

INFORME SOBRE EL PROCESO DE SELECCIÓN DE PREGUNTAS PARA EL 1ER DEBATE PRESIDENCIAL INE 2024 (FORMATO A)

Anexo 4. Compilación de gráficas de análisis semántico por tema en la muestra

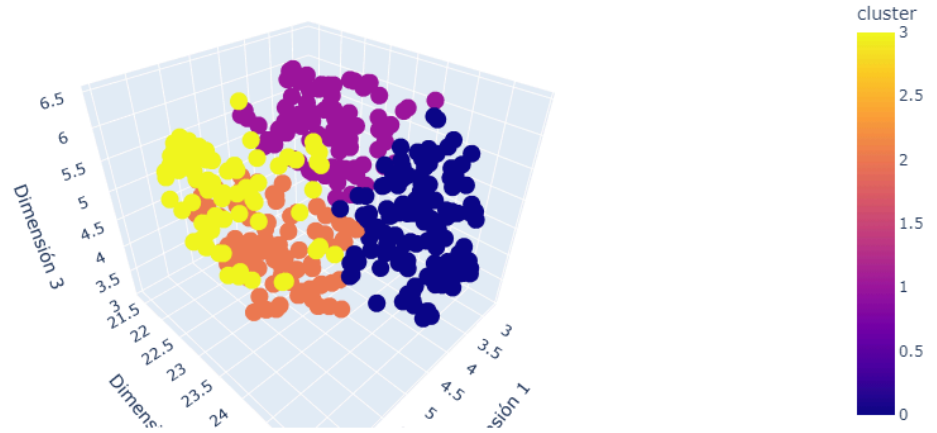
Todas las capturas, gráficas y tablas fueron generadas por el equipo de Signa_Lab ITESO



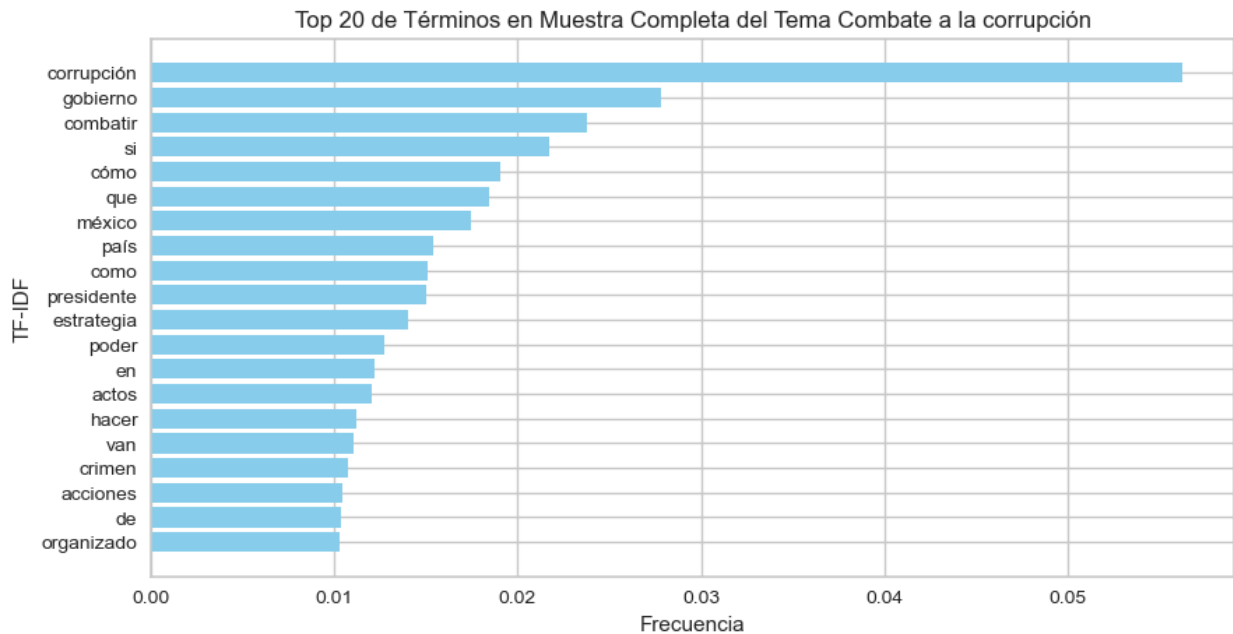
COMBATE A LA CORRUPCIÓN

1. Visualización de clusters semánticos de Combate a la Corrupción

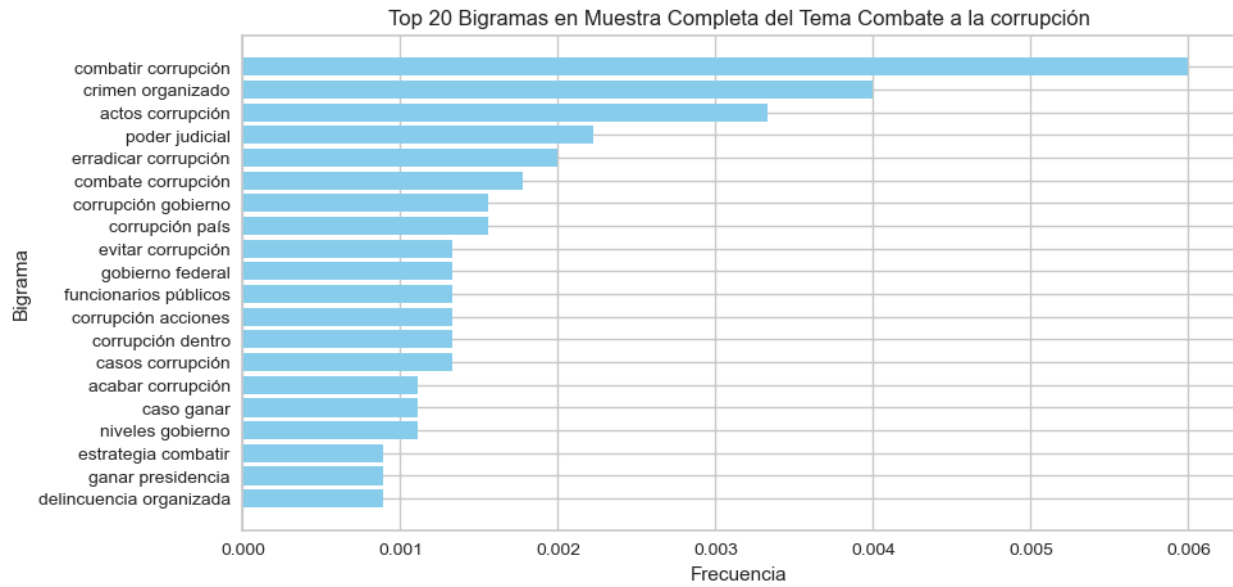
Visualización de clústers semánticos de Combate a la corrupción



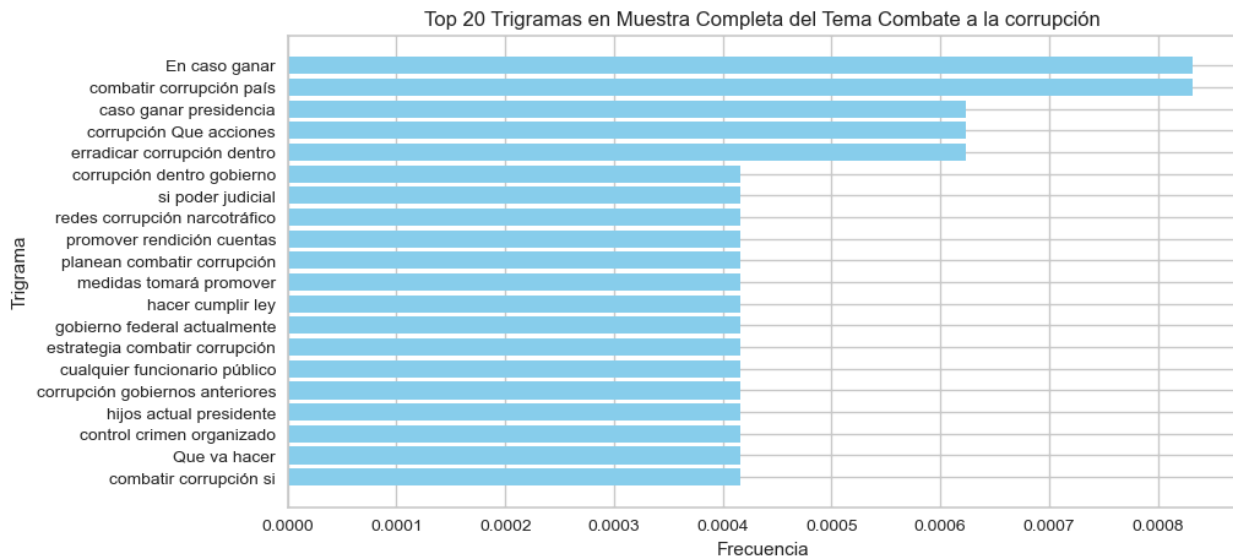
2. Top 20 términos en la muestra completa del tema Combate a la Corrupción



3. Top 20 bigramas en la muestra completa del tema Combate a la Corrupción



4. Top 20 trigramas en la muestra completa del tema Combate a la Corrupción



5. Análisis semántico por clusters (conteo de preguntas por cluster)

Análisis Semántico por Clústers. Frecuencia de términos y relaciones entre palabras por clúster:

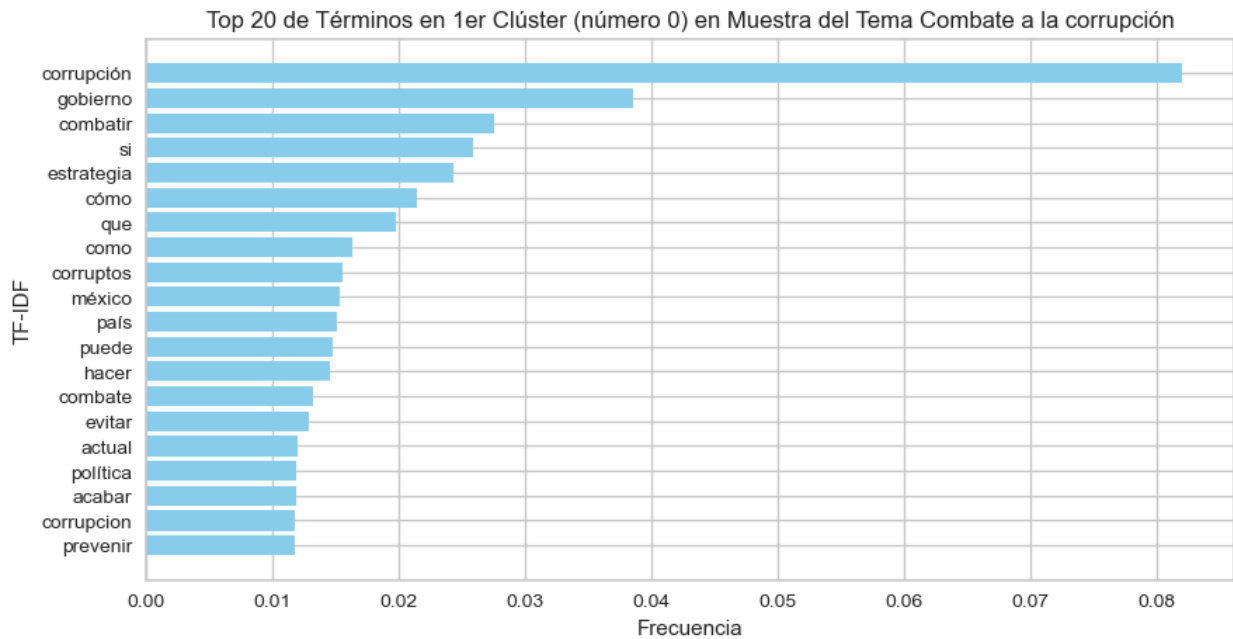
Identificar clústers más grandes en la muestra del tema (mayor frecuencia de preguntas por similitud semántica):

```
In [43]: # Identificar y enlistar clusters con mayor cantidad de preguntas (más frecuentes) del tema
for cluster, freq in dfClusterizado['cluster'].value_counts(ascending=False).head(10).items():
    print(f"Cluster {cluster} - {freq} preguntas")

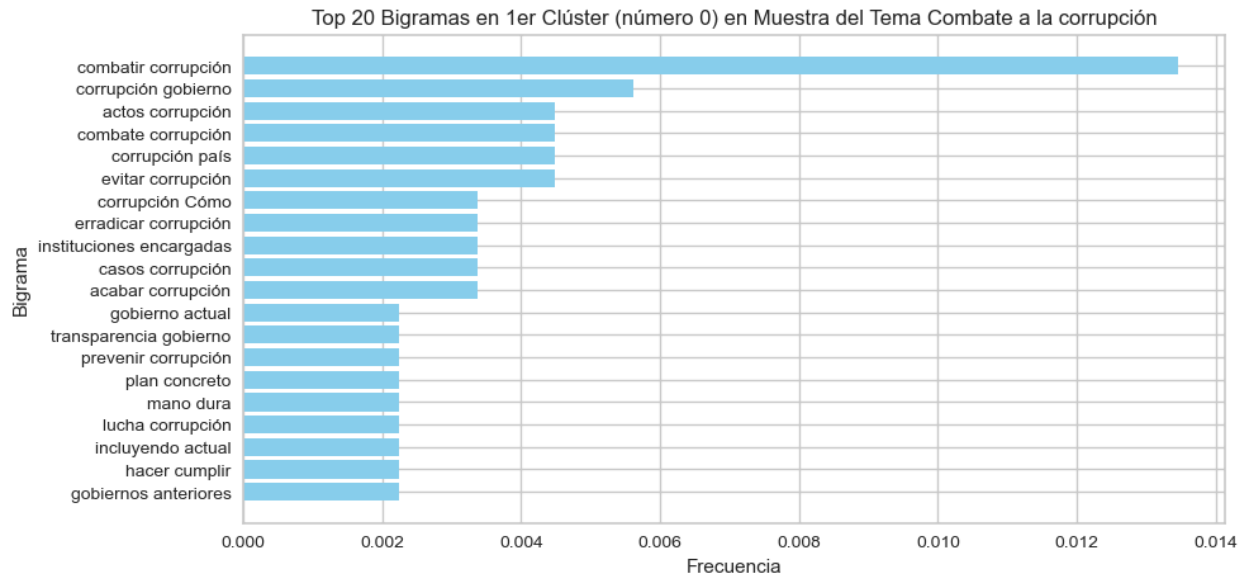
print()
```

```
Cluster 0 - 138 preguntas
Cluster 1 - 98 preguntas
Cluster 2 - 77 preguntas
Cluster 3 - 76 preguntas
```

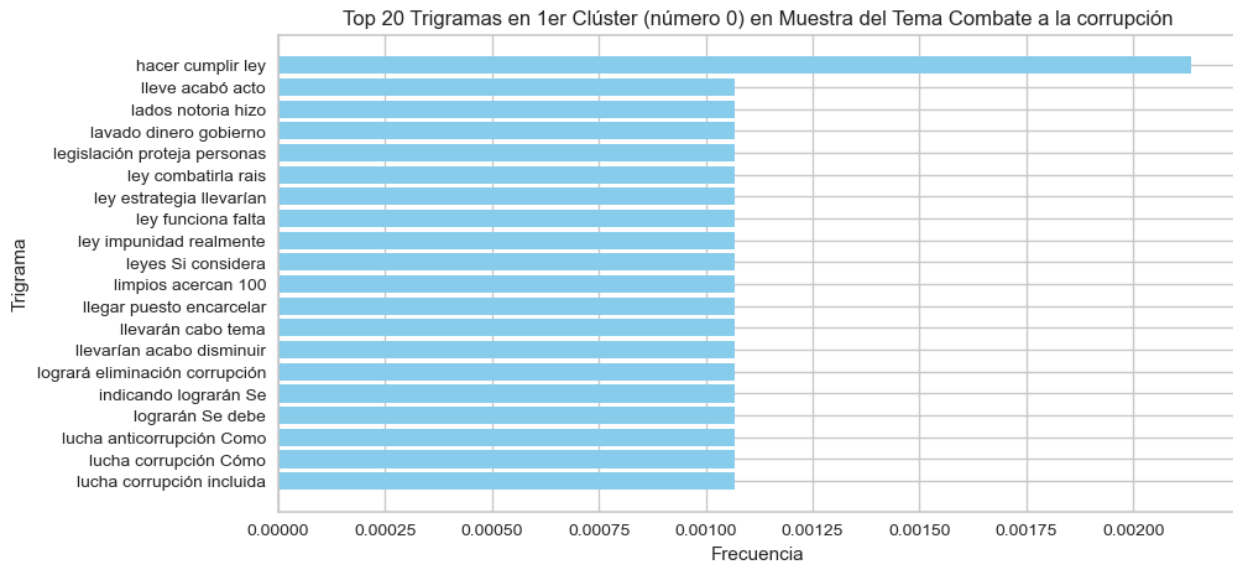
6. Top 20 términos en el primer cluster del tema Combate a la Corrupción



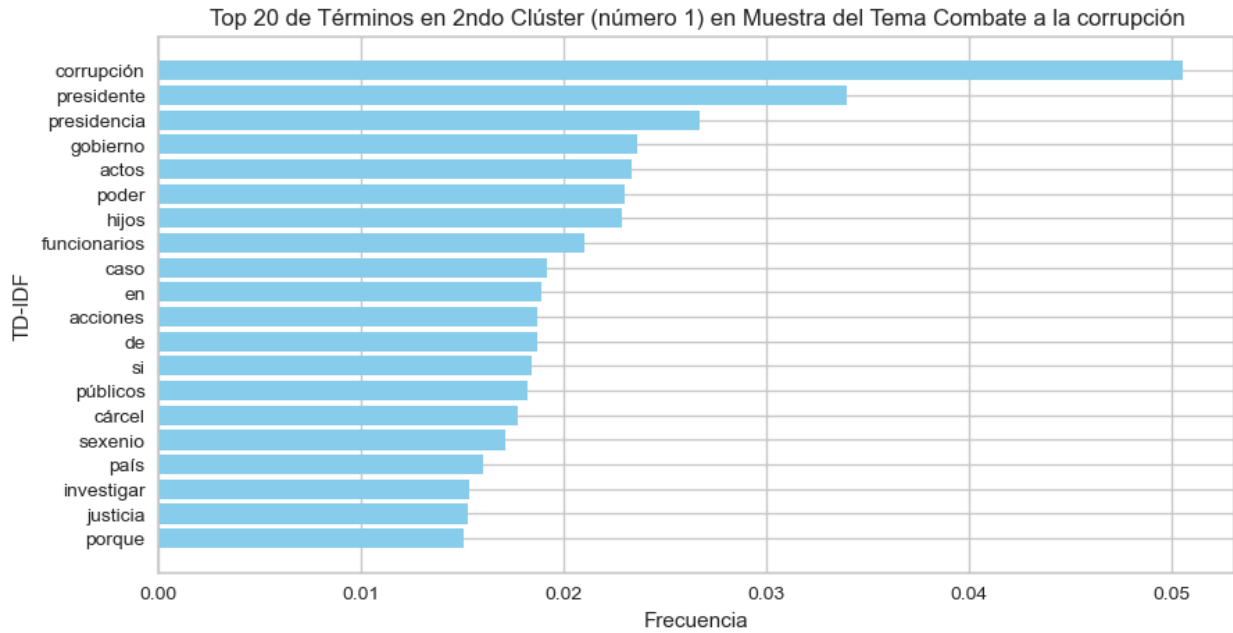
7. Top 20 bigramas en el primer cluster del tema Combate a la Corrupción



