TEXTO DIGITAL

Revista de Literatura, Linguística e Artes

Métricas Rosianas (1929-1967)

Rosian Metrics (1929-1967)

Verônica Ribas Cúrcio^a

^a Universidade Federal de Santa Catarina, Brasil - demodoco@gmail.com

Palavras-chave:
Guimarães Rosa.
Estatística Textual.
Estilometria.

Resumo: Neste artigo vamos apresentar brevemente alguns resultados oriundos de um estudo de estatística textual aplicado sobre a obra literária completa de João Guimarães Rosa, sabemos que há as epístolas de Rosa, mas essas não foram utilizadas por não possuírem caráter originalmente literário. Grosso modo, vale dizer que dois campos foram amplamente explorados: o primeiro campo retrata as características gerais e específicas do próprio vocabulário empregado por Guimarães Rosa e o segundo compara os resultados obtidos com três teses da crítica literária rosiana que retratam questões intrigantes sobre o vocabulário de Rosa e suas manifestações ao longo da vida produtiva literária do autor mineiro.

Keywords:
Guimarães Rosa.
Textual Statistics.
Stylometry.

Abstract: In this article we will briefly present some results from a study of textual statistics applied to the complete literary work of João Guimarães Rosa, we know that there are the epistles of Rosa, but these were not used because they did not have an original literary character. Grosso modo, two fields have been extensively explored: the first field portrays the general and specific characteristics of Guimarães Rosa's own vocabulary, and the second field compares the results to three theses of Rosa's literary criticism that show intriguing questions about the vocabulary of Rosa and its manifestations throughout the productive literary life of the author.

Submeter a obra literária completa de João Guimarães Rosa em um *software* de análise estatística de textos literários não foi uma tarefa muito fácil, mas com a ajuda de alguns parceiros, conseguimos ir em busca de textos mais longínquos como os primeiros contos publicados por Rosa, e assim, pudemos compor o *corpus* que apresentamos a seguir, datados com a primeira publicação: *O mistério de Highmore Hall* (1929); *Makiné* (1930); *Tempo e destino* (1930); *Caçadores de camurças* (1930); *Magma* (1936); *Sagarana* (1946); *Corpo de baile* (1956); *Grande sertão: veredas* (1956); *O último dos maçaricos* (1958)¹; *O mistério dos MMM* (1961); *Primeiras histórias* (1962); *Tutameia* (1967); *Estas histórias* (1969) e *Ave, palavra* (1970)².

Reunimos de modo abrangente o léxico (1.094.481 ocorrências, 58.647 vocábulos) de Guimarães Rosa dentro de um percurso cronológico de sua produção literária, utilizando a ferramenta informatizada chamada *Hyperbase* (desenvolvida pelo laboratório BCL da Universidade de Nice - França), e por meio dela, fizemos nossas análises sobre os dados estatísticos dos textos.

Dos resultados, podemos destacar que mais da metade do vocabulário (52,81%) não se repete³, o que demonstra bastante riqueza e diversificação. Os textos mais curtos representaram maior riqueza lexical, os textos com mais diálogos apresentaram menor diversificação. Concluimos que a maioria das obras apresenta mais déficit de vocabulário que excedente, principalmente as que retratam a temática do sertão (*Sagarana*, *Grande sertão: veredas* e *Corpo de baile*), conforme o gráfico a seguir.

1 Tradução condensada do inglês, da obra *The Last of the Curlews* (1954), de Fred Bodsworth.

2 As datas aqui apresentadas são as de primeira publicação, contudo, algumas dessas obras são coletâneas de textos que foram publicados em datas anteriores.

3 Por conta das análises de tempos verbais, optamos por não lematizar o corpus.

Fig. 1 – Riqueza lexical da obra de Guimarães Rosa

Fonte: Hyperbase, 2011.

Os resultados dos *hapaxes* também nos levaram à mesma conclusão: *Ave, palavra* (no gráfico, a seguir, está sob as datas de 1965 a 1967, ao final da linha do tempo) surge como a obra mais diversificada. Das análises dos resultados, vimos que a riqueza lexical é avaliada a partir de critérios como gênero literário, estilo e temática. Contudo, no caso de Rosa o tema não parece contribuir com a diversificação do léxico. Sabemos que o tema sertanejo é bastante presente nas obras *Tutaméia*, *Estas estórias* e *Primeiras estórias*, porém, a ênfase no vocabulário é dada nas três obras destacadas em azul no gráfico (*Sagarana*, *Grande sertão: veredas* e *Corpo de baile*). Em resumo, podemos afirmar por meio das análises de crescimento de vocabulário que, em Rosa, a riqueza vocabular não se apresenta pelo volume da obra, mas pela temática.

Fig. 2 – Hapax da obra de Guimarães Rosa

Fonte: Hyperbase, 2011.

