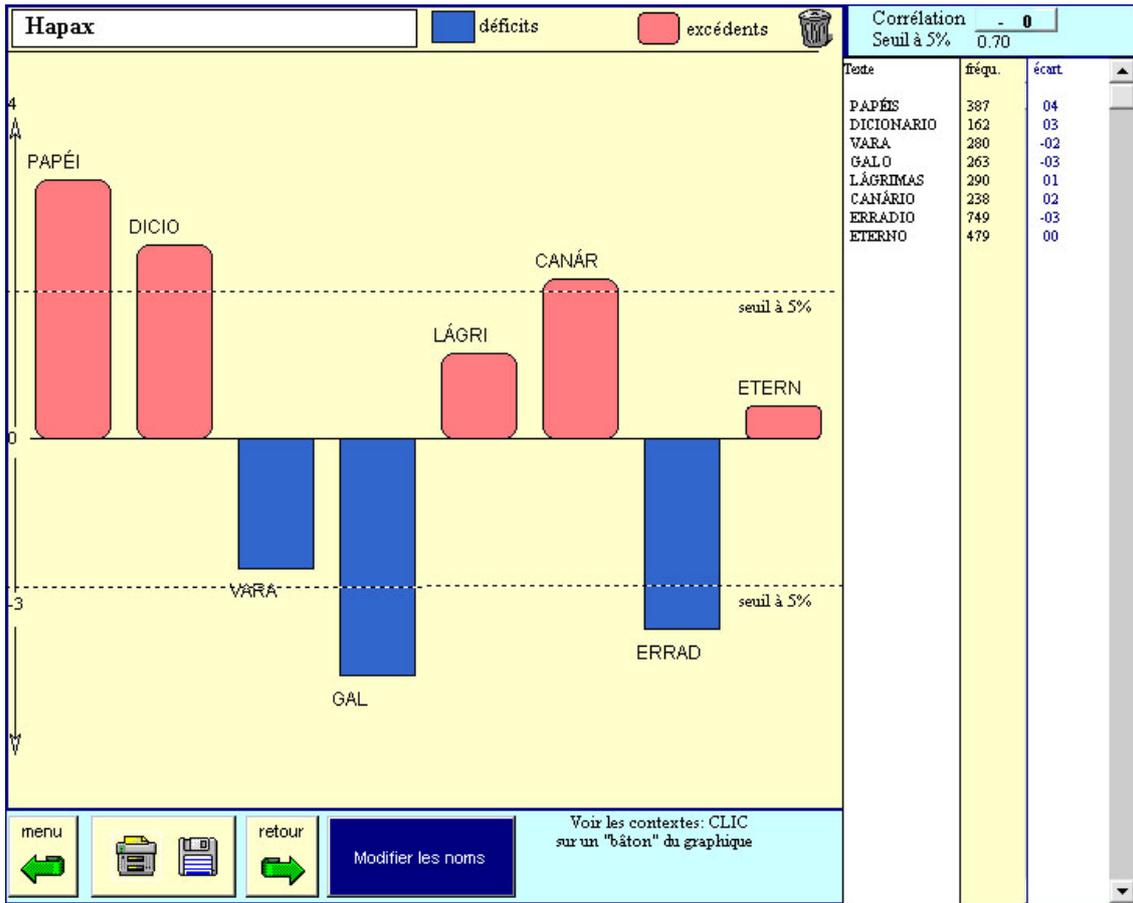


Gráfico 13 – Hápax em PR

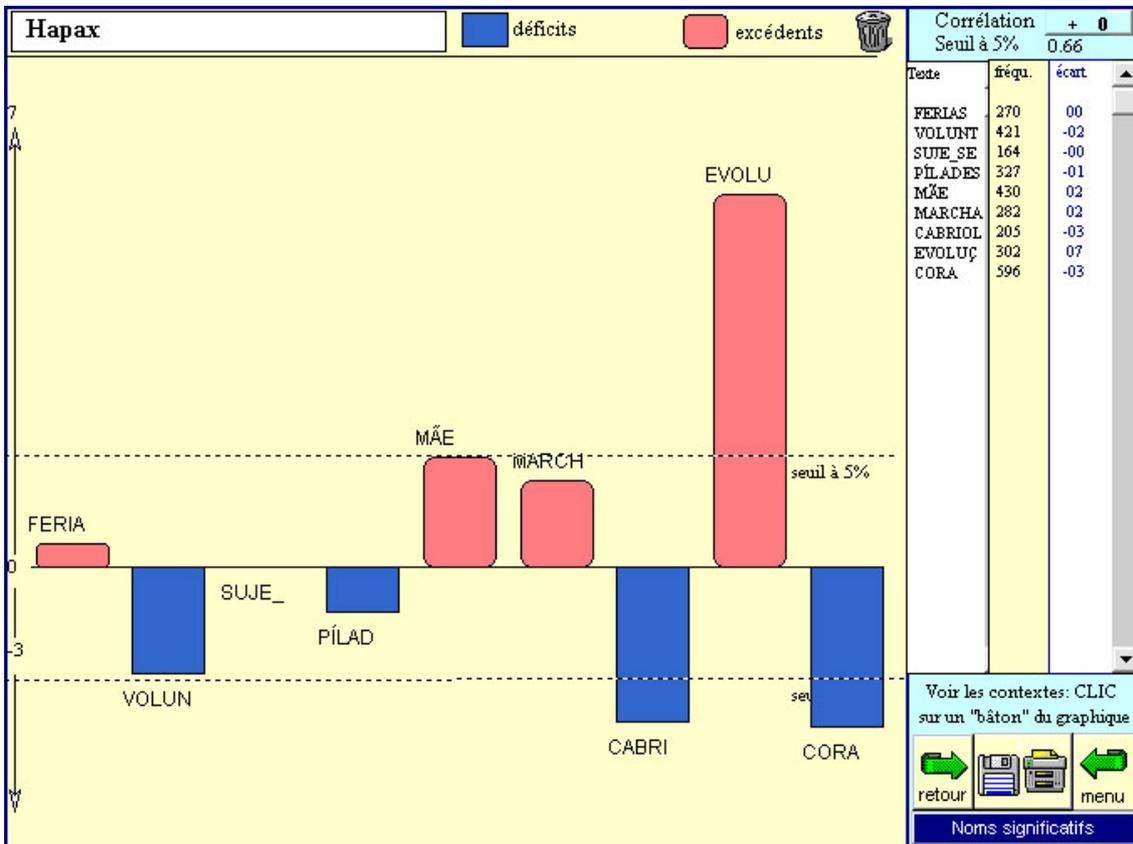


Gráfico 14 – Hápax em RCV

| Ordem Palavras |
|----------------|----------------|----------------|----------------|----------------|
| 1 10291 | 21 61 | 41 25 | 61 19 | 81 9 |
| 2 3712 | 22 97 | 42 25 | 62 13 | 82 9 |
| 3 1830 | 23 80 | 43 14 | 63 13 | 83 8 |
| 4 1237 | 24 66 | 44 21 | 64 13 | 84 4 |
| 5 801 | 25 60 | 45 26 | 65 18 | 85 7 |
| 6 638 | 26 49 | 46 20 | 66 14 | 86 7 |
| 7 501 | 27 55 | 47 22 | 67 9 | 87 11 |
| 8 404 | 28 46 | 48 25 | 68 10 | 88 9 |
| 9 348 | 29 38 | 49 12 | 69 12 | 89 8 |
| 10 273 | 30 40 | 50 16 | 70 8 | 90 3 |
| 11 230 | 31 41 | 51 19 | 71 8 | 91 7 |
| 12 204 | 32 35 | 52 14 | 72 8 | 92 8 |
| 13 190 | 33 29 | 53 21 | 73 13 | 93 3 |
| 14 155 | 34 37 | 54 15 | 74 5 | 94 5 |
| 15 131 | 35 40 | 55 12 | 75 6 | 95 8 |
| 16 129 | 36 30 | 56 14 | 76 4 | 96 10 |
| 17 105 | 37 26 | 57 12 | 77 9 | 97 7 |
| 18 85 | 38 26 | 58 12 | 78 16 | 98 6 |
| 19 93 | 39 18 | 59 18 | 79 10 | 99 2 |
| 20 85 | 40 27 | 60 15 | 80 10 | 100 2 |

Quadro 1 – Distribuição de frequências em AVG1

Ordem Palavra	21 23	42 7	63 3	84 1
1 5687	22 26	43 7	64 1	85 3
2 1672	23 20	44 6	65 3	86 1
3 781	24 16	45 5	66 4	87 1
4 487	25 20	46 7	67 2	88 1
5 312	26 18	47 6	68 4	89 2
6 246	27 9	48 5	69 8	90 0
7 175	28 11	49 6	70 1	91 0
8 133	29 13	50 7	71 6	92 6
9 118	30 12	51 4	72 5	93 0
10 82	31 14	52 3	73 3	94 2
11 78	32 7	53 4	74 3	95 2
12 78	33 10	54 3	75 3	96 1
13 73	34 9	55 6	76 2	97 0
14 65	35 10	56 4	77 2	98 0
15 43	36 14	57 6	78 5	99 2
16 43	37 2	58 3	79 1	100 1
17 34	38 9	59 4	80 2	
18 30	39 4	60 3	81 3	
19 30	40 7	61 2	82 3	
20 29	41 9	62 3	83 1	