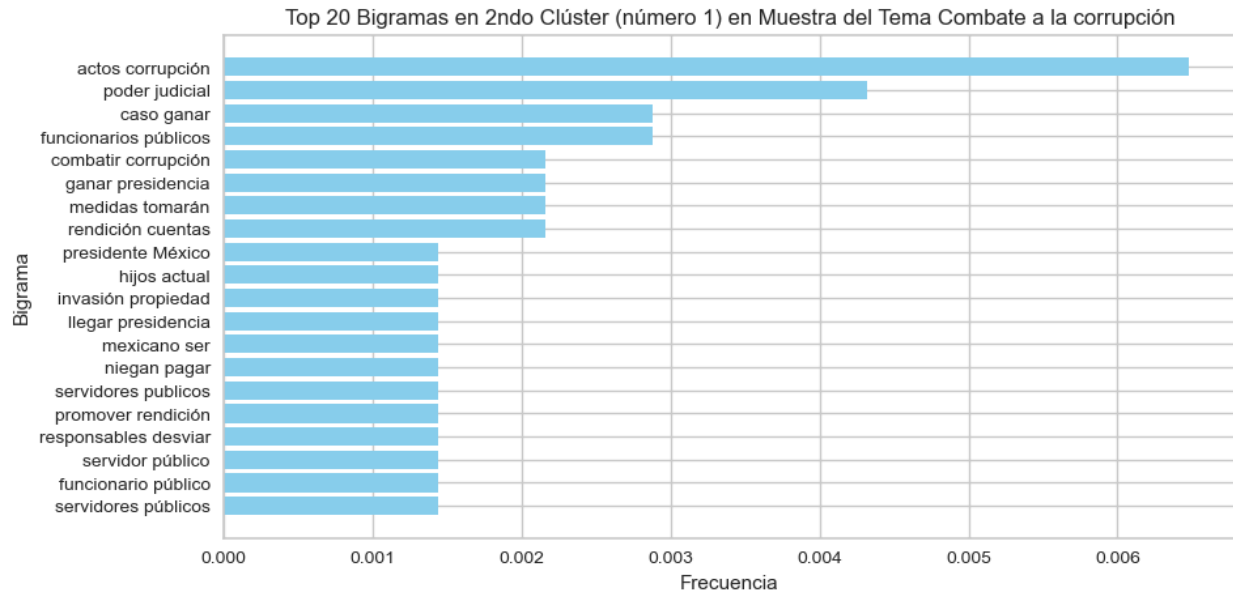
8. Top 20 trigramas en el primer cluster del tema Combate a la Corrupción



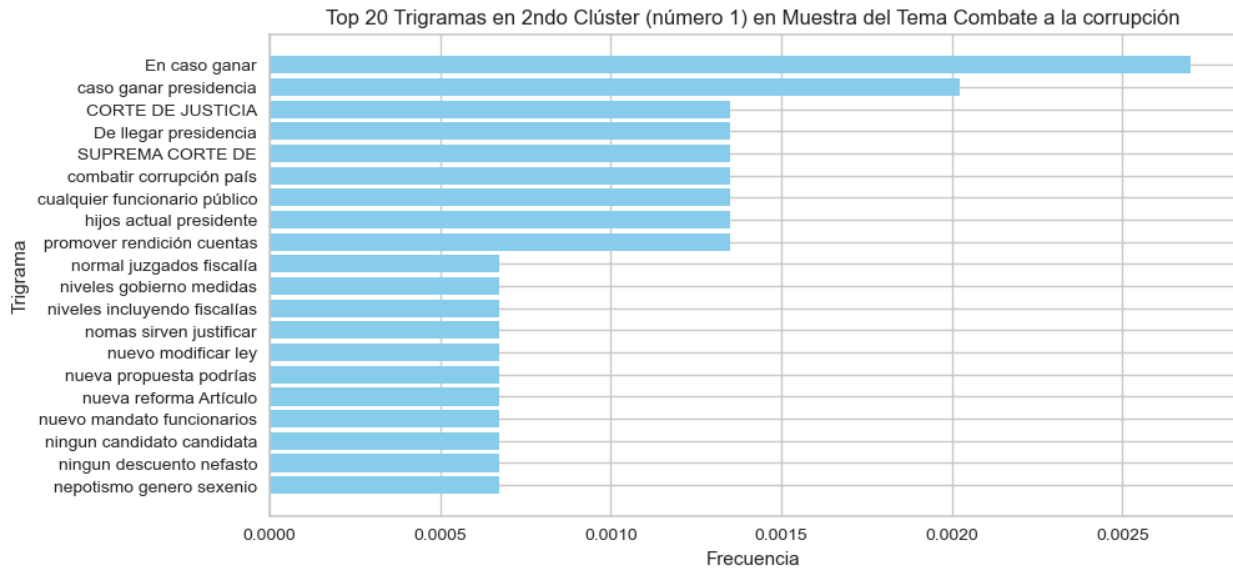
9. Top 20 términos en el segundo cluster del tema Combate a la Corrupción



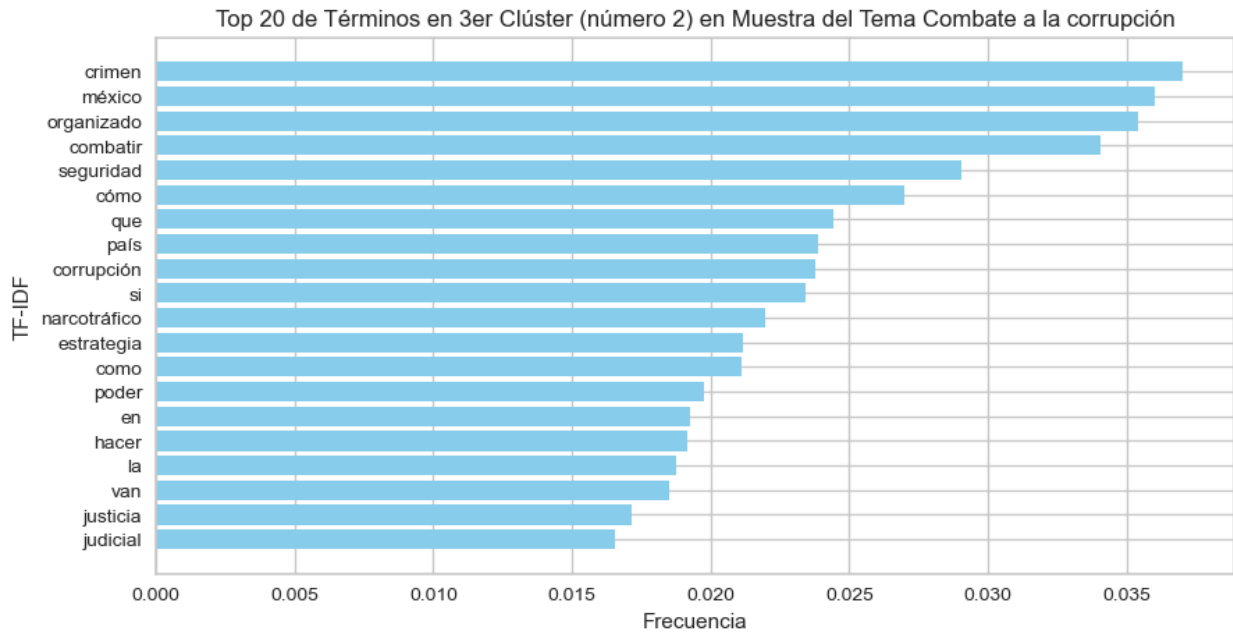
10. Top 20 bigramas en el segundo cluster del tema Combate a la Corrupción



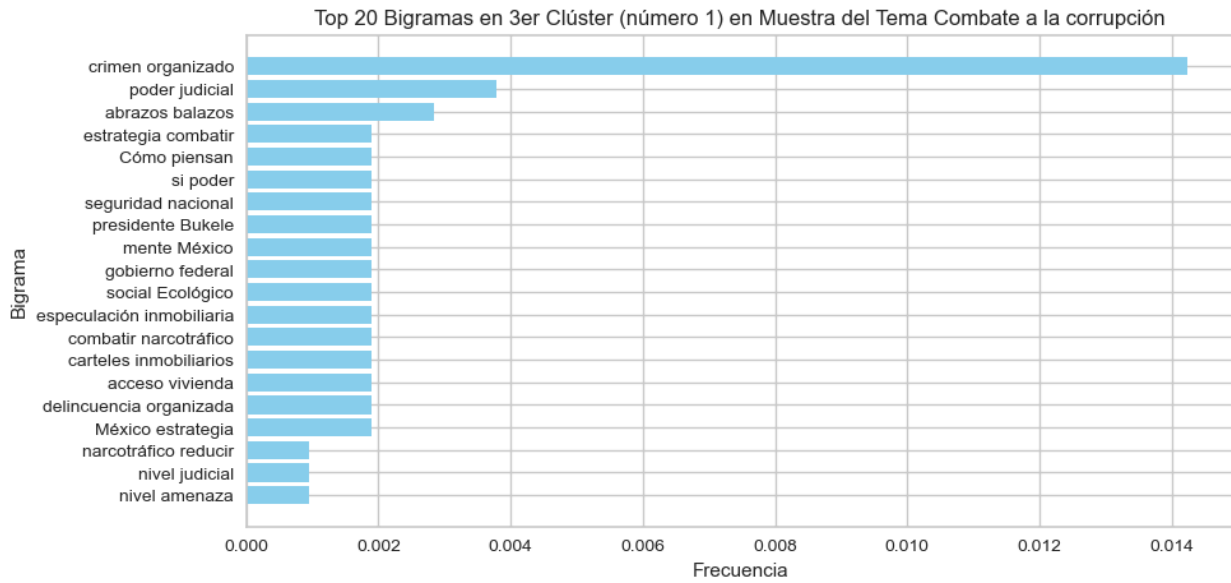
11. Top 20 trigramas en el segundo cluster del tema Combate a la Corrupción



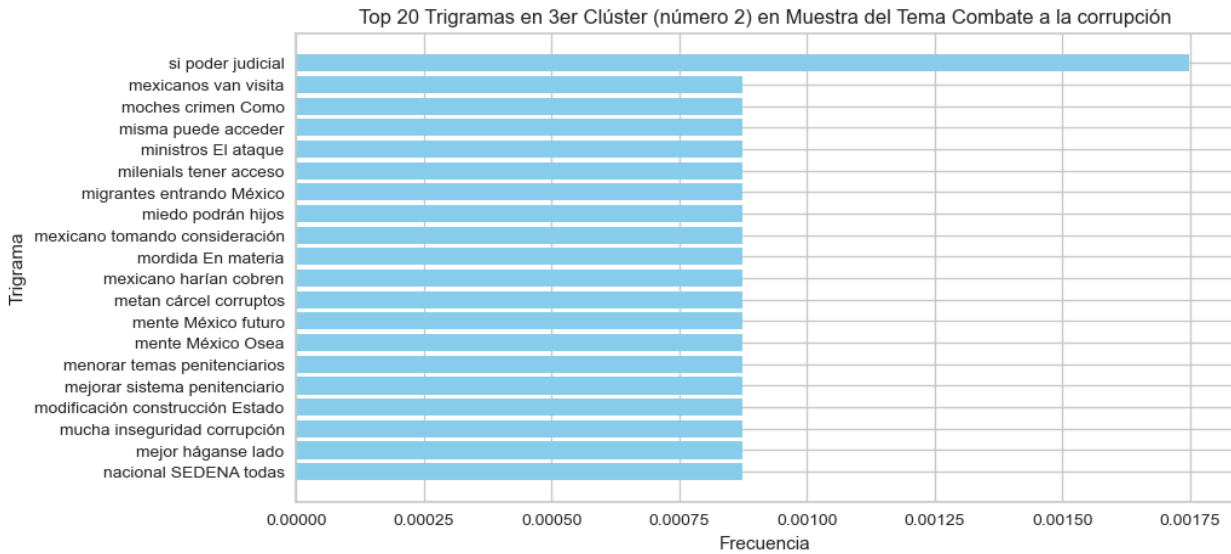
12. Top 20 términos en el tercer cluster del tema Combate a la Corrupción



13. Top 20 bigramas en el tercer cluster del tema Combate a la Corrupción



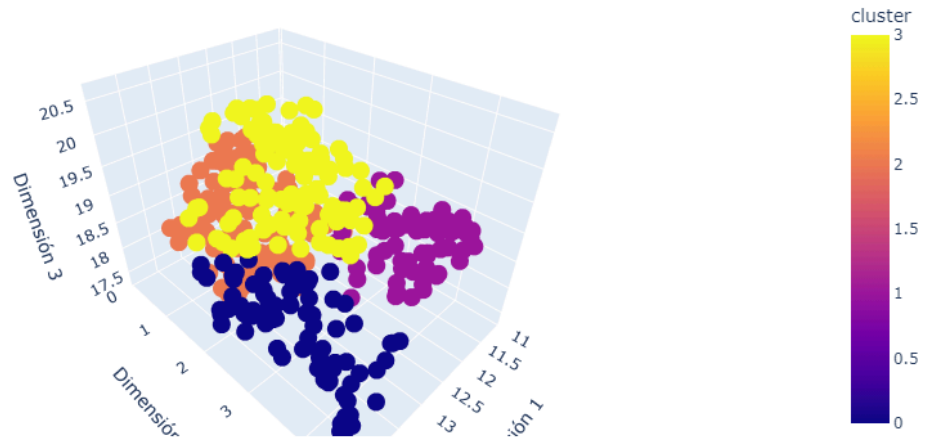
14. Top 20 trigramas en el tercer cluster del tema Combate a la Corrupción



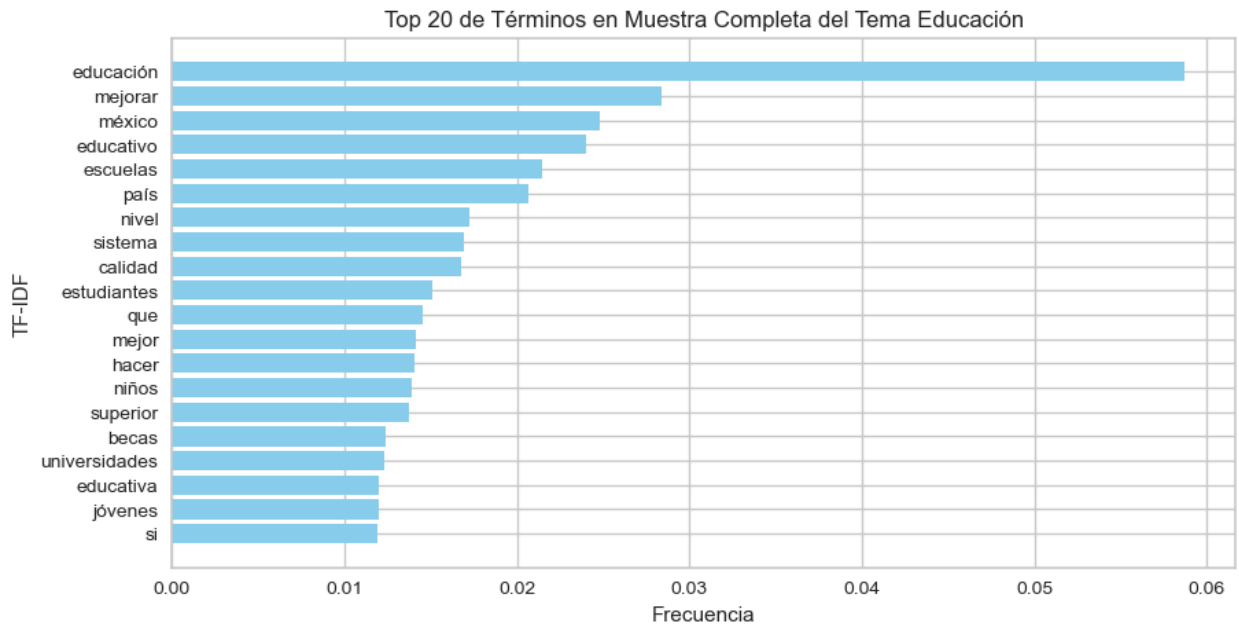
EDUCACIÓN

1. Visualización de clusters semánticos de Educación

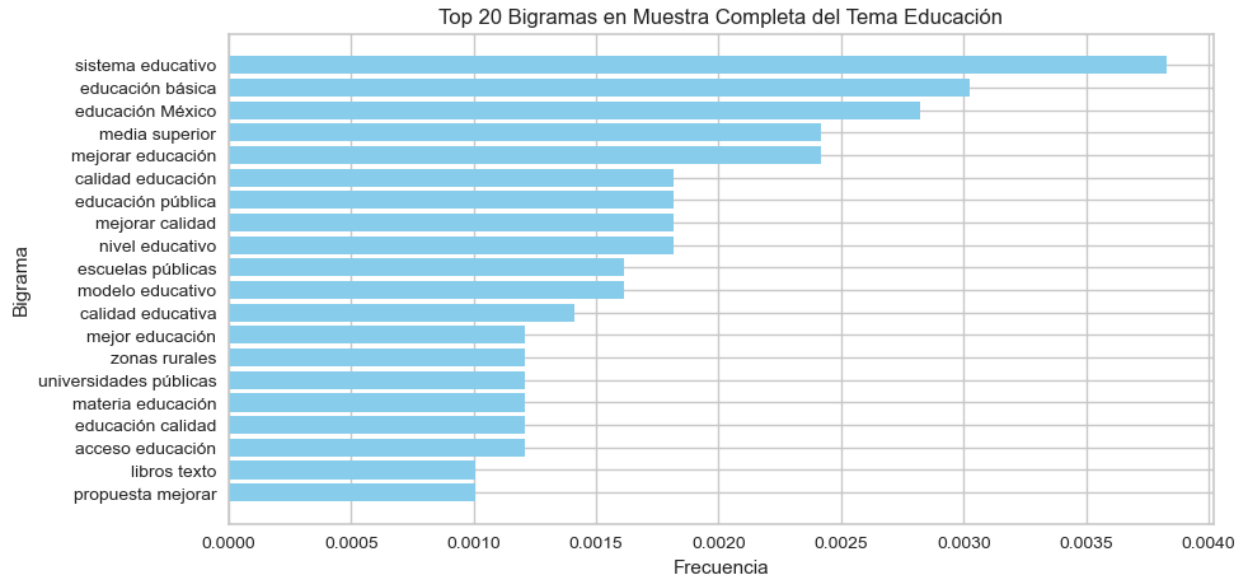
Visualización de clústers semánticos de Educación



