

## MEMÓRIA DESCRITIVA

### REINOS

Para a segunda vaga de análises de dados e visualizações relativas aos três Reinos da Natureza, focou-se nas ocorrências de bigramas com os termos literais e na forma como tais termos dinamizam pares de palavras em todo o corpus com os termos genéricos e específicos relativos aos reinos.

A primeira visualização do conjunto (Reinos da Natureza – Análise 1) mostra, em frequência absoluta distribuída por publicação, a presença dos termos literais “mineral”, “animal” e “vegetal” dentre algum dos dois termos constituintes dos pares de palavras no corpus processado em bigrama. Note-se, por exemplo, que a frequência absoluta dos bigramas será sempre o dobro da frequência absoluta dos termos isolados, dado que um bigrama considera na contagem tanto o termo antecedente como o termo posterior ao termo literal pesquisado. (p. ex. na sequência “grande animal silencioso” tem-se os bigramas “grande animal” e “animal silencioso”). A visualização em questão, organizada pela presença de termos literais, faz visível quais livros constituintes do corpus possuem total ausência de algum (ou alguns) dos termos literais elencados na pesquisa.

A seguir, as análises 2 e 3 relativas aos Reinos da Natureza, oferecem perspectivas comparáveis da relação bigramática entre pares de palavras constituídas de termo literais e de termos específicos (Reinos da Natureza – Análise 2), e os pares de palavras constituídas de termos literais e termos genéricos (Reinos da Natureza – Análise 3). Com vistas a facilitar a comparação entre a prevalência dos reinos relativos aos termos gerais e específicos, a organização dos dados e visualização priorizou a correspondência entre os reinos representados pelos termos (quer literais, genéricos ou específicos). Por exemplo, na Análise 2, há 30 ocorrências do termo literal do reino animal, mas apenas uma delas corresponde a um termo genérico relativo ao reino animal (“animal selvagem”). De igual modo, nessa mesma análise, há apenas uma ocorrência do termo literal do reino mineral dentre as 35 ocorrências de termos genéricos relacionados ao reino mineral em bigramas por todo o corpus (a saber, “mar mineral”). A Análise 3, para além de mostrar que existem menos ocorrências entre termos literais e específicos nos bigramas relacionados aos Reinos da Natureza, explicita a raridade (todas as ocorrências são únicas) e mesmo a ausência total de qualquer relação entre os termos literais e específico, como no caso do reino mineral.

Quanto à quarta análise, mais experimental, procedeu-se à adoção dum segundo modelo de análise contextual, o KWIC (acrônimo para “keywords in context”, ou em português, palavras-chave em contexto). Esse modelo prevê a identificação de palavras específicas dentro do corpus, distinguindo o contexto pela consideração de um número selecionável de palavras precedentes e posteriores ao(s) termo(s) em foco. Para tanto, foi utilizado um outro tipo de processamento de corpus, no qual não se procedeu à lematização nem à retirada das palavras de paragem, de forma que as frases fizessem sentido. Optou-se pela organização das ocorrências de modo a respeitar sua colocação cronológica quanto às publicações e dentro da sequência original do texto, considerando-se as 5 palavras antecedentes e procedentes aos termos em vista. Os termos analisados foram designados pelo radical, isto é, de forma a considerar no resultado suas diversas formas nominais e derivações. Isso foi feito através do processo de “stemming”, possibilitado pelo uso do pacote “tm” (Text Mining package) para Linguagem R. Por fim, cabe notar que desenvolver uma visualização a partir desse tipo de análise é desafiador dada a sua natureza altamente descritiva, e pode ser que seja mesmo dispensável, pela natureza matricial da apresentação de seu resultado. Uma última nota para destacar que, quando utilizada em conjunto com a primeira análise de dados bigramática do conjunto dos Reinos da Natureza, essa quarta análise contextual, complementa e especifica os dados analisados em relação a ocorrência dos bigramas em cada livro, sendo de grande utilidade sua leitura paralela de outras análises e visualizações.

Patrícia Reina