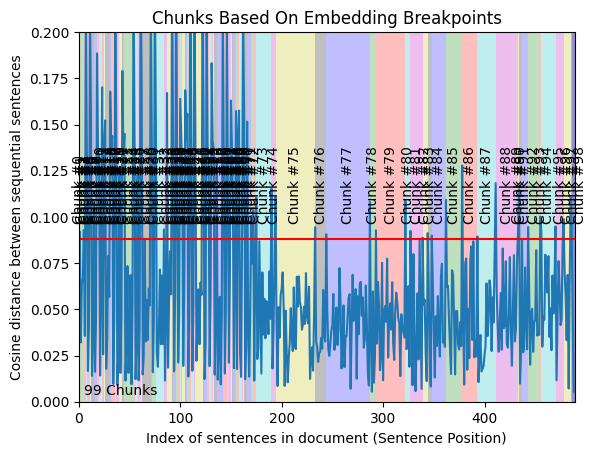
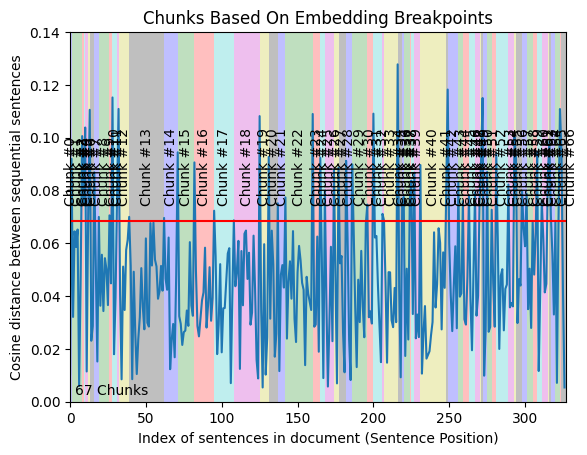
Resultado de Evaluación de Chunking Semántico version2 con un breakpoint\_percentile\_threshold = 80 con el pdf completo (incluida tabla de contenidos). Esto se produce por que cada “oración” es muy distinta a la siguiente (temas distintos) y por lo tanto sus embeddings tienen distancias coseno significativas.



Al quitar la tabla de contenidos ocurre lo siguiente:



Quedando chunks de tamaño más homogéneos lo cual se puede controla con el parámetro breakpoint\_percentile\_threshold.