Quadro 2 – Distribuição de frequências em Flumeia

Ordem Palavras	20 48	41 10	62 6	83 4
1 9015	21 57	42 11	63 4	84 4
2 3062	22 41	43 14	64 4	85 6
3 1579	23 47	44 9	65 4	86 2
4 973	24 42	45 11	66 6	87 3
5 625	25 33	46 12	67 5	88 2
6 492	26 29	47 10	68 2	89 0
7 323	27 31	48 8	69 7	90 3
8 245	28 31	49 12	70 1	91 3
9 244	29 26	50 8	71 3	92 5
10 188	30 33	51 9	72 4	93 2
11 146	31 31	52 13	73 4	94 3
12 98	32 13	53 4	74 12	95 5
13 121	33 15	54 4	75 5	96 1
14 88	34 20	55 12	76 2	97 4
15 77	35 24	56 6	77 3	98 1
16 74	36 17	57 8	78 2	99 2
17 69	37 14	58 8	79 6	100 3
18 72	38 17	59 4	80 4	
19 53	39 19	60 8	81 3	
	40 13	61 10	82 5	

Quadro 3 – Distribuição de frequências em 5VOL

Ordem Palavra	21 42	42 12	63 8	84 0
1 7659	22 41	43 8	64 10	85 2
2 2381	23 34	44 6	65 5	86 4
3 1261	24 39	45 9	66 3	87 1
4 767	25 25	46 17	67 9	88 2
5 520	26 27	47 8	68 7	89 1
6 370	27 25	48 9	69 5	90 0
7 273	28 28	49 10	70 5	91 1
8 238	29 12	50 8	71 6	92 0
9 186	30 17	51 9	72 3	93 1
10 156	31 23	52 11	73 4	94 3
11 115	32 20	53 10	74 2	95 1
12 104	33 26	54 6	75 4	96 3
13 111	34 13	55 8	76 5	97 2
14 87	35 18	56 5	77 4	98 1
15 75	36 13	57 6	78 3	99 1
16 73	37 17	58 2	79 3	100 4
17 56	38 15	59 10	80 7	
18 55	39 15	60 6	81 2	
19 58	40 11	61 7	82 6	
20 50	41 11	62 4	83 4	

Quadro 4 – Distribuição de frequências em AVG2

Ordem Palavra	21 29	42 14	63 5	84 1
1 7335	22 43	43 9	64 7	85 5
2 2507	23 28	44 10	65 4	86 3
3 1211	24 36	45 8	66 3	87 4
4 775	25 37	46 3	67 6	88 2
5 504	26 27	47 11	68 2	89 2
6 377	27 18	48 9	69 3	90 3
7 280	28 29	49 5	70 5	91 2
8 243	29 18	50 6	71 7	92 2
9 182	30 26	51 13	72 6	93 9
10 160	31 19	52 6	73 4	94 1
11 157	32 17	53 10	74 4	95 2
12 102	33 19	54 8	75 9	96 2
13 79	34 12	55 9	76 5	97 2
14 79	35 20	56 10	77 3	98 0
15 85	36 14	57 2	78 0	99 3
16 64	37 17	58 9	79 2	100 0
17 50	38 16	59 4	80 2	
18 56	39 5	60 12	81 2	
19 45	40 10	61 6	82 3	
20 27	41 8	62 3	83 4	

Quadro 5 – Distribuição de frequências em R1

OrdemPalavra	21 69	42 25	63 10	84 5
1 10344	22 56	43 24	64 6	85 5
2 3640	23 55	44 17	65 8	86 1
3 1977	24 48	45 16	66 16	87 7
4 1180	25 49	46 11	67 6	88 5
5 760	26 35	47 15	68 10	89 10
6 584	27 34	48 16	69 4	90 5
7 461	28 42	49 15	70 10	91 4
8 340	29 39	50 11	71 8	92 6
9 284	30 31	51 16	72 6	93 3
10 249	31 26	52 17	73 7	94 1
11 187	32 26	53 15	74 6	95 7
12 161	33 29	54 16	75 8	96 6
13 141	34 32	55 14	76 11	97 4
14 118	35 26	56 13	77 5	98 0
15 119	36 18	57 6	78 9	99 3
16 100	37 15	58 11	79 4	100 2
17 108	38 16	59 12	80 3	
18 80	39 21	60 10	81 4	
19 72	40 19	61 10	82 4	
20 67	41 27	62 5	83 8	

Quadro 6 – Distribuição de frequências em R2

OrdemPalavra	21 87	42 25	63 19	84 12
1 12019	22 71	43 17	64 11	85 5
2 4386	23 76	44 15	65 6	86 5
3 2436	24 65	45 17	66 16	87 8
4 1440	25 70	46 25	67 14	88 7
5 1050	26 54	47 15	68 10	89 5
6 799	27 63	48 21	69 9	90 7
7 624	28 51	49 20	70 7	91 5
8 416	29 57	50 23	71 8	92 7
9 400	30 52	51 14	72 10	93 9
10 316	31 46	52 15	73 12	94 4
11 293	32 44	53 22	74 14	95 10
12 257	33 30	54 22	75 12	96 4
13 187	34 35	55 20	76 5	97 8
14 195	35 30	56 14	77 5	98 3
15 172	36 36	57 24	78 6	99 11
16 141	37 35	58 17	79 8	100 5
17 127	38 37	59 20	80 4	
18 124	39 26	60 13	81 9	
19 99	40 31	61 11	82 11	
20 86	41 18	62 12	83 7	

Quadro 7 – Distribuição de frequências em todos os romances

OrdemPalavra	21 55	42 24	63 6	84 6
1 10639	22 48	43 12	64 5	85 3
2 3627	23 51	44 19	65 4	86 4
3 1858	24 59	45 16	66 8	87 9
4 1159	25 49	46 17	67 6	88 6
5 789	26 42	47 15	68 7	89 6
6 612	27 30	48 14	69 6	90 4
7 445	28 42	49 13	70 5	91 6
8 325	29 31	50 14	71 5	92 5
9 289	30 41	51 12	72 8	93 4
10 250	31 23	52 22	73 3	94 3
11 212	32 39	53 11	74 5	95 5
12 169	33 24	54 9	75 7	96 4
13 134	34 39	55 11	76 3	97 3
14 118	35 24	56 12	77 4	98 3
15 96	36 27	57 8	78 4	99 8
16 86	37 20	58 11	79 4	100 3
17 96	38 20	59 11	80 2	
18 69	39 29	60 11	81 6	
19 80	40 23	61 9	82 3	
20 73	41 22	62 5	83 4	

Quadro 8 – Distribuição de frequências em 7VOL

Ordem	20	15	41	1	62	4	83	2	
Palavras	21	11	42	6	63	2	84	3	
1	4108	22	12	43	6	64	1	85	4
2	1119	23	12	44	2	65	1	86	3
3	549	24	9	45	8	66	3	87	2
4	335	25	11	46	4	67	2	88	1
5	192	26	9	47	4	68	2	89	0
6	162	27	12	48	3	69	1	90	1
7	103	28	8	49	1	70	0	91	2
8	88	29	3	50	5	71	1	92	0
9	82	30	5	51	5	72	1	93	1
10	64	31	6	52	3	73	1	94	0
11	50	32	6	53	3	74	2	95	1
12	42	33	10	54	4	75	0	96	0
13	41	34	9	55	0	76	1	97	0
14	26	35	5	56	2	77	0	98	4
15	29	36	6	57	2	78	4	99	1
16	28	37	6	58	6	79	0	100	0
17	36	38	3	59	1	80	2		
18	25	39	2	60	1	81	1		
19	17	40	12	61	2	82	1		

Quadro 9 – Distribuição de frequências em CFLU

Ordem	20	13	41	4	62	0	83	0	
Palavras	21	27	42	2	63	2	84	2	
1	3833	22	16	43	7	64	0	85	1
2	1007	23	6	44	2	65	0	86	0
3	470	24	7	45	4	66	0	87	0
4	312	25	5	46	7	67	1	88	0
5	158	26	8	47	2	68	3	89	1
6	134	27	11	48	5	69	0	90	0
7	97	28	4	49	3	70	0	91	0
8	83	29	9	50	1	71	3	92	0
9	63	30	6	51	3	72	5	93	0
10	39	31	7	52	2	73	2	94	0
11	40	32	4	53	1	74	0	95	1
12	33	33	3	54	3	75	3	96	0
13	24	34	6	55	1	76	0	97	0
14	22	35	2	56	3	77	0	98	1
15	24	36	2	57	3	78	2	99	0
16	20	37	7	58	6	79	0	100	0
17	16	38	7	59	1	80	1		
18	17	39	6	60	0	81	1		
19	11	40	3	61	0	82	3		