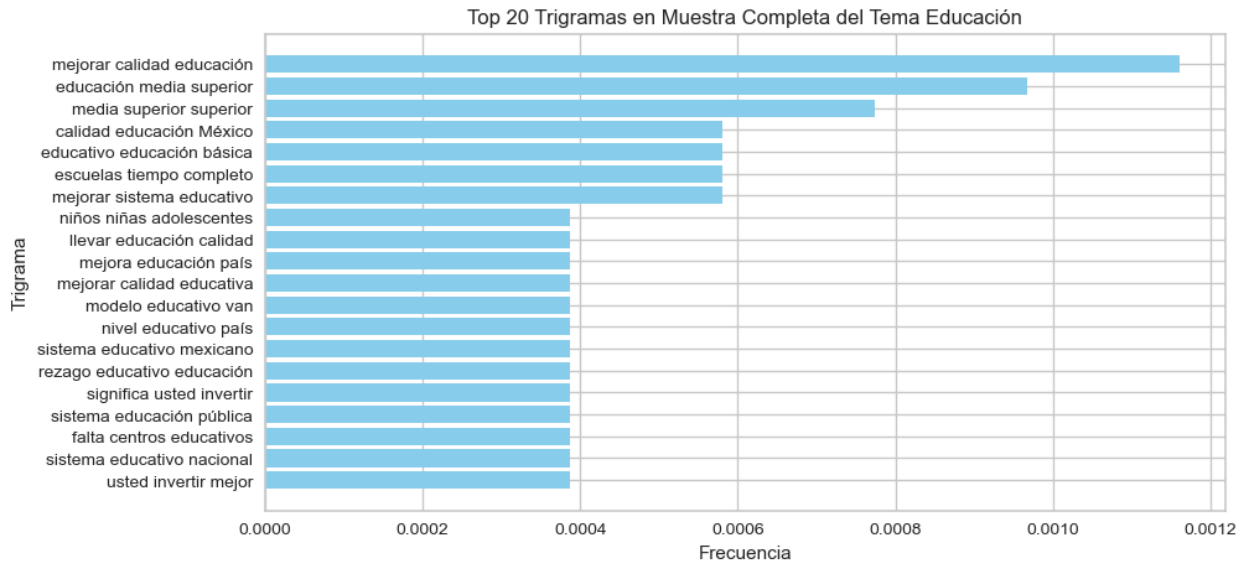
2. Top 20 términos en la muestra completa del tema Educación



3. Top 20 bigramas en la muestra completa del tema Educación



4. Top 20 trigramas en la muestra completa del tema Educación



5. Análisis semántico por clusters (conteo de preguntas por cluster)

Análisis Semántico por Clústers. Frecuencia de términos y relaciones entre palabras por clúster:

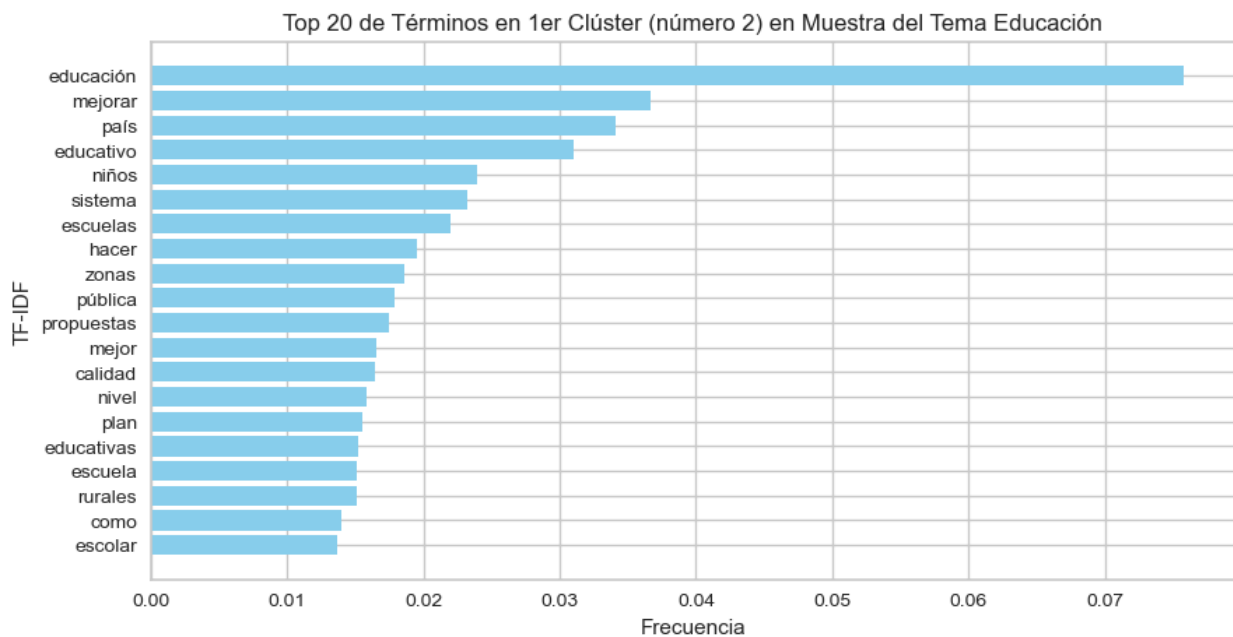
Identificar clústers más grandes en la muestra del tema (mayor frecuencia de preguntas por similitud semántica):

```
In [43]: # Identificar y enlistar clusters con mayor cantidad de preguntas (más frecuentes) del tema
for cluster, freq in dfClusterizado['cluster'].value_counts(ascending=False).head(10).items():
    print(f"Cluster {cluster} - {freq} preguntas")

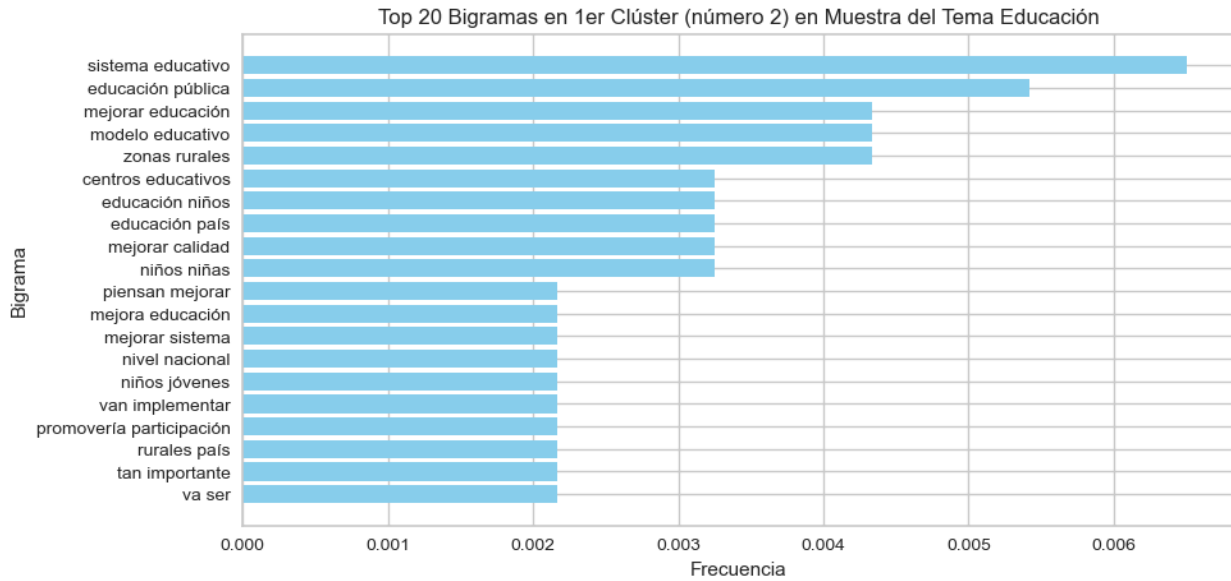
print()
```

```
Cluster 0 - 138 preguntas
Cluster 1 - 98 preguntas
Cluster 2 - 77 preguntas
Cluster 3 - 76 preguntas
```

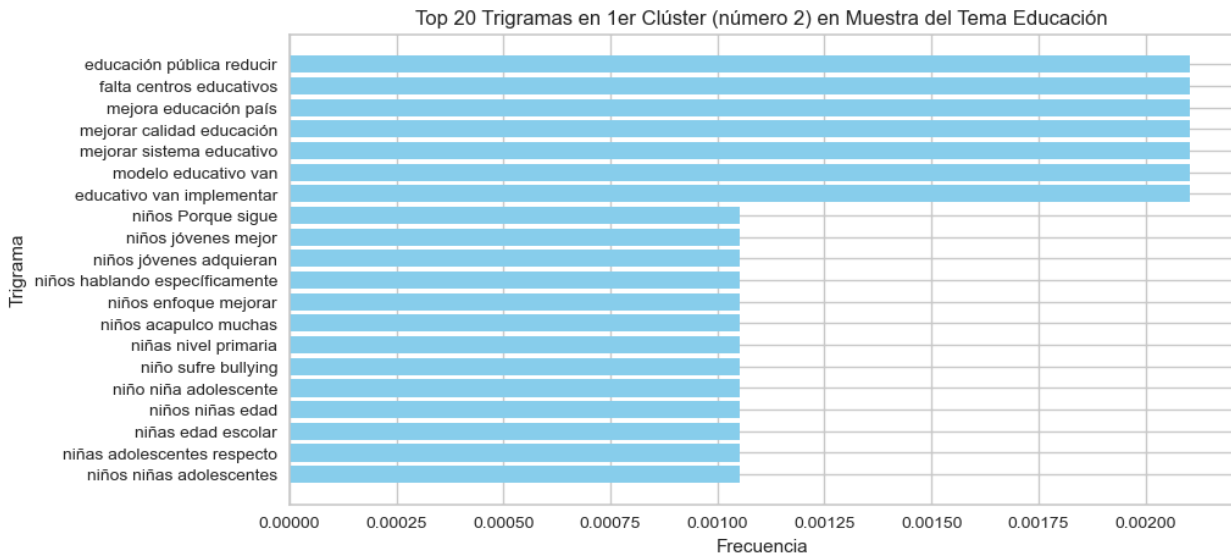
6. Top 20 términos en el primer cluster del tema Educación



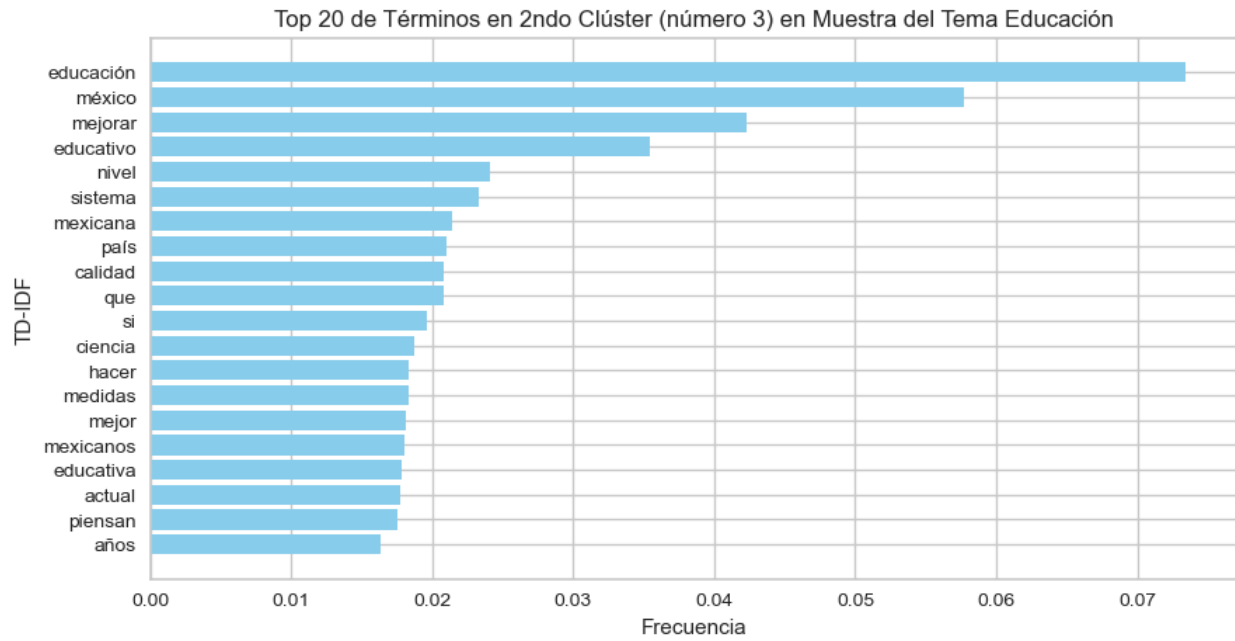
7. Top 20 bigramas en el primer cluster del tema Educación



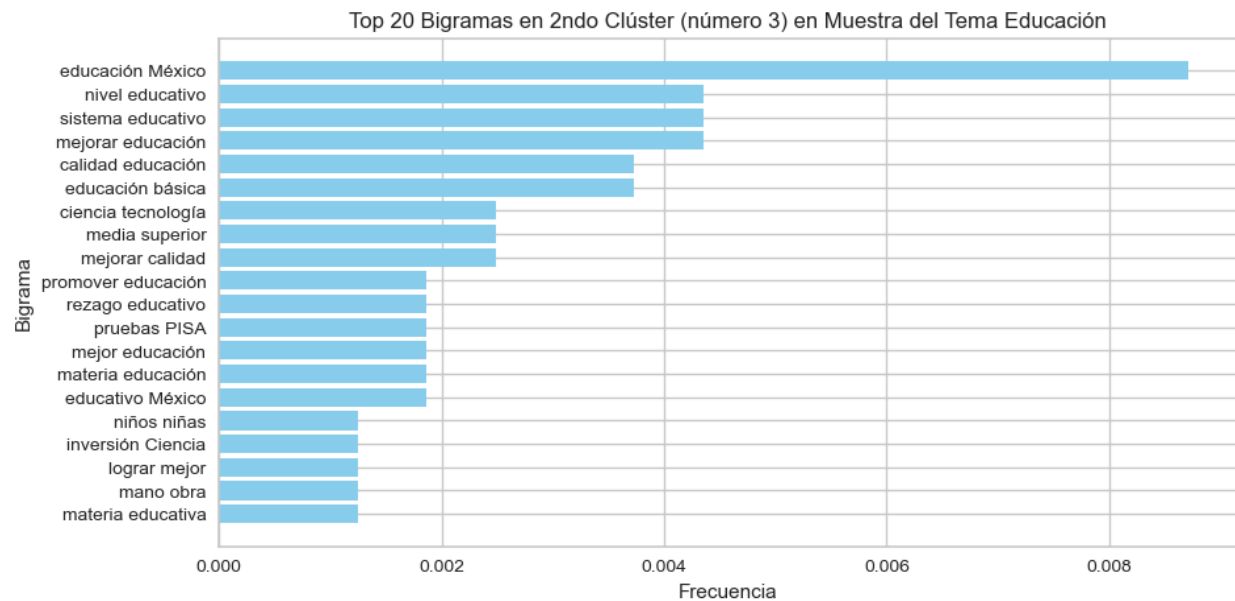
8. Top 20 trigramas en el primer cluster del tema Educación



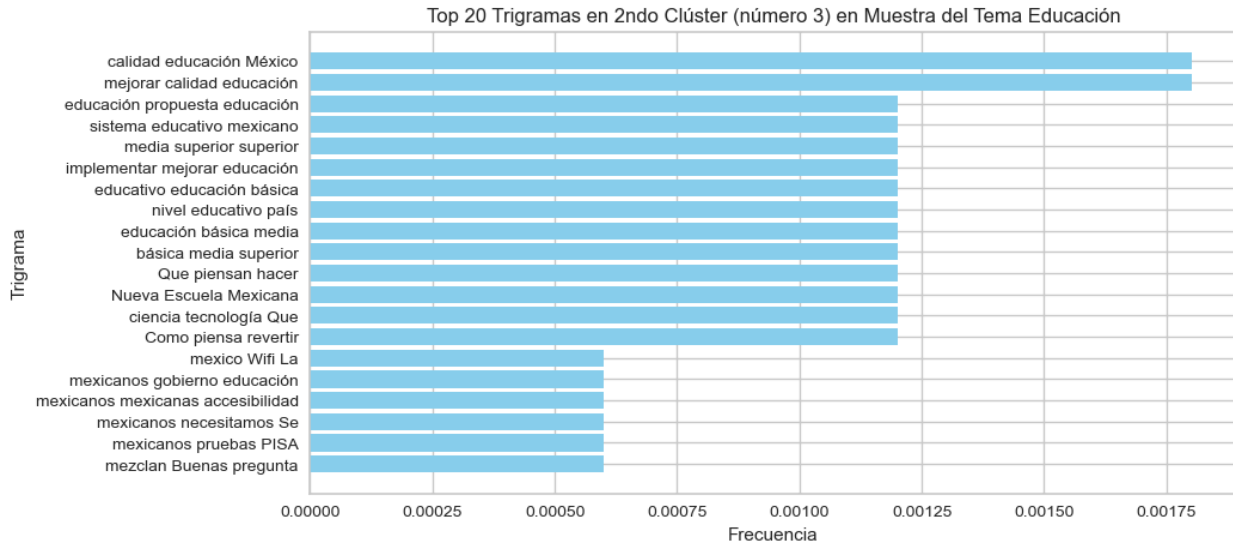
9. Top 20 términos en el segundo cluster del tema Educación