PONTUAÇÃO E VERBO

A pontuação rosiana também foi considerada nesse estudo. Por meio da ferramenta que trata da evolução do vocabulário, percebemos que, ao longo de sua produção, Rosa empregou mais a vírgula e o ponto-final, enquanto houve abandono da exclamação. A interrogação apresentou tendências deficitárias no início e no fim da linha cronológica, mas teve seu auge positivo nas obras *Grande sertão: veredas* e *Corpo de baile*.

Das análises sobre os verbos, destacamos que os tempos verbais mais utilizados foram: o subjuntivo do presente e o futuro do pretérito. Por outro lado, o pretérito imperfeito foi menos empregado no final da produção.

ZONA DE ISOLAMENTO LEXICAL

Sobre os gráficos das análises fatoriais constatamos que as obras mais distantes da linha que contorna a característica rosiana foram as de gêneros distintos: os contos de um escritor principiante (1929-1930), o capítulo de um romance policial (*MMM* - 1958) desenvolvido em parceria com outros escritores e a tradução condensada de um romance canadense (1960-

1961). Vale observar que, *Ave*, *palavra* se localiza ao meio da zona de isolamento lexical, fazendo uma ponte para os diferentes gêneros textuais.

Fig. 3 – Análise fatorial da obra de Guimarães Rosa

Fonte: Hyperbase, 2011.

EVOLUÇÃO DO VOCABULÁRIO

Dos resultados sobre os coeficientes de evolução do vocabulário, detectamos, nas altas frequências, como nos cálculos de distância lexical, mais vocábulos que se direcionam para a temática do ser e suas preocupações existenciais, tais como o *tempo*, o *amor* e a *alma*. Do mesmo modo, redefinimos a importância das palavras gramaticais, pois além das mesmas cumprirem com o funcionamento da lógica do discurso, elas também são possuidoras de carga semântica. Encontramos alguns advérbios e pronomes que identificaram dois campos semânticos: *ser* e *tempo*.

Além da garimpagem dos dados e das suas descrições sobre as características gerais e específicas do vocabulário rosiano descritas brevemente acima, buscamos confirmar e complementar intuições oriundas de leituras convencionais da crítica de Rosa. Para isso, apoiamo-nos em Sperber (1982), Covizzi (1978) e Daniel (1968).

Suzi Sperber (1982), em *Guimarães Rosa: signo e sentimento*, estuda a produção de Rosa no que diz respeito à organização da linguagem sob a luz da *Textkritik*. Nesse método, ela

verificou o comportamento do vocabulário e o amadurecimento da escrita do autor, estudando desde o rascunho à versão definitiva. Das confirmações da autora, buscamos trabalhar com aquelas que viabilizariam análises sobre o crescimento e evolução de vocabulário. Vimos com Sperber (1976), que no processo de estruturação da obra rosiana, o estilo de Rosa se sobressai ao tema. Isso foi encontrado, quando analisamos o resultado sobre a variação de vocabulário: o gráfico (Fig. 1) mostrou déficit nas obras que retratam o sertão. Nas outras obras (as não-rurais) houve uma forte diversificação do léxico. Concluímos que a temática em Rosa é um elemento de estilo, contudo ela não contribui para a diversificação do léxico.

Outro estudo que escolhemos para guiar nossas análises foi a perspectiva de Lenira Covizzi⁴ (1978), cuja tese é delinear um percurso da narrativa rosiana que vai da expressão (arte em si) à explicação (comentário, autocrítica). Para a autora, *Grande sertão: veredas* e *Corpo de baile* abordam universos bem elaborados do ponto de vista da linguagem e do tema e que resultaram da densidade que foi a produção de *Sagarana*. A partir dos textos de 1956, ocorre uma diluição do tema e em contrapartida a produção literária ganha um aumento considerável em liberdade da linguagem. Segundo Covizzi (1978), é em *Primeiras estórias* que se inicia o processo explicativo da obra.

A respeito do percurso narrativo de Rosa, a autora conclui que a partir dos textos de 1956 há uma redução do teor regional como em *Primeiras Estórias*. Para analisar essas afirmações, fizemos um levantamento de vocabulário pelas ocorrências de *hapax*, e desse modo verificamos o comportamento de entrada de novos *hapaxes*, caracterizando o processo de criação e renovação de vocabulário.

Pelo levantamento de vocabulário por ocorrências de *hapax*, detectamos que a fase expressiva de Rosa, a qual teria o seu cume em *Grande sertão: veredas*, não ocorreu em termos lexicais, e que a fase explicativa se utilizou de uma diversidade lexical muito maior que a fase expressiva. Por meio dos indícios verbais extraídos dos coeficientes positivos e negativos sobre a evolução do vocabulário, chegamos em acordo com Covizzi (1978), pois ao final da carreira literária, Rosa compôs mais locuções discursivas e comentários.