Quadro 10 – Distribuição de frequências em HMN

Ordem	20	16	41	4	62	1	83	1
Palavras	21	14	42	3	63	2	84	1
1 5274	22	12	43	3	64	1	85	1
2 1505	23	7	44	5	65	0	86	0
3 648	24	11	45	4	66	4	87	1
4 380	25	16	46	2	67	3	88	1
5 252	26	11	47	2	68	0	89	0
6 141	27	8	48	4	69	1	90	1
7 118	28	14	49	1	70	1	91	1
8 98	29	6	50	2	71	0	92	2
9 86	30	6	51	2	72	2	93	1
10 53	31	7	52	3	73	1	94	0
11 52	32	5	53	7	74	1	95	0
12 41	33	6	54	4	75	0	96	1
13 39	34	4	55	2	76	2	97	4
14 26	35	6	56	1	77	1	98	0
15 41	36	4	57	2	78	1	99	0
16 28	37	5	58	2	79	0	100	0
17 23	38	6	59	2	80	2		
18 20	39	2	60	2	81	2		
19 19	40	3	61	4	82	0		

Quadro 11 – Distribuição de frequências em PA

Ordem	20	13	41	2	62	1	83	2
Palavras	21	8	42	3	63	1	84	2
1 4898	22	9	43	4	64	2	85	0
2 1337	23	17	44	7	65	1	86	0
3 588	24	16	45	2	66	3	87	0
4 349	25	12	46	1	67	0	88	0
5 237	26	13	47	1	68	1	89	1
6 174	27	9	48	0	69	1	90	1
7 111	28	6	49	3	70	2	91	1
8 106	29	6	50	2	71	1	92	0
9 74	30	3	51	3	72	0	93	0
10 67	31	9	52	5	73	1	94	0
11 47	32	9	53	4	74	3	95	0
12 37	33	5	54	3	75	2	96	0
13 36	34	5	55	4	76	1	97	3
14 33	35	7	56	0	77	3	98	0
15 24	36	7	57	3	78	0	99	0
16 20	37	4	58	0	79	0	100	0
17 28	38	2	59	4	80	0		
18 18	39	3	60	0	81	2		
19 18	40	5	61	2	82	0		

Quadro 12 – Distribuição de frequências em HSD

Ordem	20	23	41	4	62	0	83	1	
Palavras	21	17	42	3	63	0	84	1	
1	4274	22	11	43	5	64	0	85	0
2	1164	23	7	44	6	65	3	86	2
3	502	24	13	45	3	66	0	87	1
4	316	25	6	46	3	67	0	88	0
5	199	26	5	47	3	68	0	89	0
6	130	27	11	48	1	69	1	90	1
7	82	28	2	49	2	70	3	91	0
8	73	29	4	50	3	71	4	92	1
9	79	30	10	51	5	72	0	93	0
10	49	31	4	52	3	73	0	94	0
11	43	32	8	53	1	74	2	95	3
12	39	33	2	54	1	75	0	96	0
13	35	34	7	55	0	76	1	97	1
14	20	35	3	56	1	77	0	98	0
15	29	36	3	57	4	78	0	99	1
16	23	37	7	58	1	79	1	100	2
17	17	38	6	59	2	80	1		
18	14	39	3	60	2	81	0		
19	17	40	2	61	4	82	0		

Quadro 13 – Distribuição de frequências em VH

Ordem	20	9	41	0	62	1	83	0	
Palavras	21	10	42	1	63	1	84	2	
1	2848	22	7	43	1	64	0	85	0
2	742	23	3	44	0	65	1	86	0
3	349	24	8	45	3	66	1	87	0
4	190	25	6	46	1	67	1	88	0
5	130	26	3	47	1	68	0	89	1
6	81	27	7	48	3	69	0	90	0
7	65	28	5	49	1	70	1	91	0
8	53	29	1	50	0	71	0	92	0
9	37	30	2	51	0	72	0	93	0
10	24	31	1	52	1	73	0	94	0
11	24	32	1	53	0	74	0	95	0
12	19	33	0	54	1	75	1	96	0
13	17	34	5	55	1	76	0	97	1
14	18	35	1	56	0	77	0	98	0
15	11	36	2	57	1	78	0	99	0
16	11	37	4	58	2	79	0	100	0
17	11	38	1	59	0	80	2		
18	14	39	2	60	0	81	0		
19	9	40	1	61	1	82	0		

Quadro 14 – Distribuição de frequências em PR

Ordem	20	5	41	3	62	1	83	1
Palavras	21	11	42	1	63	0	84	0
1 2997	22	8	43	4	64	3	85	1
2 866	23	11	44	3	65	1	86	2
3 353	24	8	45	2	66	1	87	2
4 207	25	6	46	3	67	0	88	0
5 153	26	3	47	1	68	1	89	0
6 102	27	4	48	4	69	0	90	0
7 89	28	4	49	1	70	2	91	0
8 59	29	8	50	2	71	1	92	0
9 58	30	7	51	0	72	0	93	0
10 36	31	7	52	2	73	1	94	0
11 41	32	4	53	2	74	1	95	0
12 28	33	3	54	0	75	0	96	1
13 26	34	6	55	0	76	1	97	0
14 26	35	2	56	1	77	0	98	0
15 7	36	0	57	1	78	0	99	0
16 13	37	2	58	2	79	1	100	0
17 11	38	1	59	1	80	1		
18 7	39	4	60	1	81	0		
19 13	40	1	61	0	82	0		

Quadro 15 – Distribuição de frequências em RCV

Ordem/ freq. / palavra		
1 3860 .	34 272 :	68 106 isso
2 3385 ,	35 264 como	69 104 nada
3 3013 -	36 253 ele	70 104 esta
4 2164 a	37 247 à	71 103 sem
5 2042 que	38 229 mais	72 101 homem
6 1880 de	39 221 na	73 99 tudo
7 1727 o	40 194 foi	74 98 soares
8 1395 e	41 193 meu	75 98 marido
9 1132 não	42 187 quando	76 98 lo
10 1002 se	43 187 adelaide	77 98 das
11 921 ;	44 178 casa	78 95 muito
12 766 um	45 153 sua	79 93 olhos
13 720 é	46 152 minha	80 91 perguntou
14 633 ?	47 150 tinha	81 91 então
15 583 do	48 148 tito	82 90 dous
16 545 uma	49 144 dos	83 88 ou
17 484 com	50 143 já	84 87 porque
18 483 da	51 143 carta	85 87 mesmo
19 479 me	52 140 estêvão	86 86 sei
20 475 mas	53 136 mendonça	87 86 isto
21 454 para	54 130 vasconcelos	88 86 bem
22 453 era	55 129 te	89 85 verdade
23 450 lhe	56 126 dia	90 85 só
24 446 em	57 125 depois	91 85 diogo
25 425 eu	58 124 há	92 85 cousa
26 386 ...	59 120 amor	93 84 nos
27 358 !	60 119 emília	94 84 mim
28 351 disse	61 116 ela	95 84 ainda
29 346 os	62 114 tempo	96 83 mulher
30 342 por	63 112 rapaz	97 83 até
31 328 no	64 110 nem	98 82 quem
32 295 ao	65 109 estava	99 81 está
33 288 as	66 108 seu	100 80 tão
	67 108 ser	

Quadro 16 – Índice hierárquico de altas frequências em CFLU

Ordem/ freq. / palavra		
1 2840 ,	34 185 por	68 73 dizer
2 2477 .	35 183 eu	69 72 tinoco
3 1626 a	36 167 disse	70 72 olhos
4 1596 que	37 158 camilo	71 72 nem
5 1517 -	38 151 foi	72 72 depois
6 1497 o	39 147 ...	73 72 até
7 1435 de	40 140 quando	74 71 todos
8 1138 e	41 137 como	75 71 meu
9 727 não	42 131 :	76 71 isabel
10 589 se	43 128 luís	77 68 tinha
11 551 do	44 128 dos	78 68 ter
12 528 ;	45 119 moça	79 68 casa
13 489 um	46 119 ela	80 67 ser
14 404 lhe	47 117 já	81 63 minha
15 390 da	48 113 ernesto	82 63 esta
16 385 com	49 104 ou	83 59 rosina
17 375 os	50 102 sua	84 58 tem
18 371 uma	51 98 tempo	85 58 outro
19 327 é	52 95 estava	86 58 homem
20 310 ?	53 89 tudo	87 58 este
21 290 para	54 85 seu	88 58 coração
22 287 era	55 84 dr	89 58 carta
23 286 em	56 84 ainda	90 57 só
24 269 ao	57 82 seus	91 57 então
25 261 me	58 82 muito	92 57 dia
26 253 mas	59 82 há	93 56 tenente
27 245 !	60 81 rapaz	94 56 pai
28 224 no	61 80 lemos	95 56 mesmo
29 218 ele	62 78 das	96 55 alguns
30 213 as	63 78 "	97 54 tão
31 211 na	64 75 sem	98 54 assim
32 205 à	65 75 quem	99 54 às
33 202 mais	66 75 d	100 53 apenas
	67 73 porque	