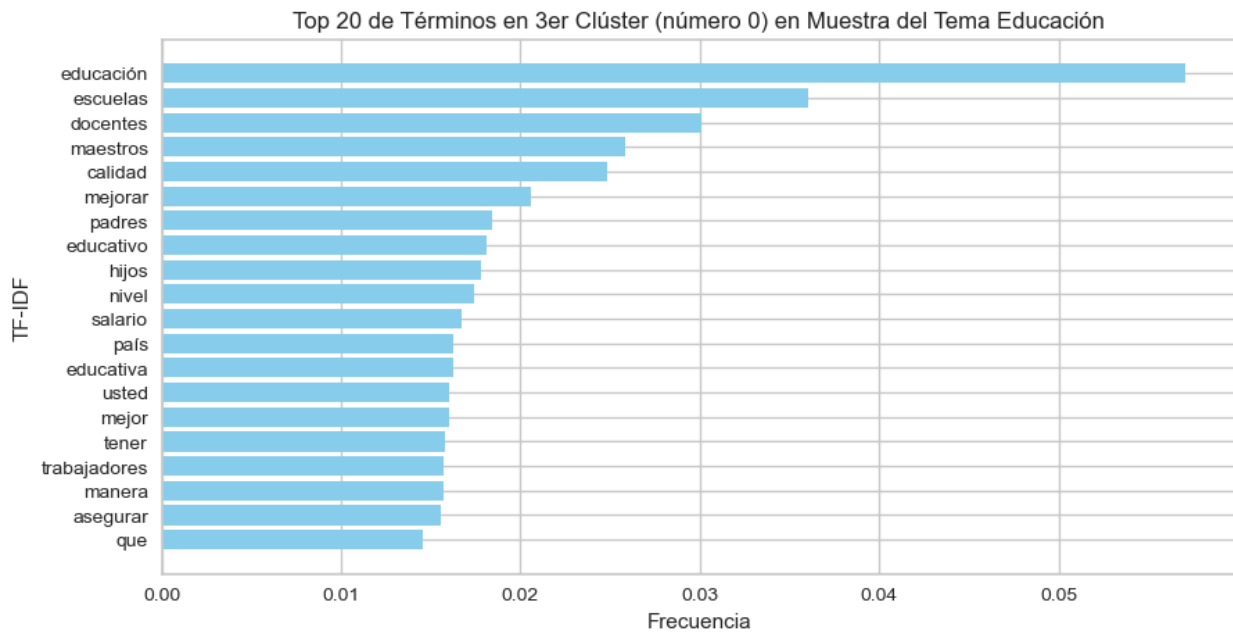
10. Top 20 bigramas en el segundo cluster del tema Educación



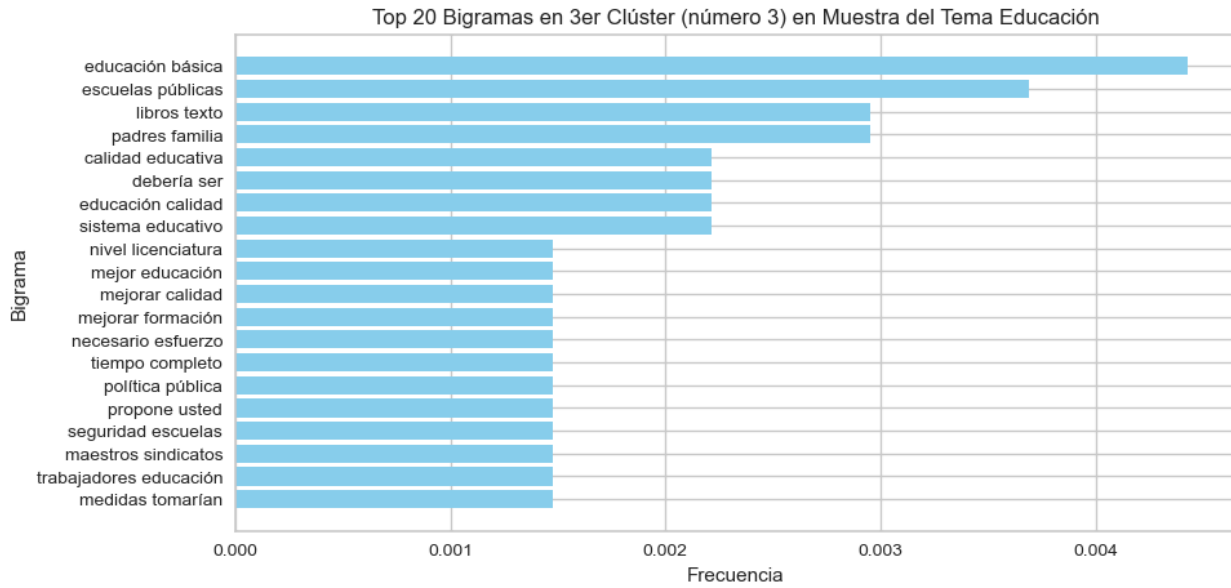
11. Top 20 trigramas en el segundo cluster del tema Educación



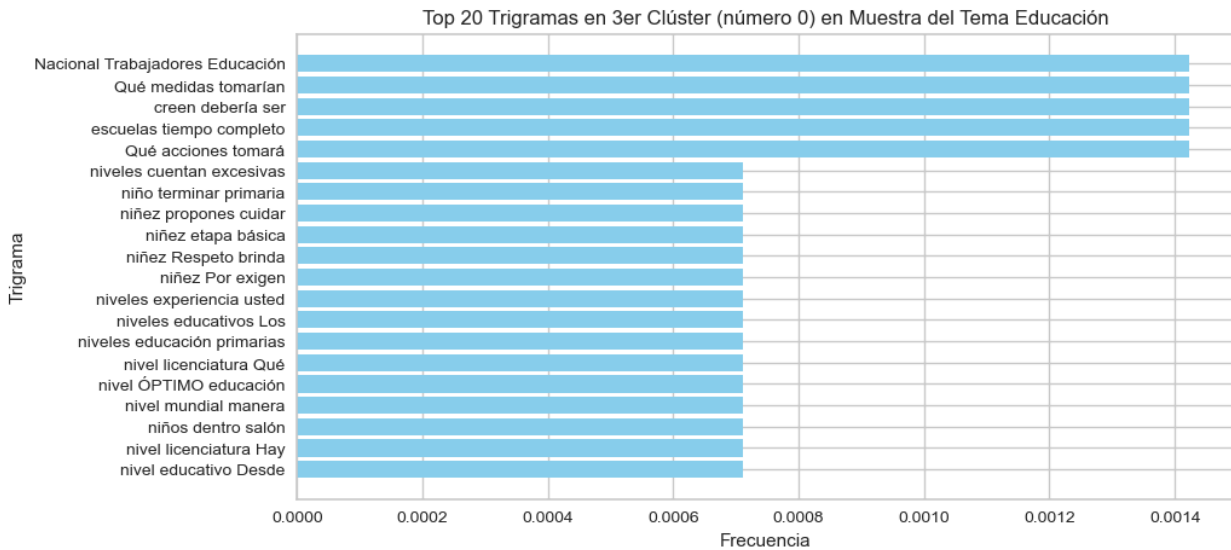
12. Top 20 términos en el tercer cluster del tema Educación



13. Top 20 bigramas en el tercer cluster del tema Educación



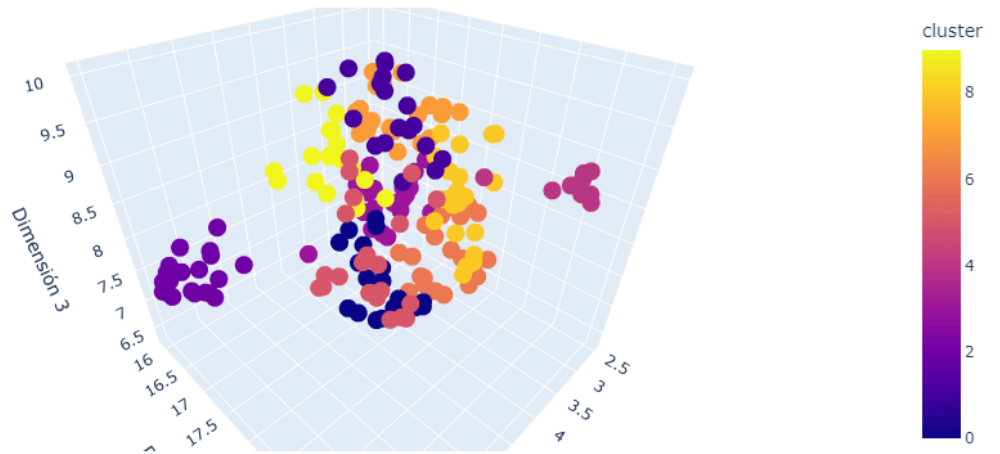
14. Top 20 trigramas en el tercer cluster del tema Educación



NO DISCRIMINACIÓN Y GRUPOS VULNERABLES

1. Visualización de clusters semánticos de No Discriminación y Grupos Vulnerables

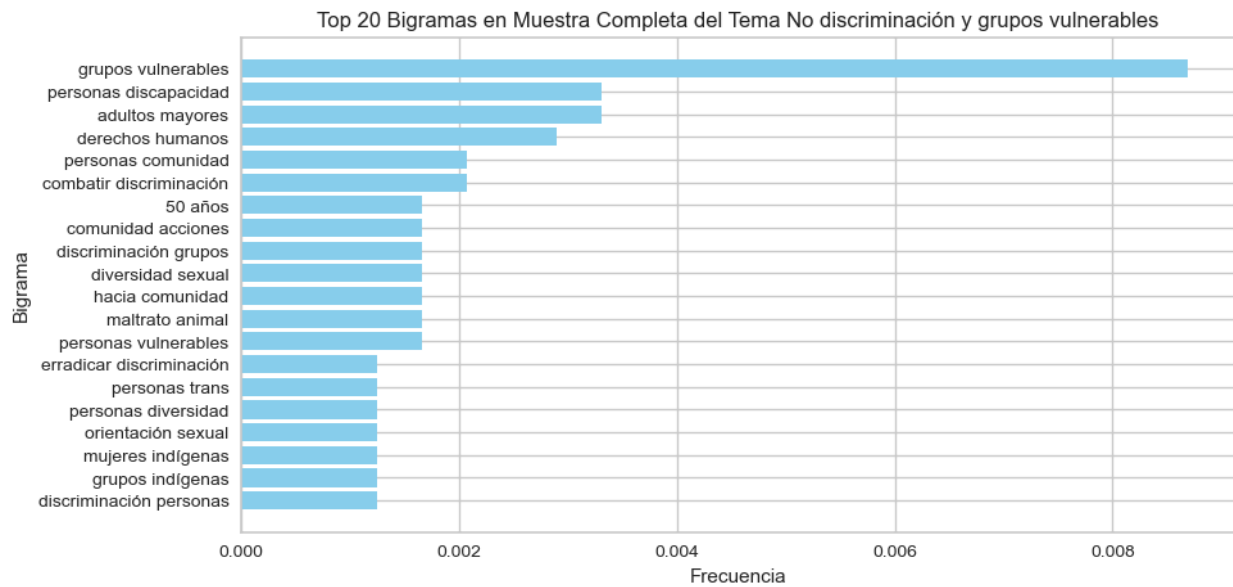
Visualización de clústers semánticos de No discriminación y grupos vulnerables



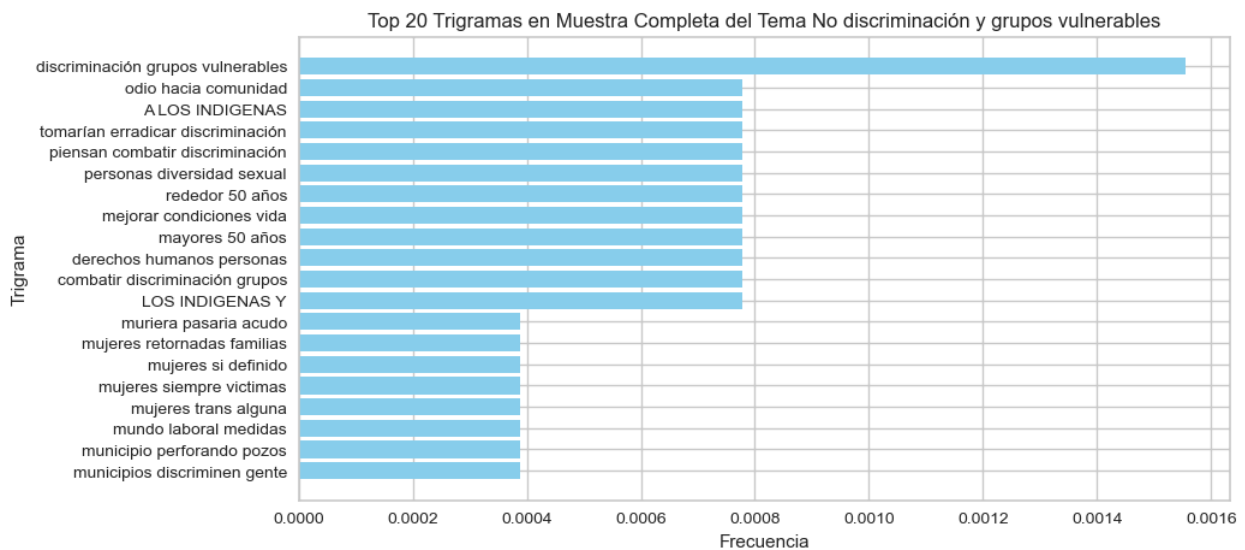
2. Top 20 términos en la muestra completa del tema No Discriminación y Grupos Vulnerables



3. Top 20 bigramas en la muestra completa del tema No Discriminación y Grupos Vulnerables



4. Top 20 trigramas en la muestra completa del tema No Discriminación y Grupos Vulnerables



5. Análisis semántico por clusters (conteo de preguntas por cluster)

Análisis Semántico por Clústers. Frecuencia de términos y relaciones entre palabras por clúster:

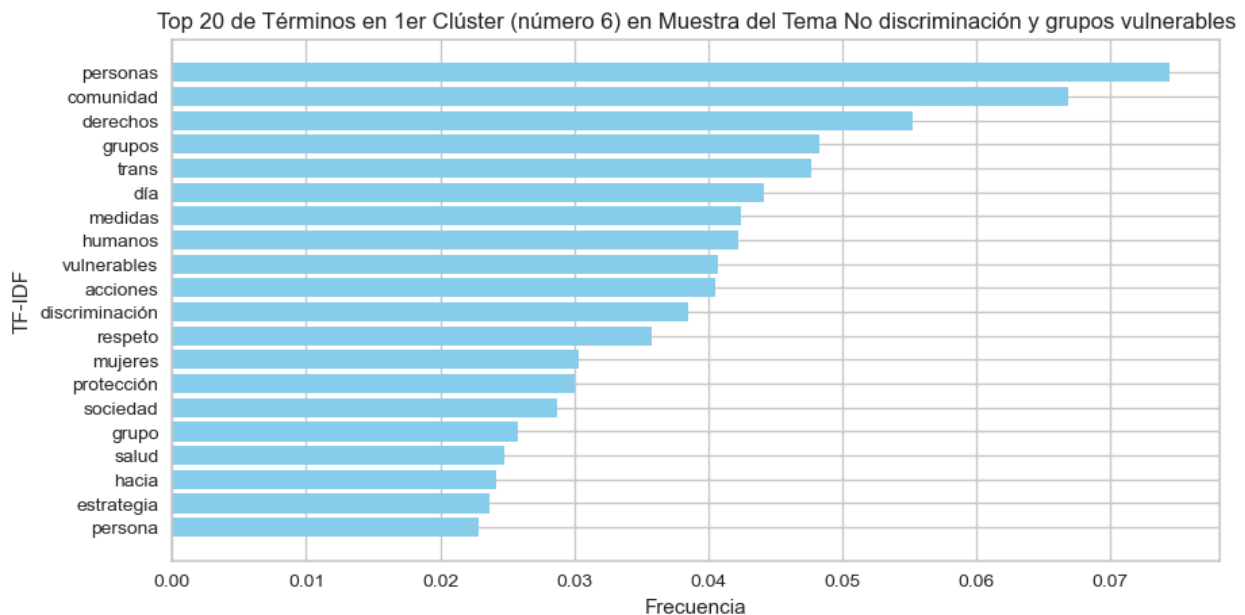
Identificar clústers más grandes en la muestra del tema (mayor frecuencia de preguntas por similitud semántica):

```
In [40]: # Identificar y enlistar clusters con mayor cantidad de preguntas (más frecuentes) del tema
for cluster, freq in dfclusterizado['cluster'].value_counts(ascending=False).head(10).items():
    print(f"Cluster {cluster} - {freq} preguntas")

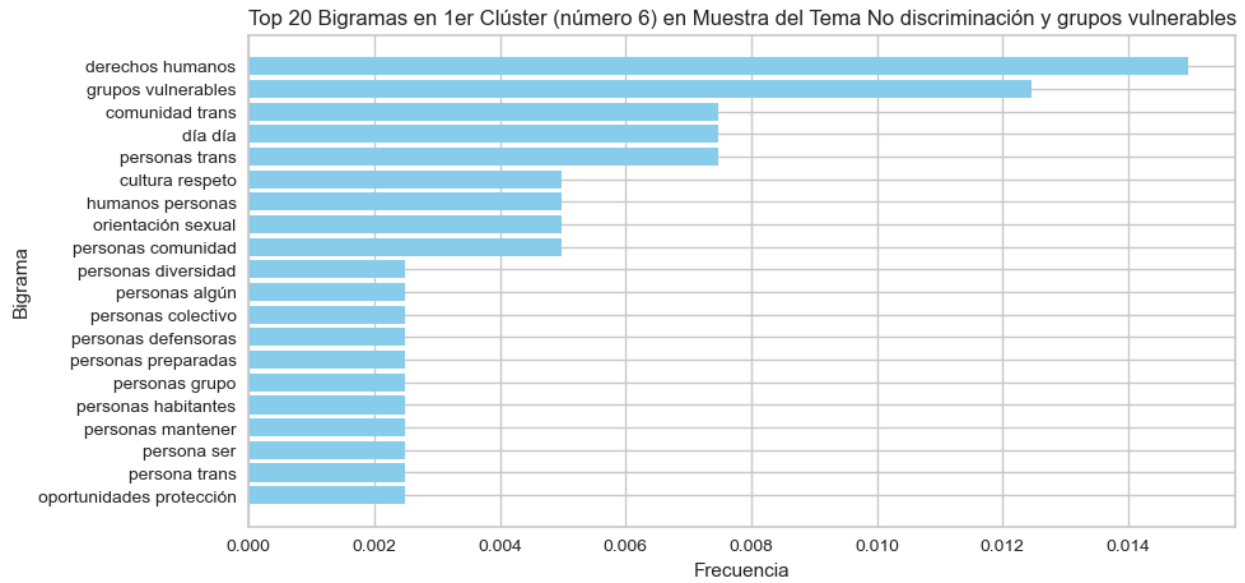
print()
```

```
Cluster 6 - 25 preguntas
Cluster 1 - 22 preguntas
Cluster 3 - 21 preguntas
Cluster 2 - 20 preguntas
Cluster 5 - 20 preguntas
Cluster 0 - 19 preguntas
Cluster 7 - 18 preguntas
Cluster 4 - 18 preguntas
Cluster 9 - 15 preguntas
Cluster 8 - 15 preguntas
```

