

Preprocesado: Selección de datos

Tras eliminar las palabras raras, creamos la matriz de términos y obtenemos gráficos para entender mejor el dominio.



Preprocesado: Bigramas

Al aparecer palabras por separado que son nombre y apellidos. Se realiza un estudio 2-gramas para ver si esto es realmente así. Al comprobarse la premisa **se unen** estas palabras.