Quadro 17 – Índice hierárquico de altas frequências em HMN

Ordem/ freq. / palavra		
1 5708 ,	34 230 na	68 81 verde
2 2710 .	35 220 me	69 81 ela
3 2137 a	36 219 :	70 80 olhos
4 1934 -	37 217 ?	71 80 estava
5 1818 o	38 205 "	72 78 tempo
6 1777 de	39 182 como	73 77 quando
7 1776 que	40 175 foi	74 76 todos
8 1713 e	41 169 d	75 76 duas
9 942 não	42 163 casa	76 74 aos
10 901 ;	43 154 das	77 73 este
11 858 um	44 150 sem	78 72 menos
12 767 do	45 146 ...	79 72 às
13 688 se	46 126 disse	80 70 esta
14 618 os	47 120 nada	81 69 nos
15 616 da	48 109 outra	82 67 três
16 485 uma	49 106 alienista	83 67 meu
17 466 com	50 104 nem	84 67 dois
18 404 lhe	51 102 tinha	85 66 ter
19 399 as	52 101 outro	86 66 também
20 396 era	53 97 tudo	87 66 coisa
21 387 em	54 97 tão	88 66 assim
22 377 ao	55 97 mesmo	89 64 logo
23 375 é	56 97 depois	90 63 la
24 374 mas	57 96 porque	91 63 anos
25 299 para	58 93 ser	92 62 podia
26 294 à	59 92 lo	93 61 isto
27 265 dos	60 92 benedita	94 61 cinco
28 264 ele	61 91 ainda	95 61 alguns
29 262 no	62 90 só	96 61 agora
30 258 por	63 88 eu	97 60 outros
31 255 mais	64 87 dia	98 60 então
32 240 ou	65 85 homem	99 59 seus
33 235 !	66 84 muito	100 59 bacamarte
	67 83 há	

Quadro 18 – Índice hierárquico de altas frequências em PA

Ordem/ freq. / palavra		
1	5479 ,	
2	2784 .	
3	2019 a	
4	1887 -	
5	1713 que	
6	1696 e	
7	1674 de	
8	1476 o	
9	977 ;	
10	917 não	
11	797 um	
12	649 se	
13	645 do	
14	602 os	
15	522 da	
16	506 lhe	
17	489 uma	
18	479 era	
19	476 com	
20	440 as	
21	412 mas	
22	404 para	
23	376 em	
24	329 é	
25	311 ao	
26	287 me	
27	284 no	
28	279 na	
29	278 à	
30	253 ele	
31	252 ?	
32	243 por	
33	233 ou	
34	212 mais	
35	202 ...	
36	201 como	
37	181 dos	
38	174 :	
39	168 ela	
40	162 foi	
41	150 nem	
42	144 ”	
43	132 !	
44	131 tinha	
45	130 das	
46	127 depois	
47	123 tudo	
48	123 nada	
49	120 muito	
50	113 casa	
51	112 disse	
52	111 eu	
53	109 eram	
54	108 tão	
55	106 outro	
56	106 mesmo	
57	103 sem	
58	101 anos	
59	97 só	
60	97 outra	
61	97 lo	
62	91 quando	
63	90 até	
64	89 ia	
65	84 três	
66	84 porque	
67	83 tempo	
68	83 olhos	
69	81 dia	
70	81 cousa	
71	77 pouco	
72	77 nos	
73	77 assim	
74	76 vez	
75	75 dous	
76	75 às	
77	74 homem	
78	74 aos	
79	74 ainda	
80	73 outras	
81	71 este	
82	70 vida	
83	70 estava	
84	69 já	
85	68 la	
86	66 também	
87	66 minha	
88	66 há	
89	65 rua	
90	64 verdade	
91	64 esta	
92	63 ser	
93	62 grande	
94	61 meu	
95	61 mariana	
96	59 marido	
97	59 então	
98	59 eles	
99	59 alguns	
100	57 noite	

Quadro 19 – Índice hierárquico de altas frequências em HSD

Ordem/ freq. / palavra
1 4975 ,
2 2631 .
3 1866 -
4 1756 a
5 1664 e
6 1604 que
7 1367 de
8 1270 o
9 782 ;
10 749 não
11 600 se
12 508 os
13 501 da
14 493 do
15 483 um
16 430 lhe
17 365 as
18 361 para
19 360 era
20 359 uma
21 348 com
22 327 mas
23 309 em
24 309 ao
25 295 me
26 270 ele
27 260 é
28 235 na
29 232 ?
30 230 no
31 217 à
32 211 por
33 211 mais

34 200 ...
35 196 foi
36 182 ou
37 179 como
38 150 ela
39 140 tudo
40 133 dos
41 131 depois
42 131 !
43 129 nem
44 127 nada
45 125 olhos
46 125 eu
47 125 :
48 115 tinha
49 115 outro
50 115 das
51 104 sem
52 104 quando
53 104 estava
54 100 tempo
55 100 casa
56 99 disse
57 97 só
58 95 muito
59 95 ia
60 95 ainda
61 92 lo
62 90 "
63 87 d
64 86 mesmo
65 86 dia
66 84 também
67 83 até

68 80 outra
69 79 então
70 76 vida
71 74 vez
72 74 assim
73 71 porque
74 71 já
75 71 cousa
76 71 anos
77 70 tão
78 70 ser
79 70 pouco
80 69 nos
81 65 rua
82 65 há
83 65 alguma
84 61 mesma
85 61 la
86 61 bem
87 61 aqui
88 60 todos
89 60 às
90 59 podia
91 59 homem
92 58 logo
93 57 toda
94 57 noite
95 57 eram
96 57 agora
97 56 ver
98 54 alguns
99 53 três
100 52 tu

Quadro 20 – Índice hierárquico de altas frequências em VH

Ordem/ freq. / palavra		
1 2454 ,	34 102 no	68 39 outra
2 1398 .	35 97 ...	69 38 ser
3 898 que	36 89 ou	70 37 rita
4 851 a	37 84 nem	71 37 outro
5 849 -	38 84 como	72 37 mesmo
6 682 de	39 80 foi	73 37 às
7 677 e	40 80 !	74 36 mim
8 660 o	41 75 ela	75 36 ”
9 492 não	42 70 tinha	76 35 sinhá
10 367 ;	43 67 disse	77 34 também
11 297 se	44 66 dos	78 34 pouco
12 289 um	45 65 :	79 34 olhos
13 263 do	46 63 das	80 34 eram
14 245 me	47 62 casa	81 34 amigo
15 217 da	48 61 elisiário	82 32 rua
16 212 os	49 58 quando	83 31 tão
17 202 uma	50 58 meu	84 30 ver
18 200 lhe	51 57 sem	85 30 porque
19 197 para	52 55 tudo	86 29 mundo
20 191 com	53 54 nos	87 28 logo
21 180 era	54 52 lo	88 28 então
22 173 as	55 49 depois	89 28 até
23 169 mas	56 48 nada	90 28 aqui
24 165 em	57 48 muito	91 28 anos
25 157 ?	58 48 já	92 27 sim
26 151 é	59 47 ia	93 27 pela
27 135 ao	60 46 ainda	94 27 noite
28 121 ele	61 45 só	95 27 marido
29 120 eu	62 45 minha	96 27 lá
30 119 mais	63 45 estava	97 27 duas
31 114 por	64 43 há	98 27 dele
32 114 na	65 42 tempo	99 26 dia
33 104 à	66 40 homem	100 26 ali
	67 39 vez	

Quadro 21 – Índice hierárquico de altas frequências em PR

Ordem/ freq. / palavra
1 3005 ,
2 1795 .
3 1155 a
4 1107 que
5 1079 -
6 962 de
7 927 e
8 872 o
9 640 não
10 458 ;
11 376 se
12 344 da
13 333 me
14 302 do
15 298 um
16 298 era
17 292 com
18 288 os
19 243 lhe
20 229 uma
21 227 em
22 225 ao
23 223 para
24 197 mas
25 194 à
26 182 as
27 177 é
28 176 por
29 161 mais
30 154 eu
31 154 ele
32 150 na
33 143 foi
34 133 como

35 127 ou
36 126 ?
37 123 nem
38 118 no
39 103 dos
40 96 ...
41 87 disse
42 87 casa
43 86 tinha
44 86 quando
45 85 "
46 83 depois
47 80 sem
48 79 maria
49 76 ela
50 74 tudo
51 73 minha
52 71 lo
53 70 ia
54 70 das
55 68 também
56 66 meu
57 65 ser
58 64 vez
59 64 nos
60 64 ainda
61 62 só
62 60 já
63 59 tempo
64 58 rua
65 58 !
66 57 :
67 56 dia
68 53 outro
69 53 alguma