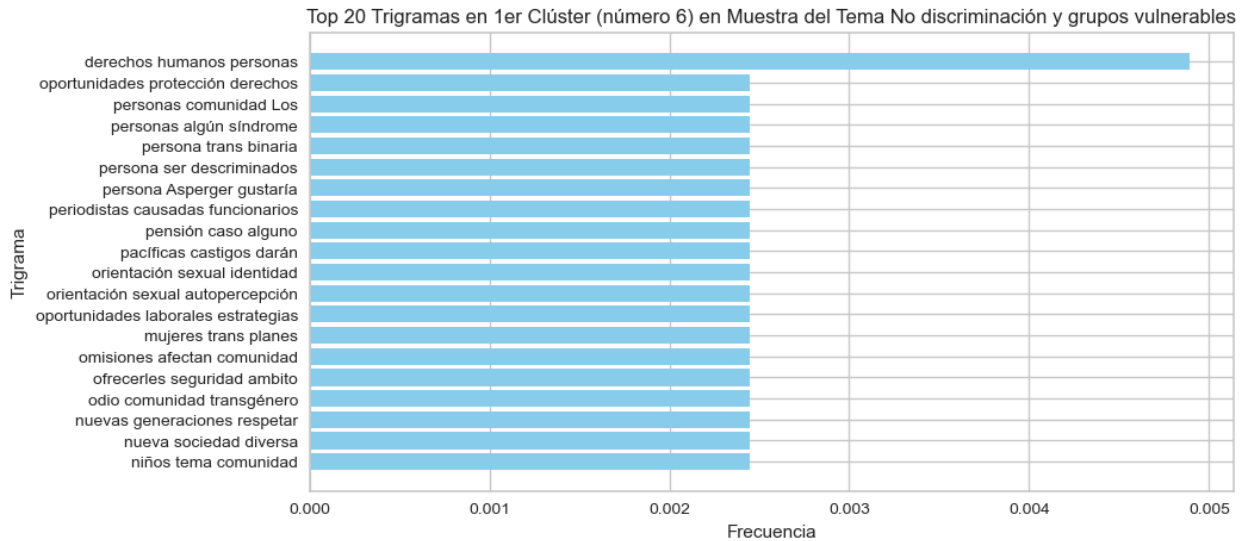
6. Top 20 términos en el primer cluster del tema No Discriminación y Grupos Vulnerables



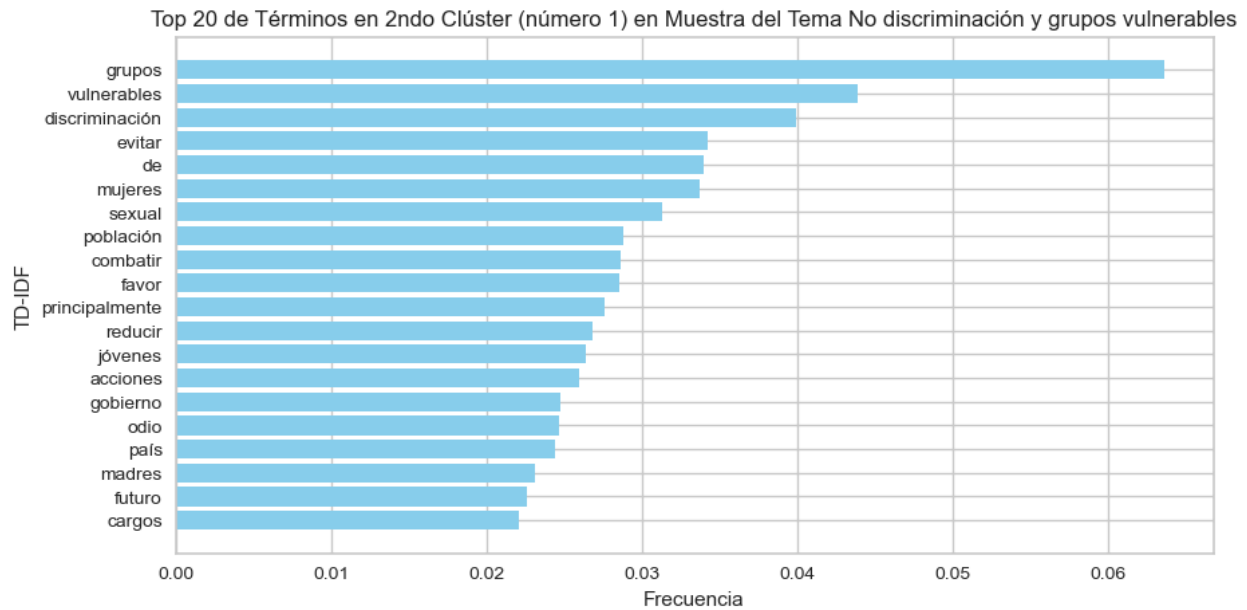
7. Top 20 bigramas en el primer cluster del tema No Discriminación y Grupos Vulnerables



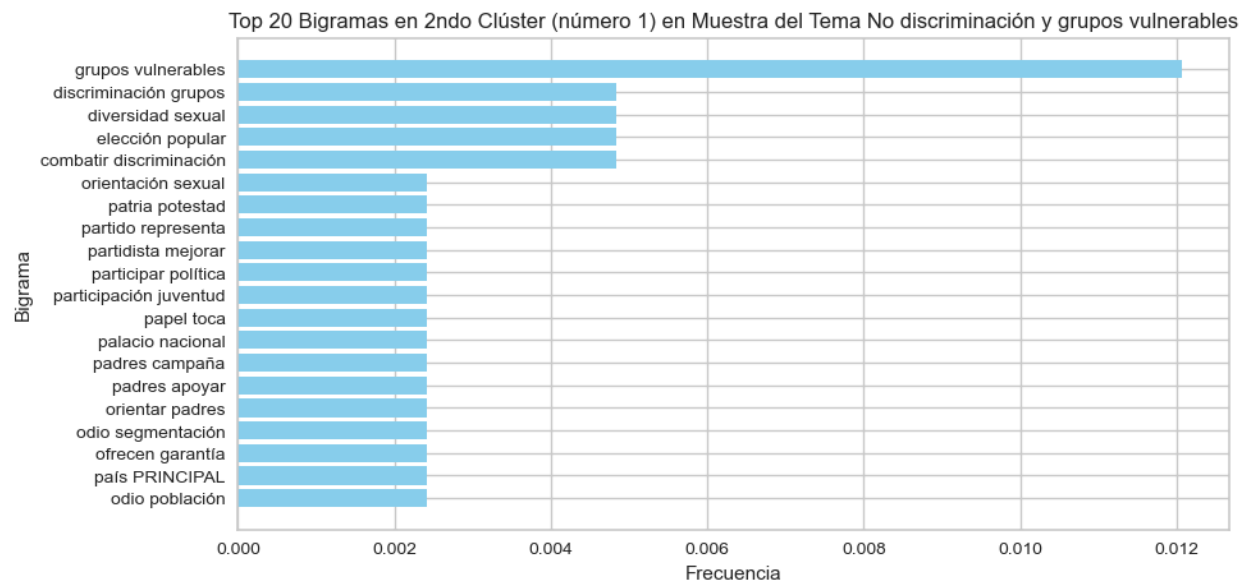
8. Top 20 trigramas en el primer cluster del tema No Discriminación y Grupos Vulnerables



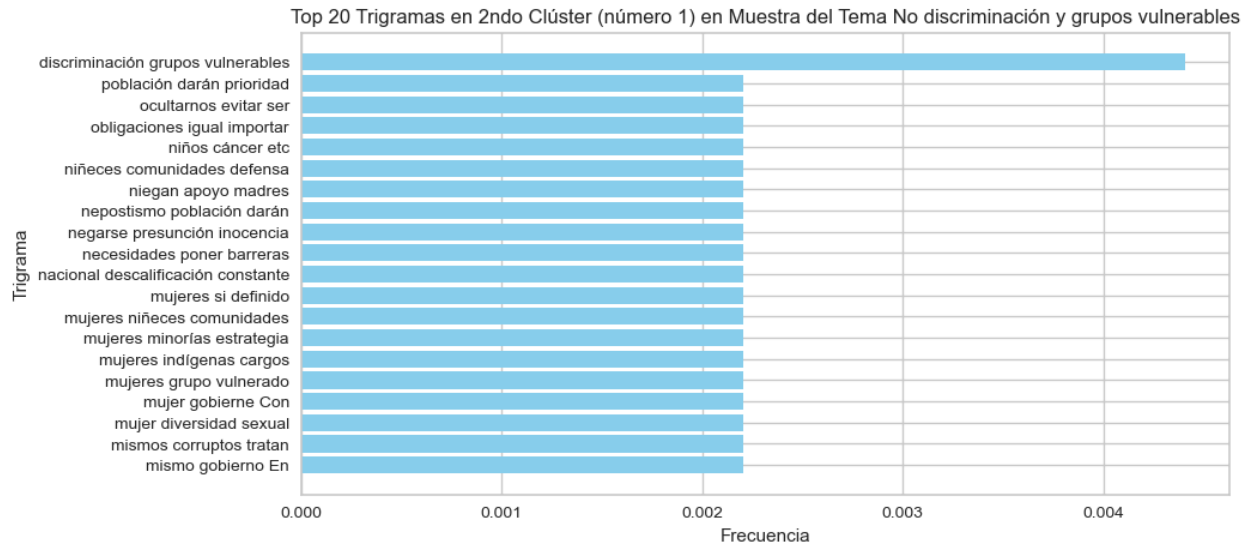
9. Top 20 términos en el segundo cluster del tema No Discriminación y Grupos Vulnerables



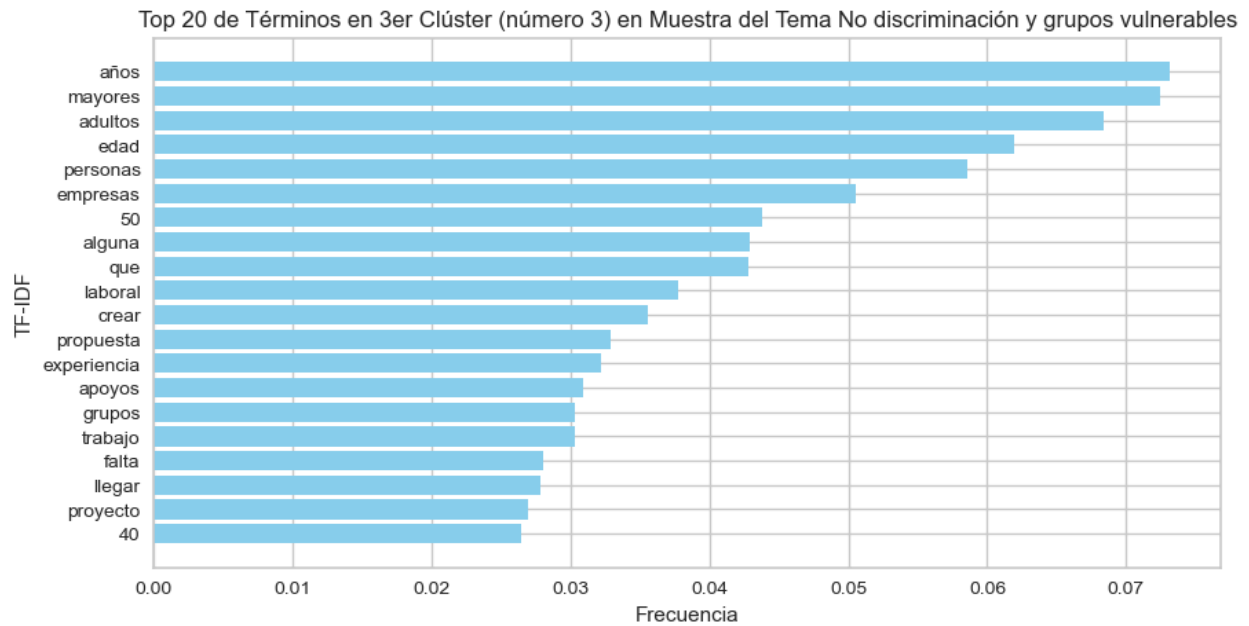
10. Top 20 bigramas en el segundo cluster del tema No Discriminación y Grupos Vulnerables



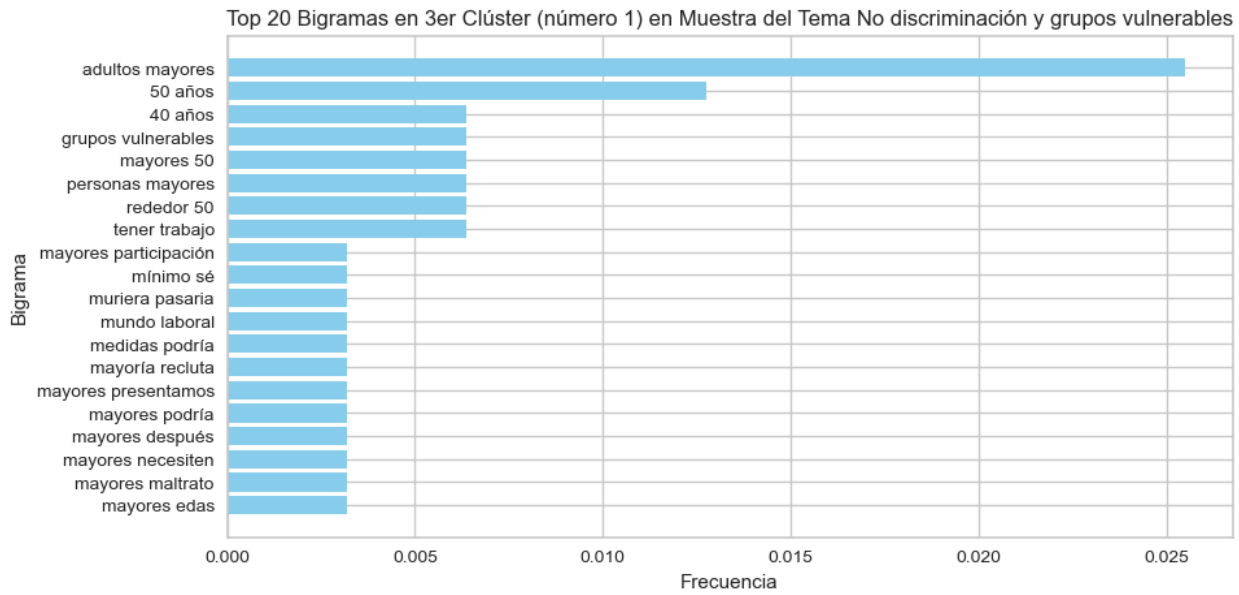
11. Top 20 trigramas en el segundo cluster del tema No Discriminación y Grupos Vulnerables



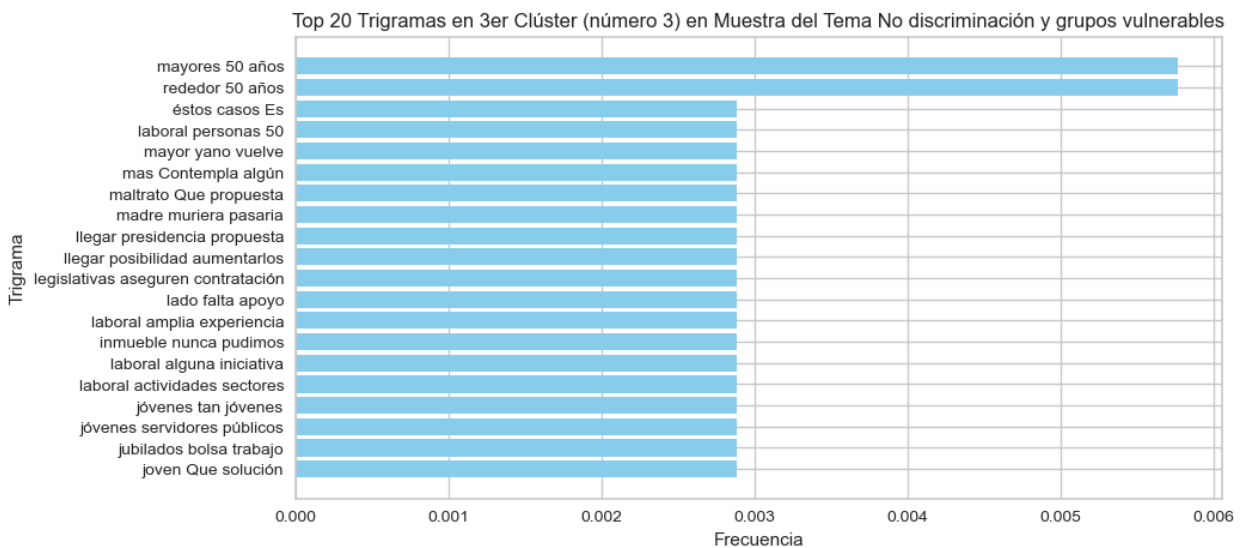
12. Top 20 términos en el tercer cluster del tema No Discriminación y Grupos Vulnerables