⁴ Vale lembrar que Covizzi (1978) considerou para a sua análise duas categorias “expressão” e “explicação”. Em nossa tese, retomamos essas categorias para prosseguir com o trabalho da pesquisadora, porém, sob outro viés e alcançando a obra completa de Guimarães Rosa (pois a autora não abarcou toda a produção).

Outro estudo que consideramos importante para as análises estilométricas foi o de Mary Lou Daniel (1968), pois suas afirmações indicam uma ruptura no léxico rosiano em duas fases, rural (com *Sagarana*, *Corpo de baile* e *Grande sertão: veredas*) e não-rural (*Primeiras estórias*). A autora estabelece uma progressão qualitativa (de obra a obra) que demonstra o movimento do potencial comunicativo, e indica *Primeiras estórias* como a obra madura do autor (ressaltamos também que este estudo de Daniel (1968) é contemporâneo à obra de Rosa, portanto, não abrangeu a obra completa do autor).

Como exercício quantitativo das análises de Daniel (1968), a premissa foi verificar como essa distinção entre rural e não-rural (urbano) se reflete no léxico. Para isso, utilizamos a ferramenta que mede a evolução e outra que aborda a distância lexical. Essa última, considera o vocabulário completo que compõe cada texto do *corpus*, sem se preocupar com a frequência dos vocábulos, mas apenas com a presença ou ausência de uma palavra. A distância lexical também verifica a conexão entre textos por meio de seus vocabulários, que podem ser comuns (permitindo a aproximação) ou não (a distância). Dos resultados obtidos, verificamos que há dois momentos de ruptura no material linguístico de Rosa, apresentando uma grande diferença de aparecimento de *hapaxes* e riqueza lexical, ou seja, de diversidade vocabular. As obras de maior fôlego (*Sagarana*, *Grande sertão: veredas* e *Corpo de baile*) apresentaram um vocabulário mais restrito, enquanto que as outras obras mostraram um vocabulário mais variado.

Muitos dados surgiram com a utilização do *Hyperbase* aplicado aos textos de Guimarães Rosa, o que trouxemos aqui é a ínfima parte do que foi estudado. Observamos que em muitos momentos a crítica rosiana “acertou” intuitivamente, foi o que o quantitativo pôde evidenciar. A estatística textual apenas soma e complementa a leitura antes feita. Outra condição que não deve ser ignorada é a possibilidade de ampliar as hipóteses de leitura de uma ou meia dúzia de obras para a obra completa, ou seja, verificar se alguma tendência que foi de maneira intuitiva detectada por meio da leitura convencional se aplica ou não à obra completa de um escritor.

Em suma, a estilometria possibilita trabalhar com uma grande massa de textos, que aliada ainda à capacidade de memória dos computadores atuais e à potencialidade das ferramentas tecnológicas para análise de textos, traduz muitos resultados. As ferramentas disponíveis criam possibilidades e potencializam a interpretação sobre a obra literária, cabe ao estudioso

elaborar suas correlações e aproveitar as variadas formas de ser ler um texto mediado pela estatística.

REFERÊNCIAS

BRUNET, Étienne. **Hyperbase** ©. Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels. Manuel de référence. Version 8.0 et 9.0. Nice, França: Université Nice Sophia Antipolis, 2011.

COVIZZI, Lenira Marques. **O insólito em Guimarães Rosa e Borges**. São Paulo: Ática, 1978.

DANIEL, Mary Lou. **João Guimarães Rosa**: Travessia literária. Rio de Janeiro: José Olympio, 1968.

ROSA, João Guimarães. **Ave, palavra**. Rio de Janeiro: José Olympio, 1970.

_____. **Carta a Harriet de Onís**: 4 nov. 1964. [S. l.]: [s. n.], 19-?.

_____. **Corpo de baile**. Rio de Janeiro: José Olympio, 1976a.

_____. **Estas estórias**. Rio de Janeiro: José Olympio, 1976b.

_____. **Ficção completa**. v. 1. Rio de Janeiro: Nova Aguilar, 1995a.

_____. **Ficção completa**. v. 2. Rio de Janeiro: Nova Aguilar, 1995b.

_____. **Grande Sertão**: Veredas. Rio de Janeiro: José Olympio, 1974.

_____. **Magma**. Rio de Janeiro: Nova Fronteira, 1997.

_____. **Primeiras estórias**. 12. ed. Rio de Janeiro: José Olympio, 1981.

_____. **Primeiras estórias**. Rio de Janeiro: José Olympio, 1974.

_____. **Sagarana**. Rio de Janeiro: Nova Fronteira, 1993.

_____. **Sagarana**. Rio de Janeiro: José Olympio, 1980.

_____. **Tutaméia**. Rio de Janeiro: José Olympio, 1967.

_____. **No Urubuquaquá no Pinhém**. Rio de Janeiro: Nova Fronteira, 1984.

SPERBER, Suzi Frankl. **Guimarães Rosa**: signo e sentimento. São Paulo: Ática, 1982.