70 52 quintanilha
71 52 estava
72 50 pouco
73 50 mesmo
74 49 nada
75 48 tia
76 48 la
77 48 este
78 48 até
79 47 gonçaves
80 46 tão
81 46 joão
82 46 então
83 45 outra
84 45 olhos
85 44 eram
86 44 dous
87 44 assim
88 43 vida
89 43 noite
90 43 cousa
91 43 agora
92 42 muito
93 41 tal
94 41 grande
95 41 dizer
96 40 amigo
97 39 podia
98 39 lá
99 39 cora
100 39 antes

Quadro 22 – Índice hierárquico de altas frequências em RCV

Palavras em progressão			Palavras em regressão		
Coef. / Freq. / Palavra			Coef. / Freq. / Palavra		
+ 0.662	202	dous	- 0.450	121	dois
+ 0.421	49	segundo	- 0.426	36	haver
+ 0.417	8	subiu	- 0.423	36	conhecia
+ 0.405	22	eis	- 0.402	60	deste
+ 0.390	14	casaram	- 0.395	13	amar
+ 0.387	231	cousa	- 0.393	77	coisas
+ 0.385	129	primeiro	- 0.371	20	passara
+ 0.376	8	modesto	- 0.363	195	coisa
+ 0.373	23	anterior	- 0.358	10	lançou
+ 0.351	9	fiel	- 0.349	24	saía
+ 0.344	14	políticas	- 0.345	25	riso
+ 0.331	715	ou	- 0.342	19	murmurou
+ 0.331	40	fiz	- 0.341	9	desapareceu
+ 0.329	35	justamente	- 0.340	16	testa
+ 0.329	19	dezoito	- 0.339	20	fechou
+ 0.328	176	tal	- 0.335	8	aceita
+ 0.326	22	estão	- 0.334	15	procurou
+ 0.326	20	destino	- 0.322	59	diante
+ 0.325	119	veio	- 0.322	27	estas
			- 0.322	18	certeza
			- 0.321	12	duras

Quadro 23 – Palavras em progressão e em regressão de uso em AVG2

Palavras em progressão	Palavras em regressão
Coef. / Freq. / Palavra	Coef. / Freq. / Palavra
+ 0.001 6225 ,	- 0.001 1591 se
+ 0.001 192 ou	- 0.001 527 por
+ 0.001 96 assim	- 0.001 246 casa
+ 0.001 92 dr	- 0.001 218 Tinha
+ 0.001 81 mão	- 0.001 183 dia
+ 0.001 73 si	- 0.001 152 amor
+ 0.001 73 logo	- 0.001 151 te
+ 0.001 72 cabeça	- 0.001 121 nos
+ 0.001 70 porém	- 0.001 112 dous
+ 0.001 57 aquele	- 0.001 77 pois
+ 0.001 57 ali	- 0.001 57 hora
+ 0.001 56 Tenente	- 0.001 48 maior
+ 0.001 55 tal	- 0.001 35 quarto
+ 0.001 49 moço	- 0.001 34 motivo
+ 0.001 46 chegou	- 0.001 34 Antônia
+ 0.001 44 respeito	- 0.001 32 sabia
+ 0.001 44 comendador	- 0.001 26 indiferente
+ 0.001 43 voz	- 0.001 25 prima
+ 0.001 43 ocasião	- 0.001 21 procurou
+ 0.001 40 vinte	- 0.001 20 segredo
+ 0.001 31 padre	- 0.001 20 chá
+ 0.001 30 notícia	- 0.001 19 jardim
+ 0.001 30 festa	- 0.001 19 chamava
+ 0.001 28 céu	- 0.001 19 beleza
+ 0.001 27 noivo	- 0.001 18 cena
+ 0.001 25 opinião	- 0.001 18 acho
+ 0.001 24 diga	- 0.001 17 vendo
+ 0.001 23 terra	- 0.001 17 presença
+ 0.001 23 longo	- 0.001 17 criado
+ 0.001 23 flor	- 0.001 17 carro
+ 0.001 23 caminho	- 0.001 17 caráter
+ 0.001 21 poderia	- 0.001 16 graças
+ 0.001 21 jovem	- 0.001 16 entrava
+ 0.001 20 véspera	- 0.001 15 visitas
+ 0.001 19 juiz	- 0.001 15 tomou
+ 0.001 18 pediu	- 0.001 14 sobrinha
+ 0.001 18 graça	- 0.001 14 compreendo
+ 0.001 18 formosa	- 0.001 13 só
+ 0.001 16 lábios	- 0.001 13 segunda
+ 0.001 14 orador	- 0.001 13 rindo
+ 0.001 14 estão	- 0.001 13 reparou
+ 0.001 14 digno	- 0.001 13 réis
+ 0.001 13 santo	- 0.001 13 pés
+ 0.001 13 imperador	- 0.001 13 frase
+ 0.001 13 entusiasmo	- 0.001 12 rir
+ 0.001 13 chama	- 0.001 12 orgulho
+ 0.001 12 irmãos	- 0.001 12 doudo

+ 0.001	12	imediatamente	- 0.001	12	chegada
+ 0.001	12	atrás	- 0.001	11	retirou
+ 0.001	11	teria	- 0.001	11	rapazes
+ 0.001	11	novidade	- 0.001	11	instantes
+ 0.001	11	insistiu	- 0.001	11	entrando
+ 0.001	10	café	- 0.001	10	rapariga
+ 0.001	10	achar	- 0.001	9	hábitos
+ 0.001	9	saudades	- 0.001	9	durou
+ 0.001	9	Providência	- 0.001	8	sofrer
+ 0.001	9	necessário	- 0.001	8	sentando
+ 0.001	9	dissesse	- 0.001	8	negócio
+ 0.001	9	confessava	- 0.001	8	mo
+ 0.001	9	andar			
+ 0.001	8	tirava			
+ 0.001	8	doce			
+ 0.001	8	chegaram			
+ 0.001	8	baixinho			
+ 0.001	8	aventurança			

Quadro 24 – Palavras em progressão e em regressão de uso em Flumeia

Palavras em progressão	Palavras em regressão
Coef. / Freq. / Palavra	Coef. / Freq. / Palavra
+ 0.495 94 sei	- 0.596 2725 um
+ 0.493 416 quando	- 0.516 58 etc
+ 0.434 18 contei	- 0.435 138 dois
+ 0.420 55 achei	- 0.426 22 acrescentou
+ 0.416 59 pegou	- 0.398 13 modesto
+ 0.415 11318 .	- 0.388 82 enfim
+ 0.414 98 quis	- 0.375 117 desde
+ 0.408 305 vez	- 0.370 106 coisa
+ 0.405 13 novamente	- 0.367 140 são
+ 0.405 12 soubesse	- 0.366 26 deve
+ 0.396 43 nomes	- 0.366 23 pura
+ 0.394 294 já	- 0.364 16 digno
+ 0.393 348 ia	- 0.362 31 povo
+ 0.391 8 pegava	- 0.358 17 inteira
+ 0.387 33 gestos	- 0.358 10 humanos
+ 0.386 10 tornei	- 0.348 37 casos
+ 0.385 21 senti	- 0.348 16 rara
+ 0.384 598 eu	- 0.345 69 obra
+ 0.383 31 ferro	- 0.342 12 leis
+ 0.381 24 perguntei	- 0.341 46 exemplo
+ 0.379 9 haviam	- 0.341 22 partes
+ 0.378 38 rir	- 0.340 12 portanto
+ 0.377 55 ouvi	- 0.337 9 definitivamente
+ 0.377 20 Nisto	- 0.333 8 capítulos
+ 0.374 88 falar	- 0.332 10 invenção
+ 0.371 173 fez	- 0.331 21 cujo
+ 0.370 10 quiser	- 0.330 10 perfeição
+ 0.369 28 deveras	- 0.329 132 caso
+ 0.366 8 desapareceu	- 0.329 18 qualidades
+ 0.364 82 gesto	- 0.326 175 cinco
+ 0.364 50 pena	- 0.326 39 maneira
+ 0.356 41 sentia	- 0.326 32 cuja
+ 0.355 27 largo	- 0.326 23 teoria
+ 0.351 85 vi	- 0.326 15 cópia
+ 0.343 100 pessoa	- 0.325 10 sistema
+ 0.342 8 barba	- 0.325 8 digna
+ 0.341 1484 para	- 0.324 41 ah
+ 0.340 3740 não	- 0.324 26 aliás
+ 0.339 8 Antônio	- 0.323 2470 do
+ 0.338 16 direção	- 0.323 35 cinquenta
+ 0.334 11 serão	
+ 0.332 14 acabasse	
+ 0.332 11 corda	
+ 0.330 171 mim	
+ 0.327 119 palavras	
+ 0.326 148 aqui	
+ 0.326 17 feita	
+ 0.322 16 rápido	
+ 0.321 61 iam	