13. Top 20 bigramas en el tercer cluster del tema No Discriminación y Grupos Vulnerables

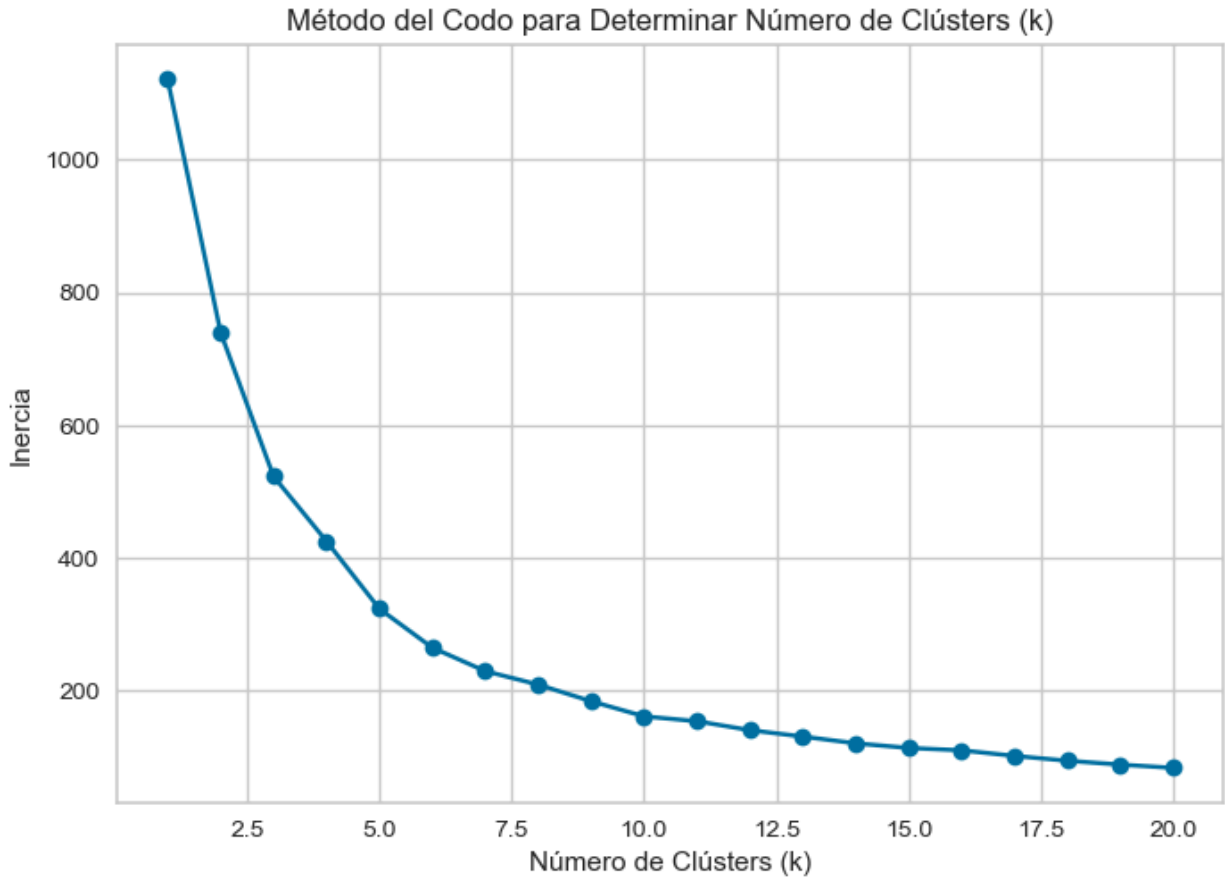


14. Top 20 trigramas en el tercer cluster del tema No Discriminación y Grupos Vulnerables

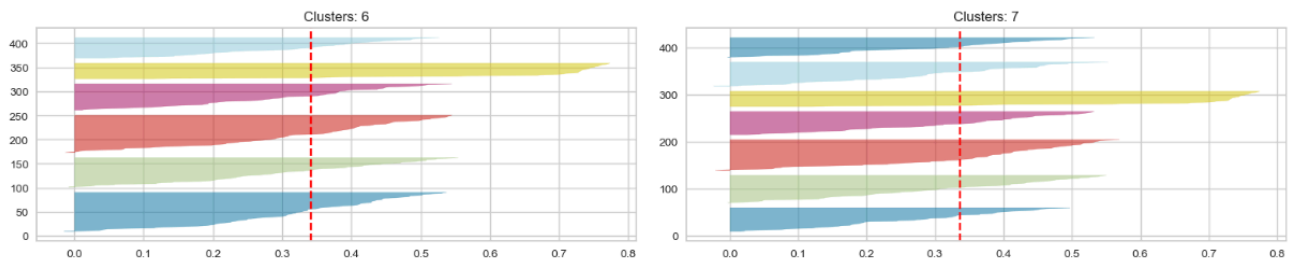


SALUD

1. Ejemplo de Método del Codo para determinar el número de clusters en el tema Salud

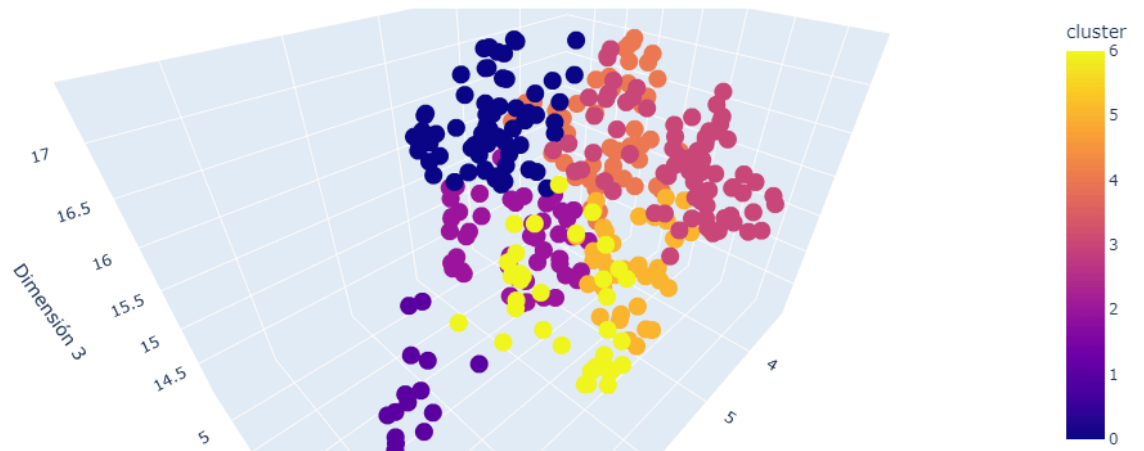


2. Ejemplo de Método de Silueta en el tema Salud

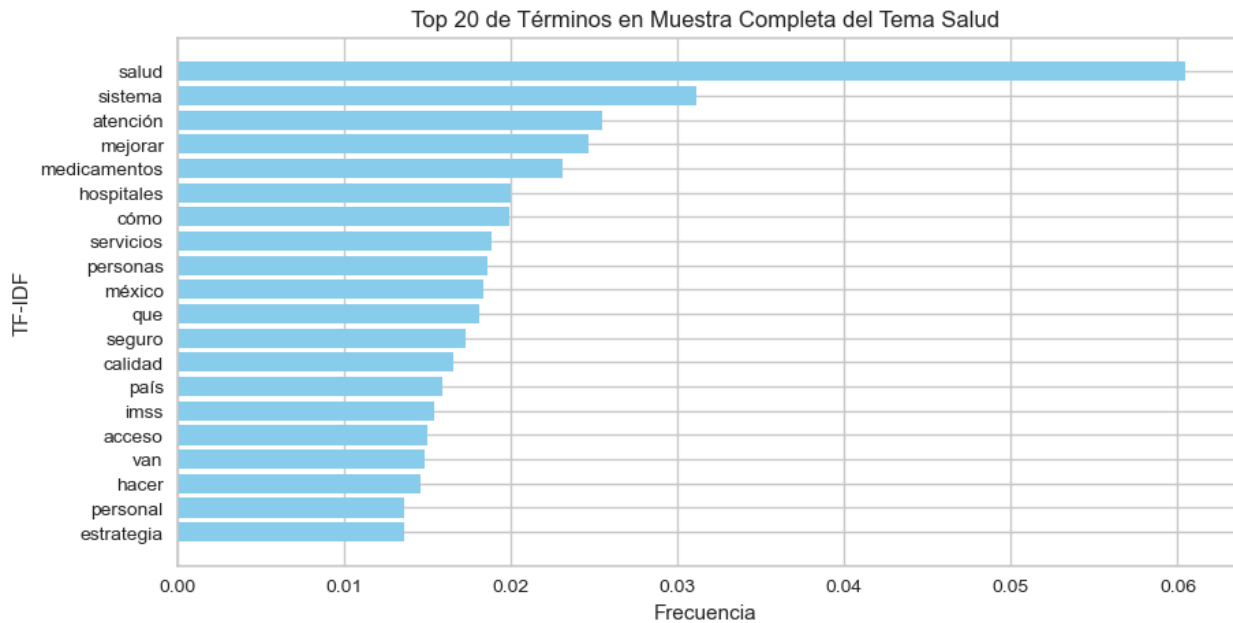


3. Visualización de clusters semánticos de Salud

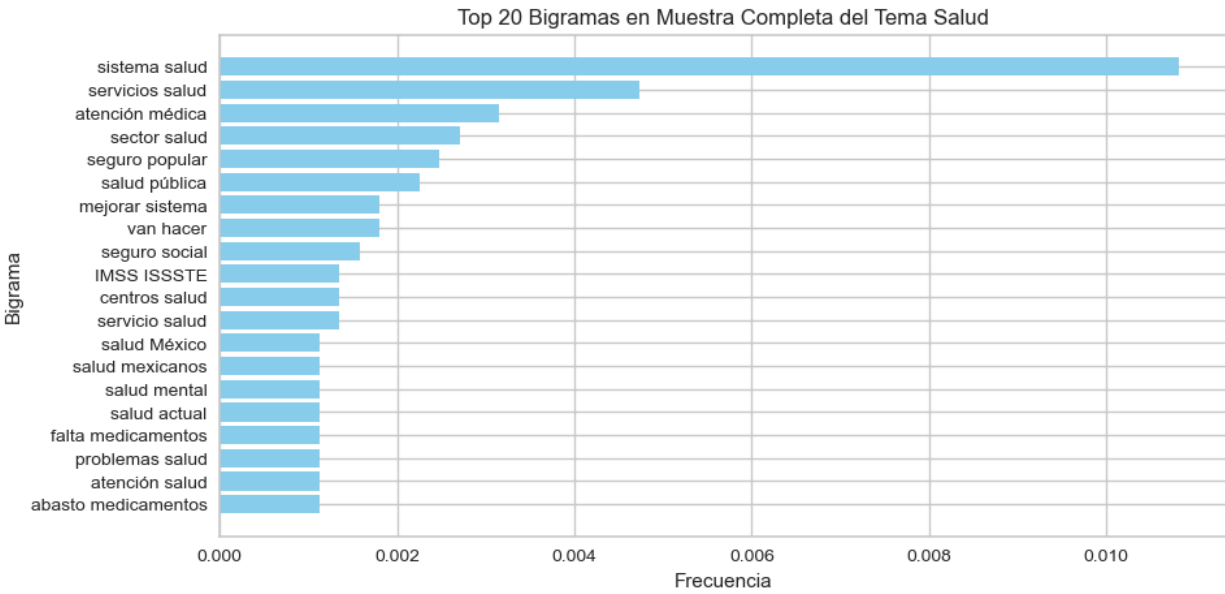
Visualización de clústers semánticos de Salud



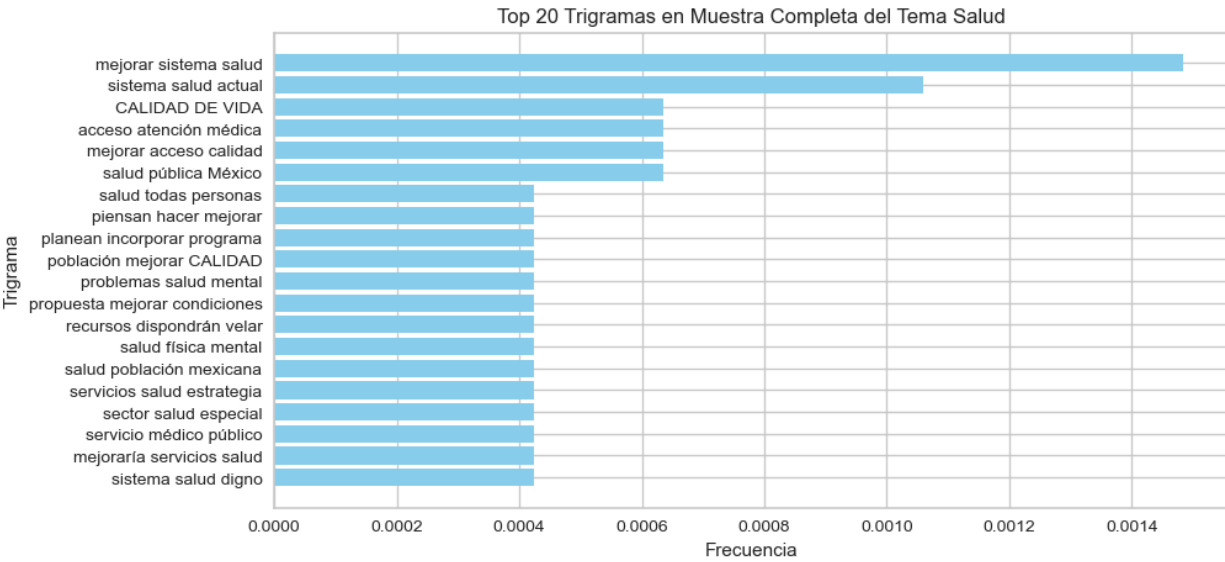
4. Top 20 términos en la muestra completa del tema Salud



5. Top 20 bigramas en la muestra completa del tema Salud



6. Top 20 trigramas en la muestra completa del tema Salud



7. Análisis semántico por clusters (conteo de preguntas por cluster)

Análisis Semántico por Clústers. Frecuencia de términos y relaciones entre palabras por clúster:

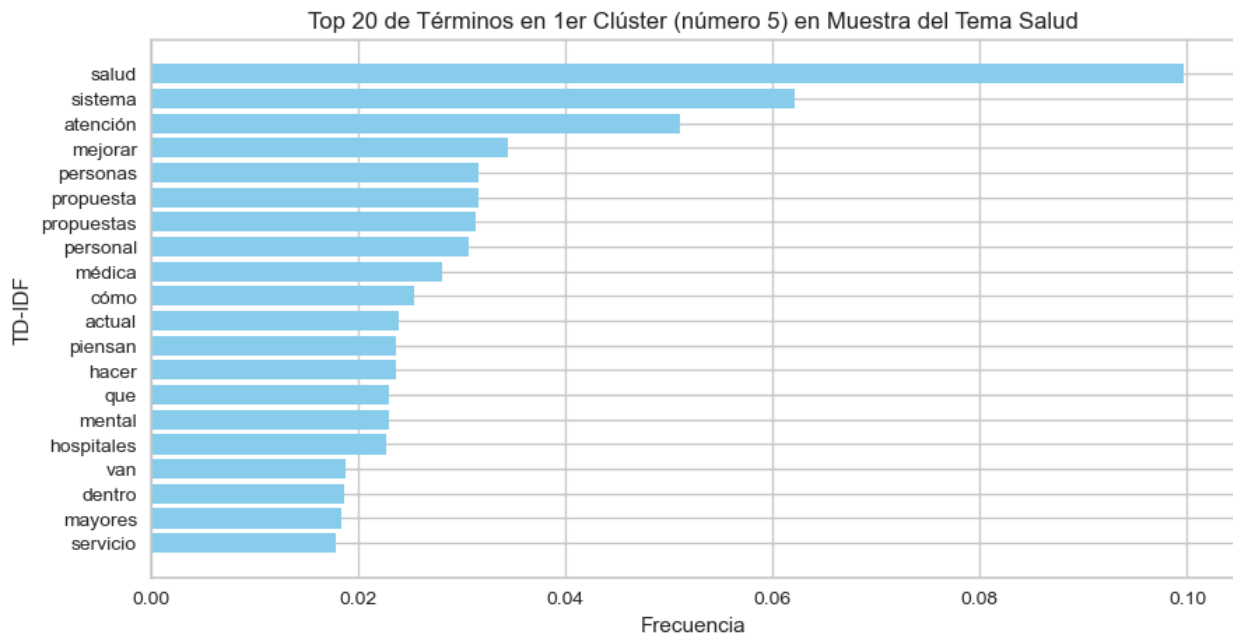
Identificar clústers más grandes en la muestra del tema (mayor frecuencia de preguntas por similitud semántica):

```
In [46]: # Identificar y enlistar clusters con mayor cantidad de preguntas (más frecuentes) del tema
for cluster, freq in dfClusterizado['cluster'].value_counts(ascending=False).head(10).items():
    print(f"Cluster {cluster} - {freq} preguntas")

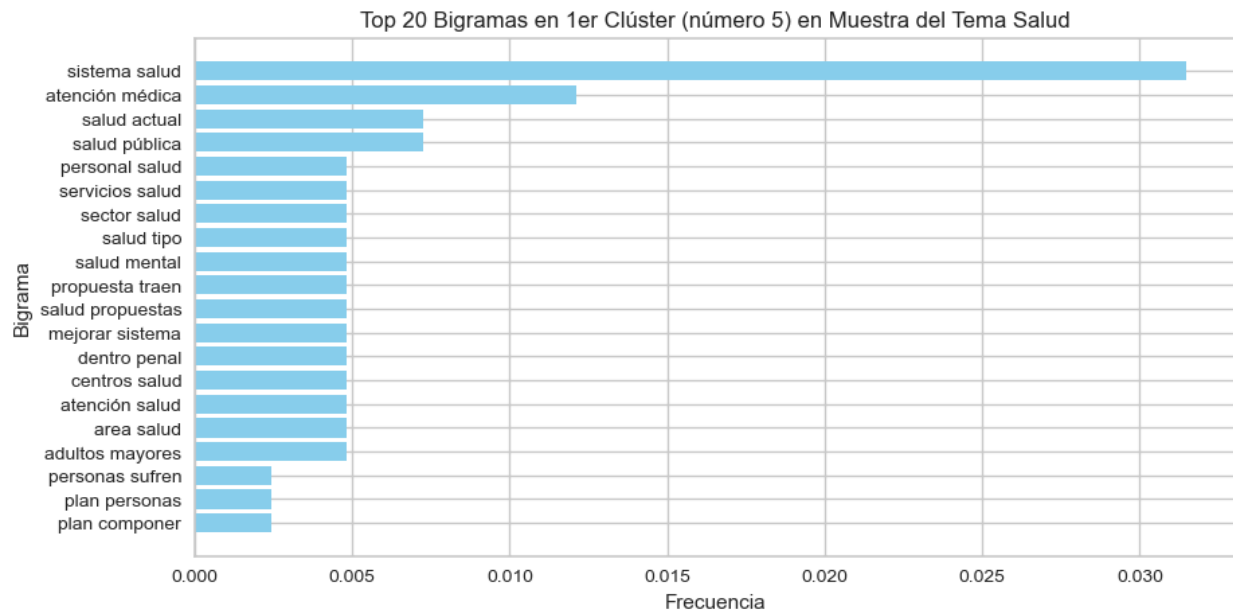
print()
```

```
Cluster 5 - 62 preguntas
Cluster 3 - 62 preguntas
Cluster 6 - 61 preguntas
Cluster 1 - 47 preguntas
Cluster 0 - 45 preguntas
Cluster 2 - 42 preguntas
Cluster 4 - 34 preguntas
```

