

MEMÓRIA DESCRITIVA

VISÃO PANORÂMICA

Depois de, num primeiro momento, analisados aspetos relativos aos dez termos mais recorrentes na obra poética de António Ramos Rosa, nomeadamente em frequências absolutas e em distribuição das frequências (ocorrências mínimas e totais por publicação), procedeu-se à análise desses mesmos dez termos em contexto de pares de palavras. Com a ajuda do modelo analítico N-grama (em inglês, *N-gram*), capaz de buscar palavras sequenciais em uma determinada amostra textual, todo o corpus sem palavras de paragem (*stop words*) e devidamente lematizado foi processado de modo a dispor os pares de palavras (bigramas, ou *bi-grams*) enquanto dados pontos para posterior análise.

Através da utilização do pacote “quanteda”, uma ferramenta de análise quantitativa disponível para Linguagem R, procedeu-se à manipulação do corpus de forma que cada palavra do conjunto já processado fosse reconhecido como um *token*, ou seja, como a unidade textual mais atômica possível. A partir desse passo de tokenização, o corpus foi reprocessado utilizando-se como modelo a divisão em bigramas, isto é, cada *token* passou a ser constituído não apenas por uma palavra, mas por duas: um bigrama. Nessa altura, pode-se analisar quais os bigramas mais frequentes em todo o corpus em frequência absoluta (Visão Panorâmica - Análise 1), constatando-se a forte presença de algumas das dez palavras mais frequentes (em cor) em todo o corpus no contexto dos pares de palavras mais frequentes em todo corpus. Cabe ressaltar que, nessa análise, admitiu-se a presença de pares de palavras repetidas, como “corpo” seguido de “corpo”, e “ser” seguido de “ser”, entre outros observáveis no gráfico.

Observada essa presença significativa, procedeu-se a uma segunda análise, mais específica, para entender a correspondência entre os bigramas que tivessem, quer como primeiro quer como segundo termo, alguma das 10 palavras mais frequentes em todo corpus (Visão Panorâmica - Análise 2). Assim, seria possível compreender como as 10 palavras mais frequentes em todo o corpus relacionavam-se entre si num contexto de proximidade textual. Dada o objetivo de perceber certa correlação, todos os pares de palavras repetidas foram suprimidos dessa análise.

Por fim, os bigramas de todo o corpus foram analisados mais uma vez, porém num contexto ainda mais panorâmico, a considerar os 50 pares de palavras mais frequentes, sem considerar os pares de palavras repetidos (Visão Panorâmica - Análise 3). Com esse recorte, como pode-se notar facilmente na visualização de dados, apreende-se de forma mais clara a presença das palavras mais frequentes do corpus (em cor) como primeiro ou segundo termo entre os bigramas mais recorrentes. Essa terceira análise e visualização permite não só identificar quais seriam esses tais termos bigramáticos, mas, sobretudo, observar a articulação entre tais termos e a relevância de tais ocorrências em todo o corpus.