Quadro 25 – Palavras em progressão e em regressão de uso em 5VOL

Palavras em progressão			Palavras em regressão		
Coef. / Freq. / Palavra			Coef. / Freq. / Palavra		
+ 0.411	78	deveras	- 0.420	57	fogo
+ 0.368	54	loja	- 0.361	19	anjos
+ 0.358	53	trinta	- 0.360	189	quero
+ 0.355	17	replicou	- 0.337	37	sombra
+ 0.342	46	ombros	- 0.333	37	ideal
+ 0.333	123	cousa	- 0.320	28	solidão
+ 0.333	22	comunicou	- 0.318	260	aquele
+ 0.331	22	obedecer	- 0.313	28	seguir
+ 0.324	2266	ele	- 0.309	53	menor
+ 0.318	408	fez	- 0.300	97	adeus
+ 0.313	45	recusa			
+ 0.307	302	José			
+ 0.306	207	ar			
+ 0.302	19	nuvens			
+ 0.301	58	dous			

Quadro 26 – Palavras em progressão e em regressão de uso em AVG1

Tabela 1 – Indicadores de riqueza lexical e Hápax em Flumeia

n°	réal	théo	Ecart	réduit	Hápax	réduit	Titre
1	2274	2389	-115	-002	521	003	MISS
2	1926	2233	-307	-006	391	-001	LUÍS
3	2992	3638	-646	-011	723	-006	LINHA
4	1026	1093	-67	-002	177	002	FREI
5	2025	2352	-327	-007	442	-000	CONFISSÕES
6	1928	2390	-462	-009	376	-004	SEGREDO
7	2246	2555	-309	-006	457	-002	MULHER
8	858	1086	-228	-007	125	-002	RELÓGIO
9	2091	2330	-239	-005	461	001	ERNESTO
10	2051	2133	-82	-002	504	006	AURORA
11	1785	1872	-87	-002	413	005	DUARTE
12	1242	1630	-388	-010	196	-004	PONTO
13	3388	3524	-136	-002	901	003	PARASITA
Total	10781				5687		

Tabela 2 – Indicadores de riqueza lexical e Hápax em AVG1

n°	réel	théo	écart	réduit	Hápax	réduit	Titre	n°	réel	théo	écart	réduit	Hápax	réduit	Titre
1	1896	2260	-364	-7.66	126	-2.26	ALMAS	39	1940	2488	-548	-10.99	134	-3.37	LOBO
2	1140	1481	-341	-8.86	84	0.78	FOGO	40	3366	3849	-483	-7.79	412	3.19	QUESTÃO
3	2625	3421	-796	-13.61	186	-6.36	RAFAEL	41	2578	2686	-108	-2.08	323	8.67	LEÃO
4	1880	2115	-235	-5.11	166	2.35	DONZELAS	42	2123	2483	-360	-7.22	223	3.39	SILVESTR
5	862	1140	-278	-8.23	103	8.10	PLATINA	43	1502	1871	-369	-8.53	91	-2.16	ETERNA
6	2663	3200	-537	-9.49	256	-0.51	CAMINHO	44	2217	2672	-455	-8.80	221	1.53	POETA
7	2071	2519	-448	-8.93	164	-1.36	MENDONÇA	45	322	760	-438	-15.89	11	-1.40	TESOUROS
8	1458	1825	-367	-8.59	84	-2.45	CARRO	46	745	1120	-375	-11.21	40	-0.85	RECEITA
9	994	1194	-200	-5.79	180	17.81	LATA	47	1279	1773	-494	-11.73	70	-3.38	CASA
10	1428	1712	-284	-6.86	114	1.56	ÁGUIA	48	1685	2038	-353	-7.82	120	-0.93	MARIANA
11	598	1072	-474	-14.48	52	1.60	MISSA	49	2054	2369	-315	-6.47	169	0.27	MILOCA
12	2065	2472	-407	-8.19	169	-0.60	CASES	50	1214	1591	-377	-9.45	85	-0.24	MACHETE
13	1631	2026	-395	-8.78	106	-2.07	CINCO	51	1120	1357	-237	-6.43	104	4.68	SAINETE
14	1460	1783	-323	-7.65	119	1.34	DECADÊNC	52	1151	1590	-439	-11.01	90	0.31	CRISE
15	1241	1570	-329	-8.30	87	0.19	DIANA	53	1488	1963	-475	-10.72	101	-2.00	ESQUELET
16	910	1188	-278	-8.07	158	14.88	ELOGIO	54	202	639	-437	-17.29	16	3.14	MILAGROS
17	2583	3161	-578	-10.28	216	-2.71	ENCHER	55	3341	3527	-186	-3.13	239	-3.97	MÔNICA
18	1720	2133	-413	-8.94	173	2.77	ENTRUDO	56	1605	1874	-269	-6.21	159	4.22	VIRGINIU
19	2170	2569	-399	-7.87	225	2.74	CASAMENT	57	1224	1428	-204	-5.40	102	3.50	AYRES
20	1733	2011	-278	-6.20	124	-0.35	FERNANDO	58	1535	1818	-283	-6.64	134	2.43	CASADA
21	754	1161	-407	-11.94	98	6.97	FILOSOFI	59	1698	2096	-398	-8.69	157	1.76	AVARO
22	1520	1868	-348	-8.05	97	-1.57	FRANCISC	60	2200	2464	-264	-5.32	245	5.25	DÍVIDA
23	1319	1705	-386	-9.35	83	-1.51	LÁGRIMA	61	1469	1974	-505	-11.37	97	-2.45	FITA
24	1368	1852	-484	-11.25	85	-2.57	LONGE	62	2751	3521	-770	-12.98	238	-3.99	OUTRA
25	1688	2097	-409	-8.93	117	-1.70	NOIVAS	63	1527	1990	-463	-10.38	90	-3.21	ONZE
26	1851	2199	-348	-7.42	138	-0.77	ANOS	64	1282	1783	-501	-11.86	75	-2.98	CONTA
27	1643	2132	-489	-10.59	98	-3.60	ASNO	65	1508	1893	-385	-8.85	110	-0.57	TOBE
28	778	1148	-370	-10.92	33	-2.18	ORÁCULO	66	1483	1766	-283	-6.73	87	-1.65	SUPERIOR
29	2316	2882	-566	-10.54	191	-2.26	OPAI	67	956	1357	-401	-10.89	53	-1.61	HERANÇA
30	1101	1505	-404	-10.41	49	-3.42	PASSADO	68	1228	1609	-381	-9.50	65	-2.54	COSME
31	1773	2016	-243	-5.41	153	2.18	ONDA	69	954	1323	-369	-10.14	51	-1.53	FOLHA
32	1734	2123	-389	-8.44	145	0.48	ANTÃO	70	1141	1482	-341	-8.86	111	3.86	ASTRÓLOG
33	1759	2200	-441	-9.40	110	-3.10	LOUREIRA	71	1741	1910	-169	-3.87	26	-8.51	QUIMERAS
34	2135	2596	-461	-9.05	177	-1.02	PIANISTA	72	1845	2199	-354	-7.55	179	2.63	OLHOS
35	1565	2042	-477	-10.56	123	-0.71	MOÇAS	73	1606	2009	-403	-8.99	156	2.52	ALMOÇO
36	1513	1874	-361	-8.34	92	-2.09	CAIPORAS	74	1884	2276	-392	-8.22	177	1.74	AMBICIOS
37	1870	2252	-382	-8.05	120	-2.68	POSSÍVEL	75	2333	2594	-261	-5.12	215	1.78	VALÉRIO
38	3204	3974	-770	-12.21	295	-4.04	QUAL	76	1879	2263	-384	-8.07	219	5.27	CHAVE
Tot 23591								10291							