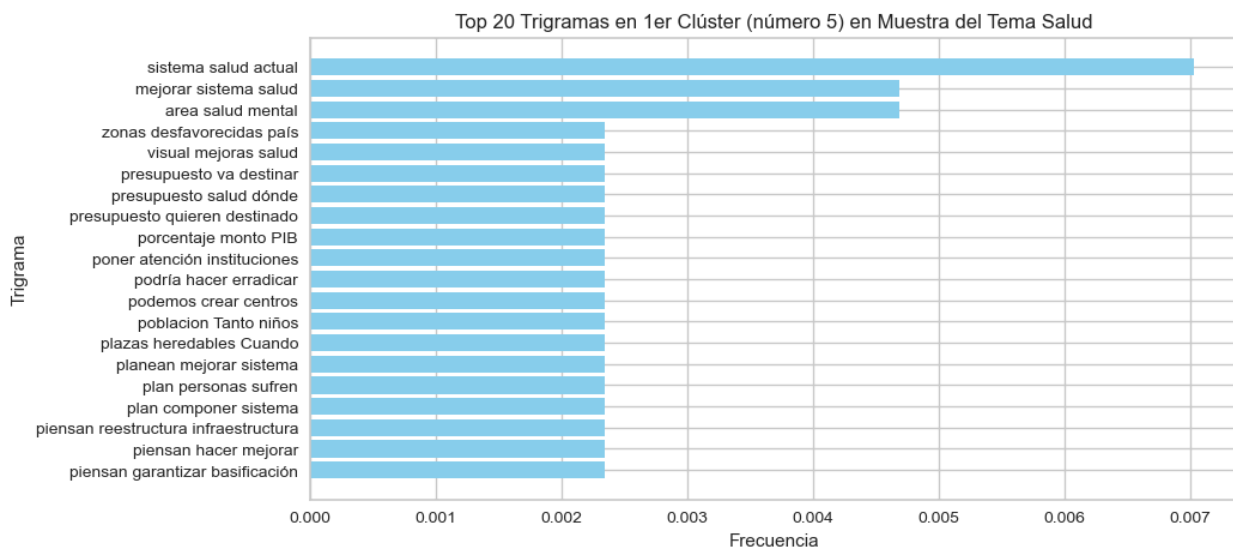
8. Top 20 términos en el primer cluster del tema Salud



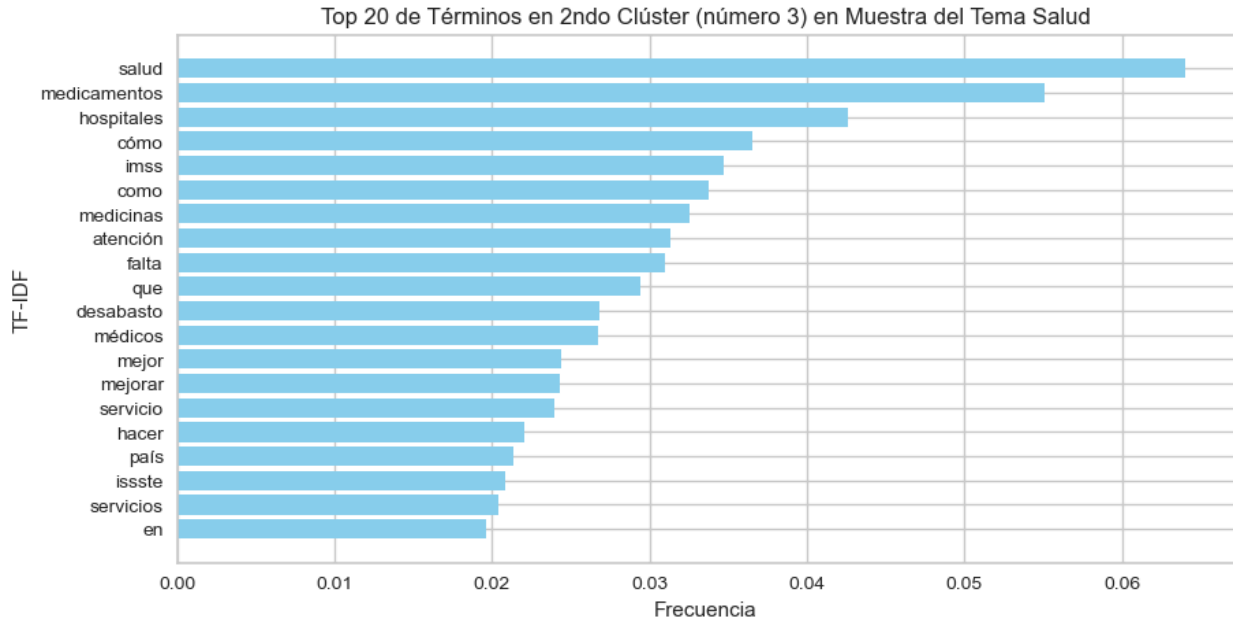
9. Top 20 bigramas en el primer cluster del tema Salud



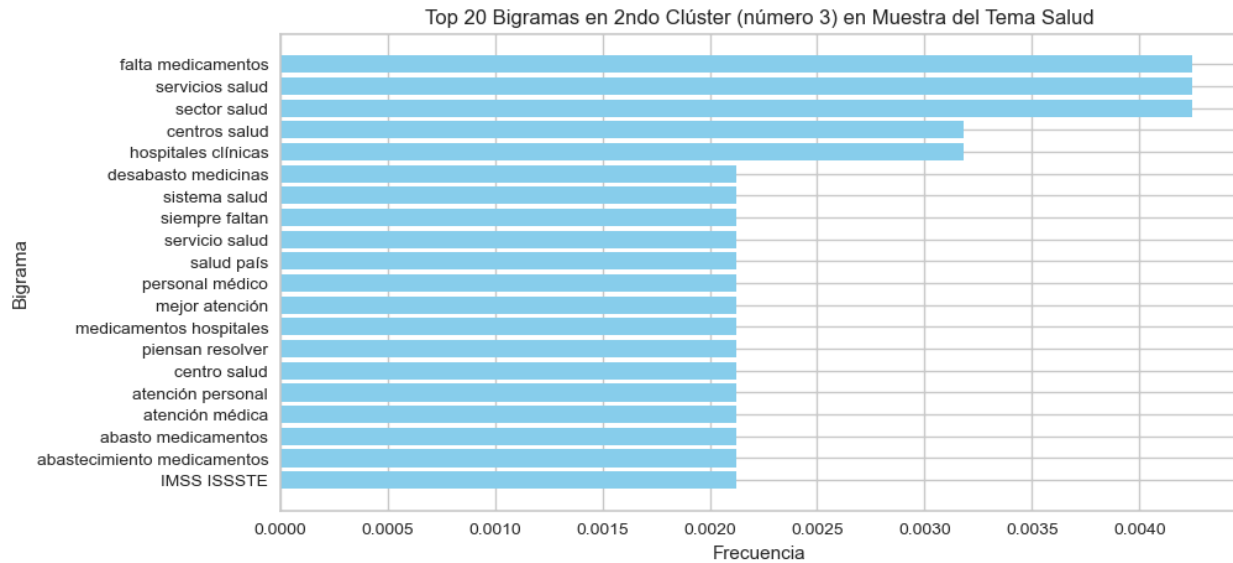
10. Top 20 trigramas en el primer cluster del tema Salud



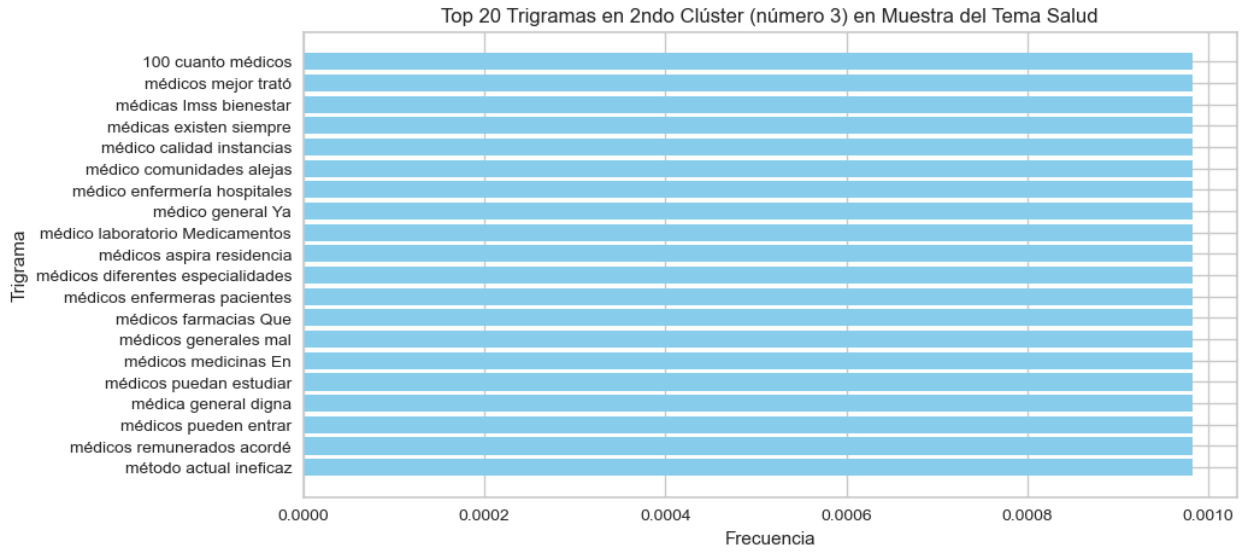
11. Top 20 términos en el segundo cluster del tema Salud



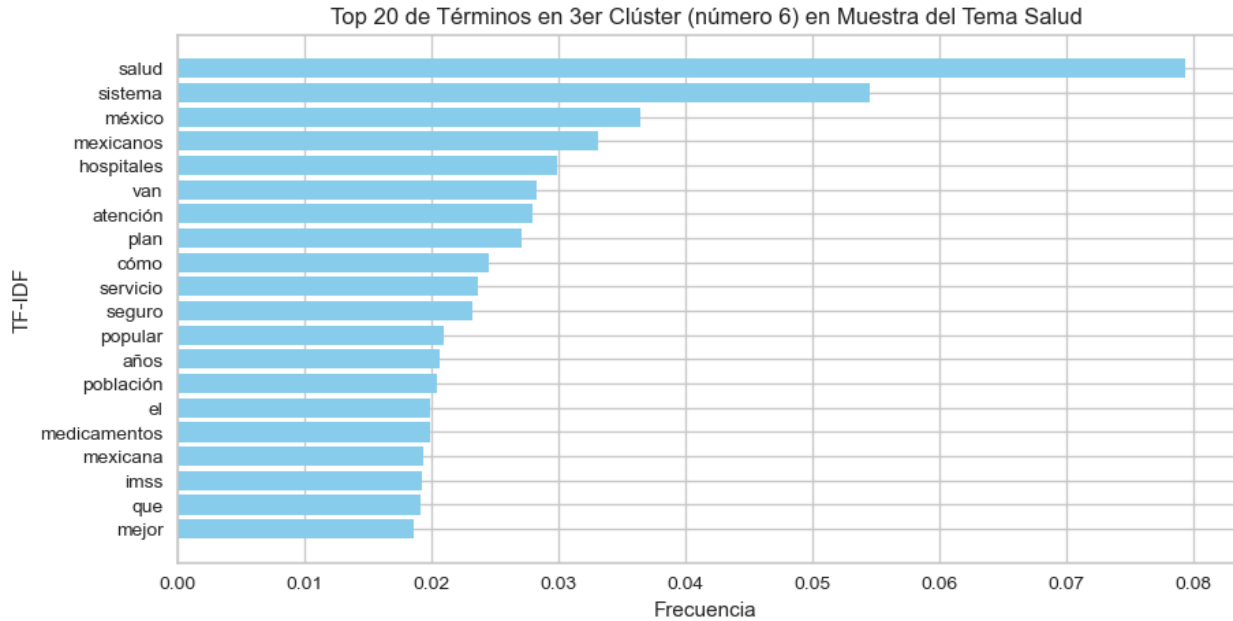
12. Top 20 bigramas en el segundo cluster del tema Salud



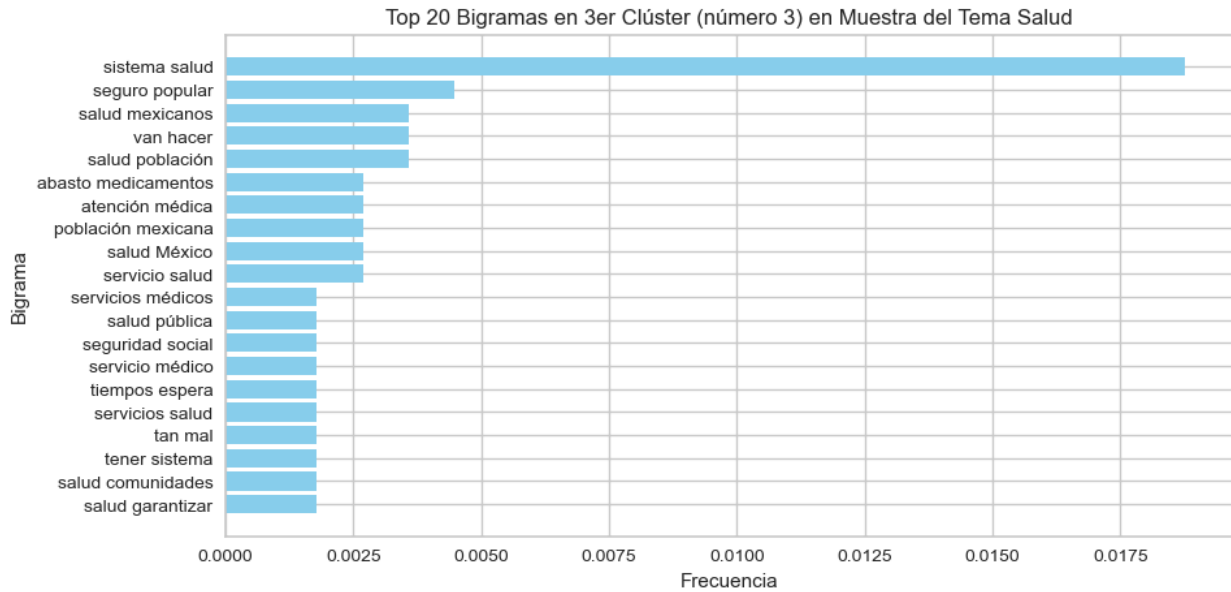
13. Top 20 trigramas en el segundo cluster del tema Salud



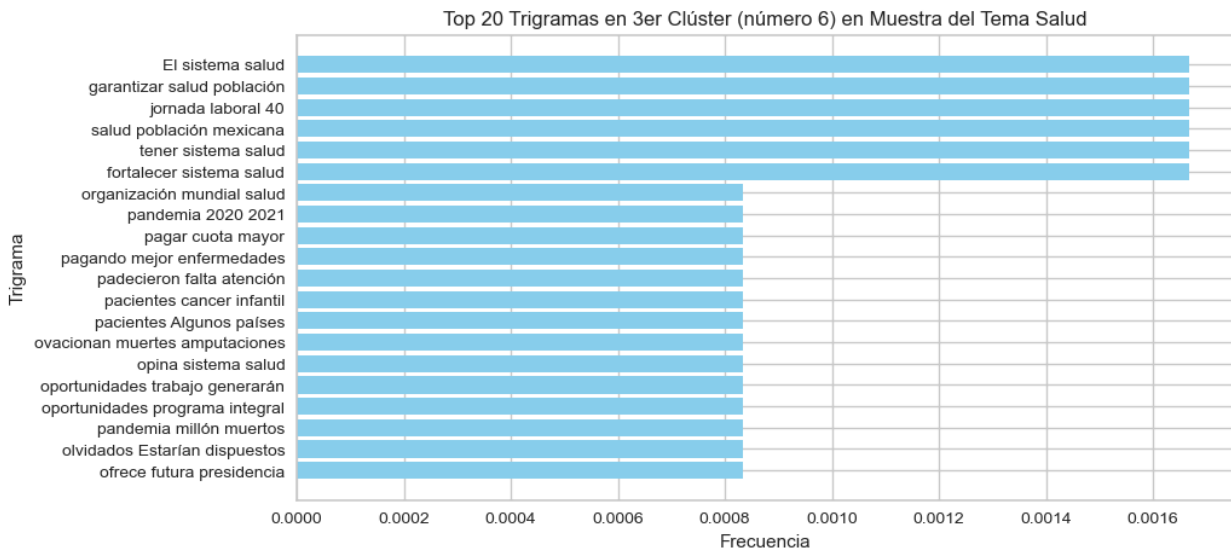
14. Top 20 términos en el tercer cluster del tema Salud