Tabela 3 – Indicadores de riqueza lexical e Hápax em 5VOL

n°	réal	théo	écart	réduit	Hápax	réduit	Titre
1	1037	1202	-165	-4.76	152	3.34	ALCIBÍAD
2	1068	1237	-169	-4.81	165	4.04	SEGREDO
3	1197	1416	-219	-5.82	164	1.52	ESPELHO
4	1065	1269	-204	-5.73	131	0.51	EMPRESTI
5	1069	1272	-203	-5.69	146	1.83	POLÍCRAT
6	4096	4417	-321	-4.83	871	5.21	ALIENIST
7	698	1092	-394	-11.92	115	1.34	ARCA
8	2003	2600	-597	-11.71	258	-4.49	BENEDITA
9	1181	1263	-82	-2.31	243	10.70	SERENÍSS
10	1372	1476	-104	-2.71	184	2.40	VERBA
11	1289	1494	-205	-5.30	164	0.57	CHINELA
12	1257	1305	-48	-1.33	244	10.03	MEDALHÃO
13	1225	1361	-136	-3.69	156	1.54	CAPITULO
14	936	1239	-303	-8.61	65	-5.14	SINGULAR
15	1212	1393	-181	-4.85	153	0.89	SAPUCAIA
16	1165	1355	-190	-5.16	150	1.10	LAPSO
17	872	1131	-259	-7.70	59	-4.65	ALMIRANT
18	1059	1226	-167	-4.77	138	1.71	MANUSCRI
19	1208	1318	-110	-3.03	149	1.48	GALERIA
20	959	1073	-114	-3.48	104	0.50	FULANO
21	1057	1280	-223	-6.23	107	-1.78	CATHEDRA
22	1193	1308	-115	-3.18	201	6.20	ALEXANDR
23	1561	1894	-333	-7.65	208	-0.67	CHAPEUS
24	613	798	-185	-6.55	40	-3.07	ESPONSAI
25	1063	1200	-137	-3.95	146	2.81	ACADEMIA
26	1177	1426	-249	-6.59	124	-1.93	PECUNIAR
27	1013	1226	-213	-6.08	111	-0.78	GALVAO
28	1093	1287	-194	-5.41	131	0.29	SEGUNDA
29	1125	1217	-92	-2.64	163	4.14	IGREJA
30	1063	1243	-180	-5.11	131	0.84	SENHORA
31	1137	1434	-297	-7.84	142	-0.52	BRACOS
32	1220	1468	-248	-6.47	147	-0.51	CÉLEBRE
33	969	1167	-198	-5.80	104	-0.71	TRIO
34	1183	1397	-214	-5.73	131	-1.02	ENFERMEI
35	1236	1486	-250	-6.49	134	-1.76	DIPLOMAT
36	822	957	-135	-4.36	150	7.20	CONEGO
37	1103	1349	-246	-6.70	118	-1.59	MARIANA
38	1093	1316	-223	-6.15	137	0.46	SANTOS
39	1108	1411	-303	-8.07	88	-4.78	PAULA
40	1012	1316	-304	-8.38	98	-2.96	ESCOLA
41	752	971	-219	-7.03	103	1.84	ADÃO
42	1130	1372	-242	-6.53	107	-2.78	DESEJADA
43	1229	1461	-232	-6.07	126	-2.14	SECRETA
44	890	1195	-305	-8.82	155	3.72	VIVER
45	1134	1411	-277	-7.37	123	-1.85	CARTOMAN
46	297	501	-204	-9.11	25	-0.91	APOLOGO
47	1036	1204	-168	-4.84	117	0.05	PAPÉIS
48	522	650	-128	-5.02	64	2.48	DICIONAR
49	909	1163	-254	-7.45	86	-2.39	VARA
50	942	1187	-245	-7.11	70	-4.17	GALO
51	804	1057	-253	-7.78	100	0.30	XERXES
52	741	875	-134	-4.53	77	0.34	CANÁRIO
53	1915	2334	-419	-8.67	232	-3.53	ERRADIO
54	1348	1571	-223	-5.63	172	0.29	ETERNO
55	971	1202	-231	-6.66	107	-0.85	FERIAS
56	1455	1803	-348	-8.20	136	-4.78	VOLUNTÁR
57	672	862	-190	-6.47	61	-1.35	GORDO
58	1179	1452	-273	-7.16	96	-4.51	PÍLADES
59	1328	1604	-276	-6.89	133	-3.08	PAI
60	1001	1181	-180	-5.24	96	-1.65	MARCHA
61	871	1155	-284	-8.36	74	-3.45	CABRIOLE
62	865	1008	-143	-4.50	110	2.03	EVOLUÇÃO
63	1787	2346	-559	-11.54	223	-4.17	CORA
Tot	18703				9015		

Tabela 4 – Indicadores de riqueza lexical e Hápax em 5VOL

n°	réel	théo	écart	réduit	Hápax	réduit	Titre
1	552	704	-152	-5.73	60	1.14	TARPÉIA
2	1731	1962	-231	-5.22	292	3.80	ASTÚCIAS
3	598	770	-172	-6.20	77	2.24	INVENTAR
4	1310	1486	-176	-4.57	164	0.53	BARRETO
5	1674	2139	-465	-10.05	185	-5.05	VIÚVA
6	1250	1461	-211	-5.52	168	1.17	ESCRIVÃO
7	1348	1564	-216	-5.46	162	-0.57	INGLESIN
8	662	789	-127	-4.52	83	2.66	JOÃO
9	774	963	-189	-6.09	65	-2.09	REMÉDIO
10	397	622	-225	-9.02	76	5.22	METAFISI
11	1602	1843	-241	-5.61	213	-0.10	UMA
12	1447	1718	-271	-6.54	182	-0.86	MULHER
13	405	557	-152	-6.44	36	0.22	ORAI
14	1007	1194	-187	-5.41	135	1.88	CARDEAL
15	1028	1262	-234	-6.59	126	0.13	TERPÍCO
16	1579	1857	-278	-6.45	173	-3.02	QUARTO
17	1377	1639	-262	-6.47	164	-1.29	SONHO
18	1161	1451	-290	-7.61	130	-1.81	TRINA
19	1092	1189	-97	-2.81	182	6.38	EZEQUIEL
20	942	1223	-281	-8.04	78	-3.79	SOBRAL
21	4014	4860	-846	-12.14	839	-3.30	CASA
22	1306	1443	-137	-3.61	165	1.16	JUCUNDA
23	687	848	-161	-5.53	58	-1.37	FLOR
24	865	1012	-147	-4.62	122	3.34	HABILIDO
25	1286	1444	-158	-4.16	235	6.90	IDENTIDA
26	1076	1264	-188	-5.29	130	0.47	BICHO
27	1611	2043	-432	-9.56	181	-4.37	ROMUALDO
28	540	690	-150	-5.71	56	0.83	CONTRATO
29	1127	1176	-49	-1.43	187	7.07	SALES
30	1142	1232	-90	-2.56	154	3.10	SÓ
31	433	570	-137	-5.74	38	0.30	DISTICO
32	1533	1737	-204	-4.89	190	-0.50	NOITE
33	2081	2532	-451	-8.96	281	-3.31	OUTRA
34	1173	1400	-227	-6.07	112	-2.73	DIVINA
35	1219	1373	-154	-4.16	171	2.59	VIAGEM
36	661	836	-175	-6.05	68	0.01	VIDROS
37	546	733	-187	-6.91	39	-2.20	CARTEIRA
38	450	608	-158	-6.41	23	-2.77	CURTA
39	1252	1550	-298	-7.57	131	-2.83	FINOCA
40	719	871	-152	-5.15	60	-1.45	CARTA
41	2189	2383	-194	-3.97	433	7.02	IMORTAL
42	407	681	-274	-10.50	19	-4.31	REMÉDIO
43	2174	2337	-163	-3.37	358	3.24	PROGRAMA
44	593	813	-220	-7.72	55	-1.25	MARIDOS
45	569	748	-179	-6.54	51	-0.81	TRÊS
46	1311	1551	-240	-6.09	174	0.52	TROCA
47	644	782	-138	-4.93	62	0.09	INCENDIO
48	641	797	-156	-5.53	66	0.36	VINTE
49	556	717	-161	-6.01	54	0.08	DESTINAD
50	477	646	-169	-6.65	39	-0.89	HISTÓRIA
51	1255	1466	-211	-5.51	138	-1.33	LETRA
52	1048	1326	-278	-7.63	102	-2.77	CANTIGA
53	1092	1332	-240	-6.58	117	-1.53	ENTRE
Tot	15605				7659		

Tabela 5 – Indicadores de riqueza lexical e Hápax em 7VOL

n°	réel	théo	écart	réduit	Hápax	réduit	Titre	n°	réel	théo	écart	réduit	Hápax	réduit	Titre
1	2274	2632	-358	-007	254	-001	MISS	39	1181	1455	-274	-007	115	-000	ANEDOTA
2	1926	2453	-527	-011	206	-003	LUÍS	40	1016	1267	-251	-007	97	000	GALVÃO
3	2992	4089	-1097	-017	388	-006	LINHA	41	1097	1324	-227	-006	114	001	SEGUNDA
4	1026	1216	-190	-005	80	-001	FREI	42	1129	1259	-130	-004	138	005	IGREJA
5	2025	2589	-564	-011	240	-002	CONFISSÕ	43	1067	1283	-216	-006	111	001	SENHORA
6	1928	2633	-705	-014	174	-006	SEGREDO	44	1142	1463	-321	-008	119	000	BRAÇOS
7	2246	2823	-577	-011	240	-004	MULHER	45	1225	1496	-271	-007	119	-000	CÉLEBRE
8	858	1209	-351	-010	52	-004	RELÓGIO	46	972	1211	-239	-007	84	-001	TRIO
9	2091	2564	-473	-009	225	-003	ERNESTO	47	1187	1428	-241	-006	107	-001	ENFERMEI
10	2051	2341	-290	-006	258	002	AURORA	48	1241	1512	-271	-007	120	-000	DIPLOMAT
11	1785	2048	-263	-006	230	003	DUARTE	49	824	1015	-191	-006	130	008	CÔNEGO
12	1242	1781	-539	-013	104	-004	PONTO	50	1107	1383	-276	-007	92	-002	MARIANA
13	3388	3955	-567	-009	490	-001	PARASITA	51	1097	1351	-254	-007	126	002	SANTOS
14	1041	1244	-203	-006	135	004	ALCIBÍAD	52	1112	1441	-329	-009	74	-004	PAULA
15	1072	1277	-205	-006	132	004	BONZO	53	1016	1351	-335	-009	81	-002	ESCOLA
16	1201	1446	-245	-006	132	001	ESPELHO	54	754	1028	-274	-009	94	003	ADÃO
17	1068	1307	-239	-007	116	002	EMPRÉSTI	55	1134	1405	-271	-007	95	-002	DESEJADA
18	1072	1309	-237	-007	123	002	POLÍCRAT	56	1234	1488	-254	-007	97	-002	SECRETA
19	4121	4377	-256	-004	751	007	ALIENIST	57	893	1238	-345	-010	137	005	VIVER
20	701	1141	-440	-013	94	001	ARCA	58	1138	1442	-304	-008	110	-001	CARTOMAN
21	2014	2589	-575	-011	214	-003	BENEDITA	59	298	599	-301	-012	19	-001	APÓLOGO
22	1185	1302	-117	-003	218	012	SERENÍSS	60	1039	1246	-207	-006	108	002	PAPÉIS
23	1376	1503	-127	-003	156	003	VERBA	61	523	734	-211	-008	56	003	DICIONÁR
24	1294	1520	-226	-006	120	-000	CHINELA	62	912	1208	-296	-009	70	-002	CASO
25	1261	1341	-80	-002	217	011	MEDALHÃO	63	945	1230	-285	-008	58	-003	MISSA
26	1229	1394	-165	-004	133	002	ÚLTIMO	64	807	1109	-302	-009	89	001	LÁGRIMAS
27	939	1279	-340	-010	56	-004	SINGULAR	65	743	940	-197	-006	67	001	CANÁRIO
28	1216	1424	-208	-006	131	002	SAPUCAIA	66	1925	2329	-404	-008	205	-002	ERRADIO
29	1169	1388	-219	-006	136	003	LAPSO	67	1353	1593	-240	-006	149	001	ETERNO
30	875	1178	-303	-009	51	-004	ALMIRANT	68	974	1244	-270	-008	91	-000	FÉRIAS
31	1063	1267	-204	-006	127	003	MANUSCRI	69	1462	1815	-353	-008	119	-003	VOLUNTÁR
32	1212	1353	-141	-004	124	002	GALERIA	70	674	929	-255	-008	53	-001	GORDO
33	962	1123	-161	-005	82	000	FULANO	71	1184	1480	-296	-008	85	-003	PÍLADES
34	1061	1318	-257	-007	90	-001	CATHEDRA	72	1334	1625	-291	-007	110	-002	PAI
35	1197	1344	-147	-004	174	007	CONTO	73	1004	1225	-221	-006	85	-001	MARCHA
36	1568	1903	-335	-008	176	000	CAPÍTULO	74	875	1201	-326	-009	61	-003	CABRIOLE
37	615	869	-254	-009	29	-003	ESPONSAI	75	867	1063	-196	-006	94	003	EVOLUÇÃO
38	1067	1243	-176	-005	123	003	ACADEMIA	76	1797	2341	-544	-011	179	-004	CORA
Tot 22678								10639							

Tabela 6 – Indicadores de riqueza lexical e Hápax em Portext

n°	réal	théo	écart	réduit	Hápax	réduit	Titre	n°	réal	théo	écart	réduit	Hápax	réduit	Titre
1	14255	17834	-3579	-26.80	2169	31.48	GrMA	39	4582	6818	-2236	-27.08	215	0.45	FALE
2	16800	27639	-10839	-65.20	3256	18.42	SeAV	40	4085	6564	-2479	-30.60	128	-4.64	ALIE
3	7925	10997	-3072	-29.29	524	0.80	Cara	41	9630	12456	-2826	-25.32	480	-5.79	Cabe
4	2800	5135	-2335	-32.58	81	-2.19	Urag	42	8710	12228	-3518	-31.81	547	-2.36	INOC
5	3578	6518	-2940	-36.42	135	-3.95	PoEs	43	10398	13699	-3301	-28.20	866	5.09	MiFo
6	5760	8505	-2745	-29.76	321	-0.06	CaCh	44	8426	10984	-2558	-24.41	390	-5.15	ERMI
7	3576	6013	-2437	-31.43	145	-0.97	MeRJ	45	14271	13483	788	6.79	1174	17.52	ATEN
8	3502	6305	-2803	-35.30	193	1.31	Macá	46	13115	18013	-4898	-36.49	787	-10.74	Mula
9	3857	6819	-2962	-35.87	206	-0.18	NoTa	47	11119	17496	-6377	-48.21	468	-19.13	GiAm
10	3911	6163	-2252	-28.69	182	1.20	PoAA	48	9069	12551	-3482	-31.08	521	-4.43	FiBo
11	5130	7467	-2337	-27.04	235	-1.03	FaVa	49	12301	17776	-5475	-41.06	734	-11.71	CaPe
12	6954	11617	-4663	-43.26	305	-10.69	MORE	50	11196	15984	-4788	-37.87	641	-9.87	CORT
13	10767	20000	-9233	-65.29	480	-24.34	MoLo	51	8167	12204	-4037	-36.54	409	-7.95	LISO
14	8553	12516	-3963	-35.42	378	-10.07	LuMá	52	5101	8093	-2992	-33.26	162	-7.71	MoAl
15	2353	5190	-2837	-39.38	83	-2.28	DiMA	53	2624	5533	-2909	-39.11	88	-3.50	Escr
16	5559	8376	-2817	-30.78	242	-4.03	LaRi	54	3892	8347	-4455	-48.76	216	-5.41	CaFe
17	6902	9586	-2684	-27.41	542	7.17	SuPo	55	9842	13484	-3642	-31.36	455	-9.68	NORM
18	7779	13124	-5345	-46.66	271	-15.81	SaMi	56	1734	4571	-2837	-41.96	33	-4.39	CoVe
19	3899	6672	-2773	-33.95	55	-10.24	Aviu	57	9783	12334	-2551	-22.97	605	-0.34	MORT
20	3257	6157	-2900	-36.96	62	-8.10	CMJA	58	8812	12786	-3974	-35.14	425	-8.93	INTR
21	6456	9340	-2884	-29.84	268	-5.84	DIVA	59	11046	13581	-2535	-21.75	982	9.88	MoLi
22	5470	8654	-3184	-34.23	317	-0.86	IRAC	60	8189	8661	-472	-5.07	868	29.43	MaHi
23	6200	9571	-3371	-34.46	150	-12.50	PATA	61	10902	10631	271	2.63	1077	27.57	CoCo
24	7731	11221	-3490	-32.95	264	-11.42	Lucí	62	8802	10350	-1548	-15.22	1546	51.28	PeBo
25	11463	15326	-3863	-31.20	649	-7.80	SENH	63	22682	23741	-1059	-6.87	3014	28.43	Sert
26	5495	8720	-3225	-34.54	124	-11.66	Enca	64	11873	13312	-1439	-12.47	1141	17.01	ObS6
27	5102	7374	-2272	-26.46	274	1.86	CAAI	65	13702	14900	-1198	-9.81	1548	24.78	ObS7
28	7456	12806	-5350	-47.28	202	-17.79	CoFl	66	17146	17462	-316	-2.39	2249	35.66	ObS8
29	5617	9594	-3977	-40.60	141	-13.02	LUVA	67	2572	6347	-3775	-47.38	76	-7.68	Novi
30	9229	14837	-5608	-46.04	347	-16.98	JACÓ	68	6844	8827	-1983	-21.11	536	10.34	PoIn
31	7811	12978	-5167	-45.36	184	-18.86	IAIÁ	69	10472	13135	-2663	-23.24	610	-2.74	ISAÍ
32	9816	13573	-3757	-32.25	489	-8.64	BRÁS	70	10753	14138	-3385	-28.47	591	-6.47	PoQu
33	8629	13904	-5275	-44.74	322	-15.71	DOMC	71	8640	10531	-1891	-18.43	682	9.83	BRUZ
34	6310	11919	-5609	-51.38	153	-17.84	AIRE	72	6906	8808	-1902	-20.27	356	0.68	NoCA
35	4577	6756	-2179	-26.51	201	-0.25	AMER	73	8591	12095	-3504	-31.86	397	-8.13	ClAn
36	3989	7646	-3657	-41.82	82	-11.20	CaVe	74	8107	11570	-3463	-32.19	431	-5.16	CeVi
37	3904	6132	-2228	-28.45	129	-2.79	CRIS	75	11677	13343	-1666	-14.42	1045	13.21	DiIn
38	2973	5059	-2086	-29.33	117	1.85	Ocid	76	6473	9492	-3019	-30.99	997	30.65	SimõesLNeto
Tot 106574								41768							

Ficha Catalográfica

Freitas, Deise J. T. de

A composição do estilo do contista Machado de Assis. Deise J. T. de Freitas ; orientador, Alckmar Luiz dos Santos. – Florianópolis, 2007.

211 f.

Inclui bibliografia

Tese (doutorado) – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Literatura, 2007.

1. Machado de Assis. 2. Conto. 3. Estatística Textual. 4. Estilo Informática. I. Santos, Alckmar Luiz dos. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Literatura. III. Título.