15. Top 20 bigramas en el tercer cluster del tema Salud



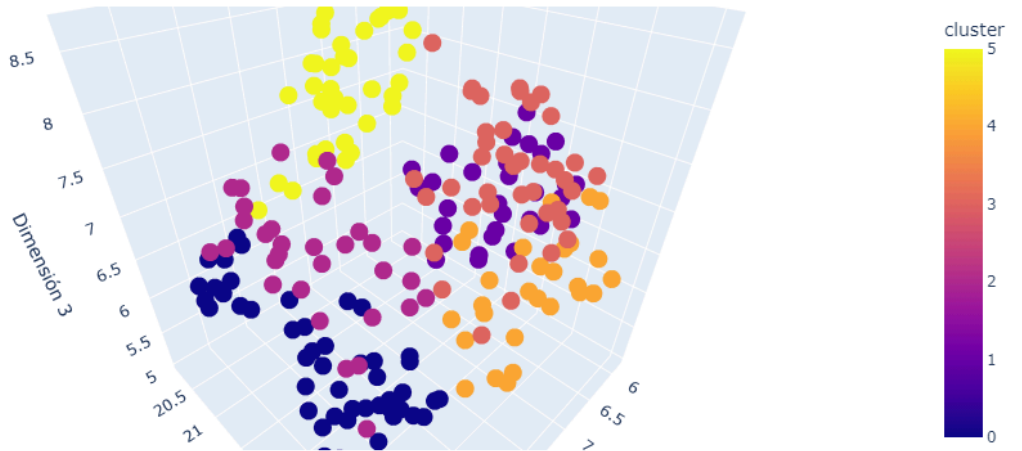
16. Top 20 trigramas en el tercer cluster del tema Salud



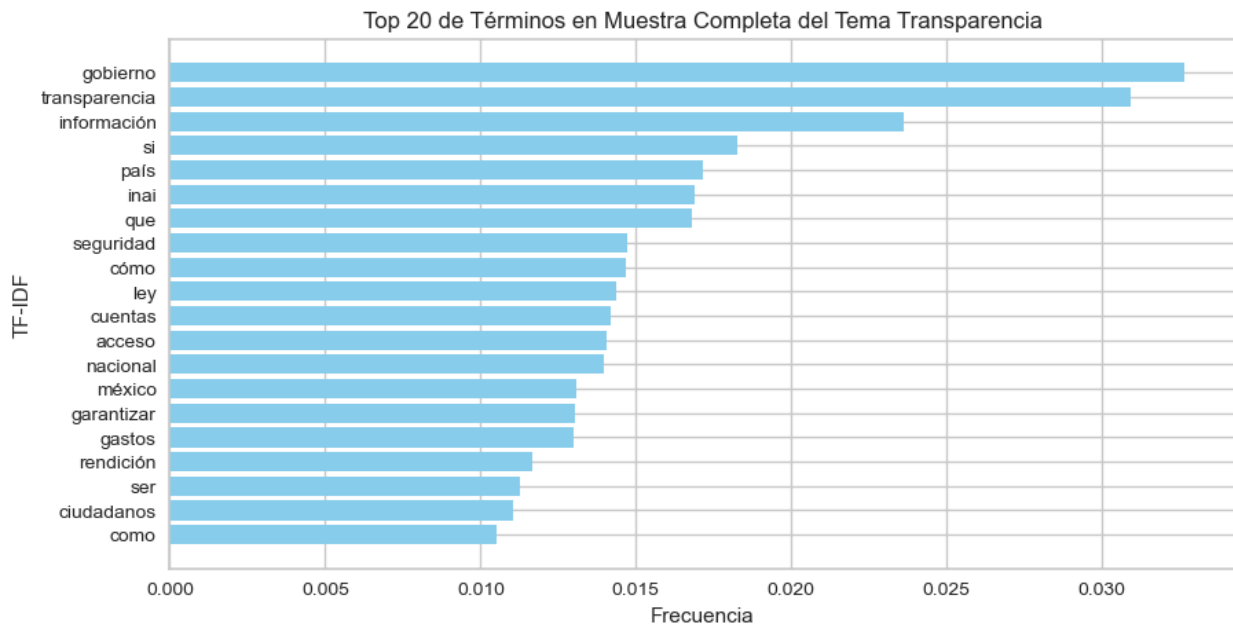
TRANSPARENCIA

1. Visualización de clusters semánticos de Transparencia

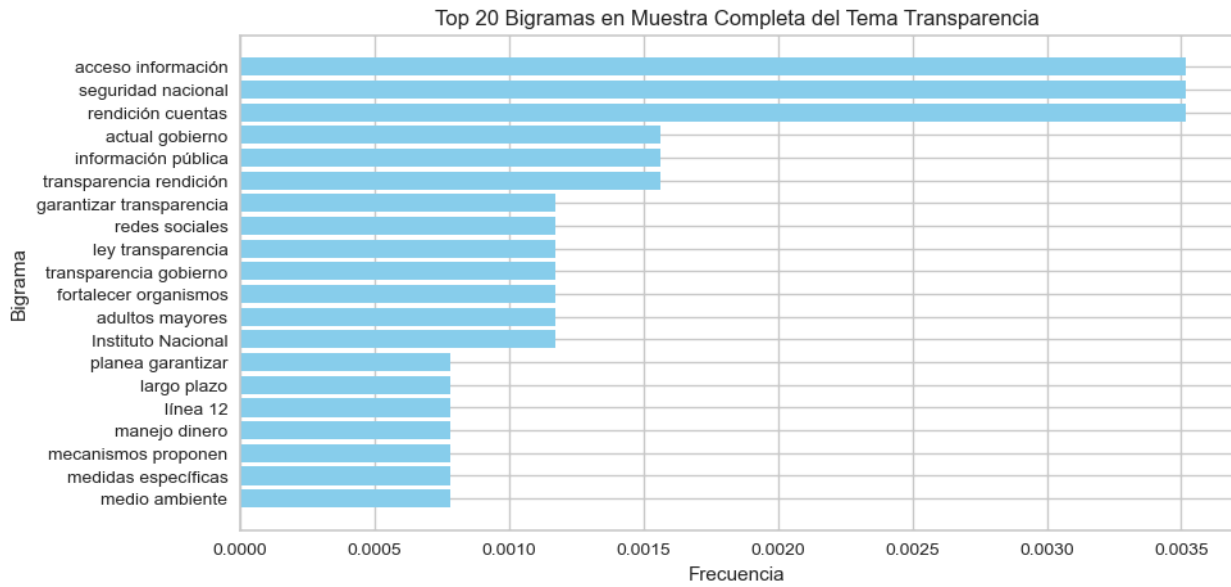
Visualización de clústers semánticos de Transparencia



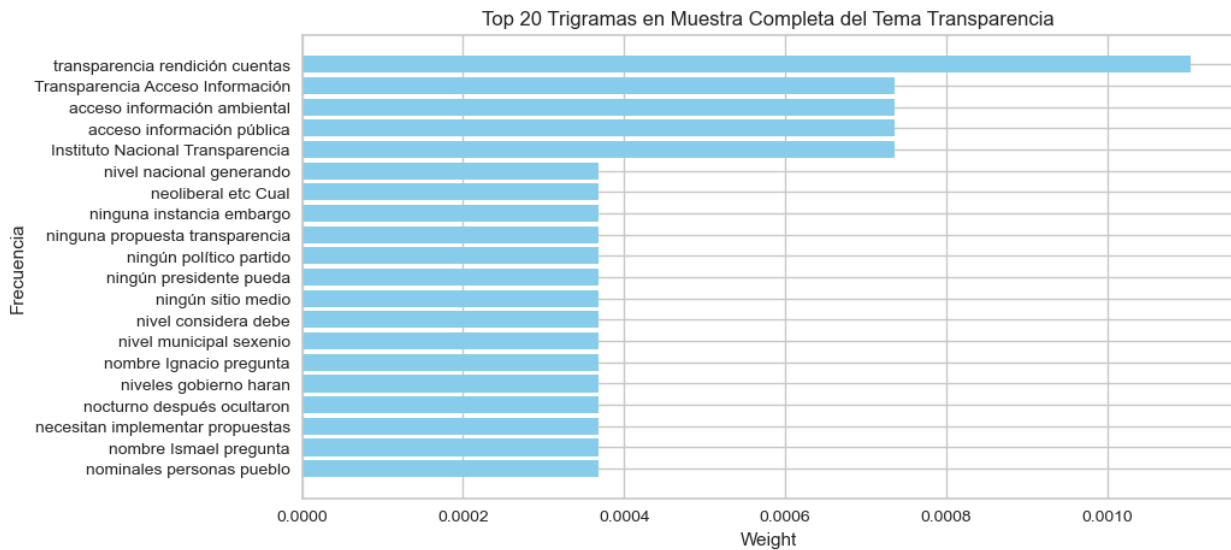
2. Top 20 términos en la muestra completa del tema Transparencia



3. Top 20 bigramas en la muestra completa del tema Transparencia



4. Top 20 trigramas en la muestra completa del tema Transparencia



5. Análisis semántico por clusters (conteo de preguntas por cluster)

Análisis Semántico por Clústers. Frecuencia de términos y relaciones entre palabras por clúster:

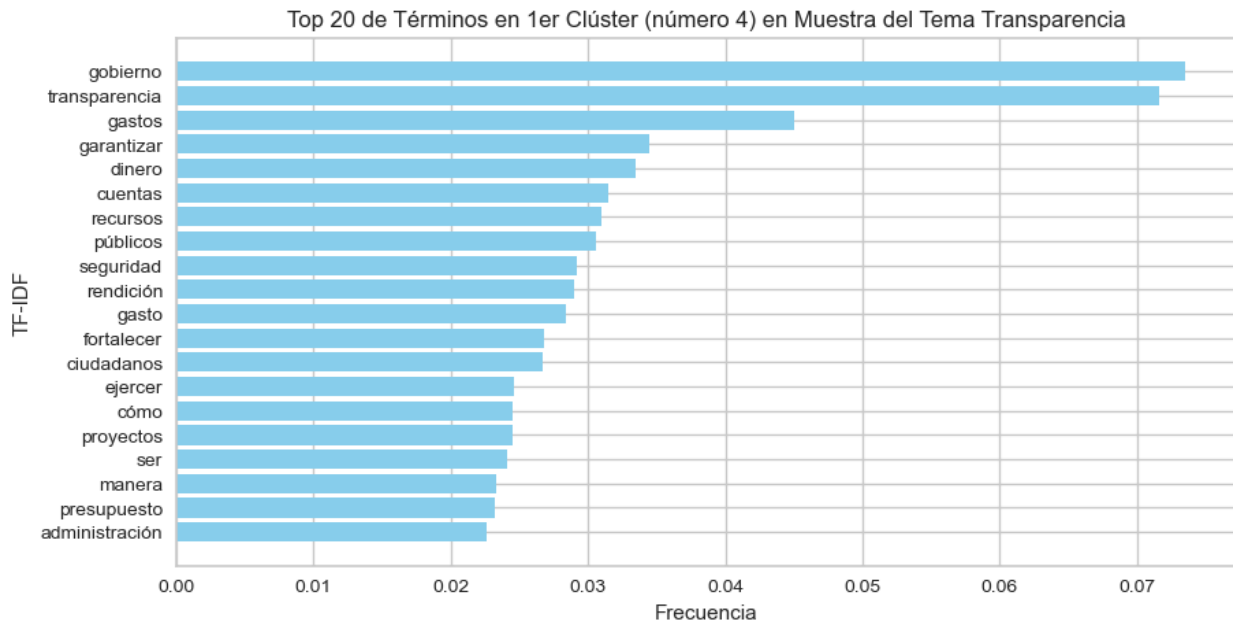
Identificar clústers más grandes en la muestra del tema (mayor frecuencia de preguntas por similitud semántica):

```
In [42]: # Identificar y enlistar clusters con mayor cantidad de preguntas (más frecuentes) del tema
for cluster, freq in dfClusterizado['cluster'].value_counts(ascending=False).head(10).items():
    print(f"Cluster {cluster} - {freq} preguntas")

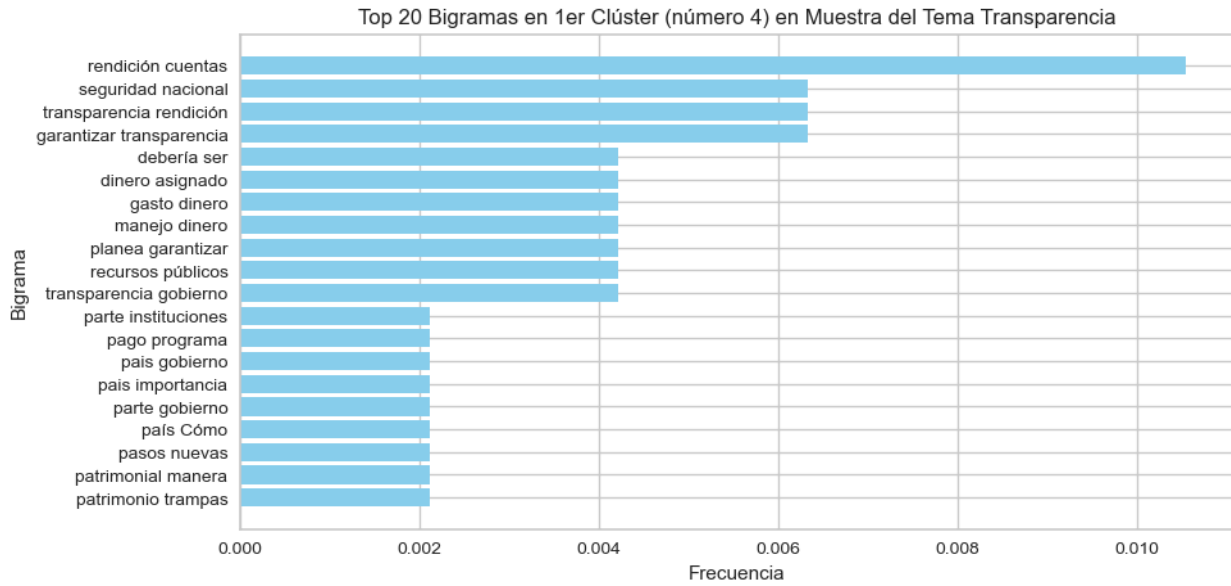
print()
```

```
Cluster 4 - 47 preguntas
Cluster 0 - 43 preguntas
Cluster 1 - 42 preguntas
Cluster 2 - 35 preguntas
Cluster 3 - 27 preguntas
Cluster 5 - 25 preguntas
```

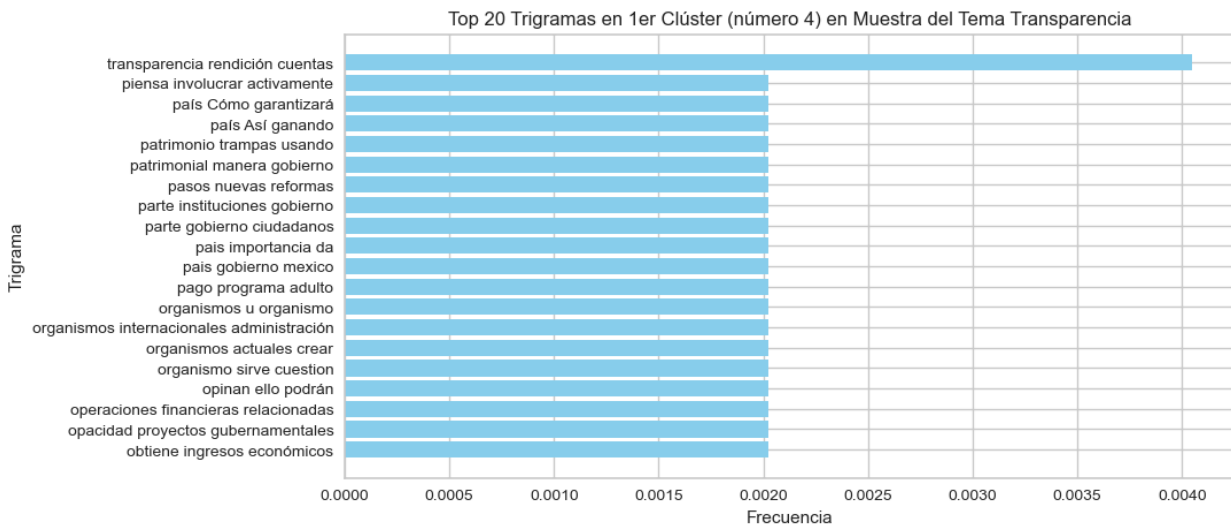
6. Top 20 términos en el primer cluster del tema Transparencia



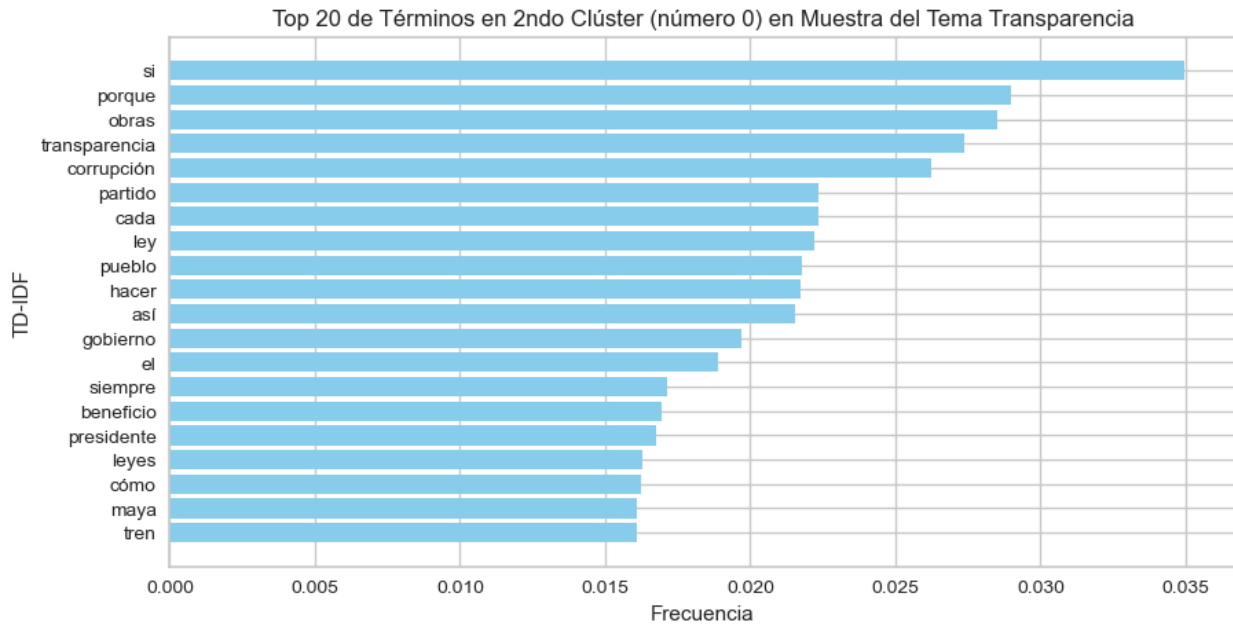
7. Top 20 bigramas en el primer cluster del tema Transparencia



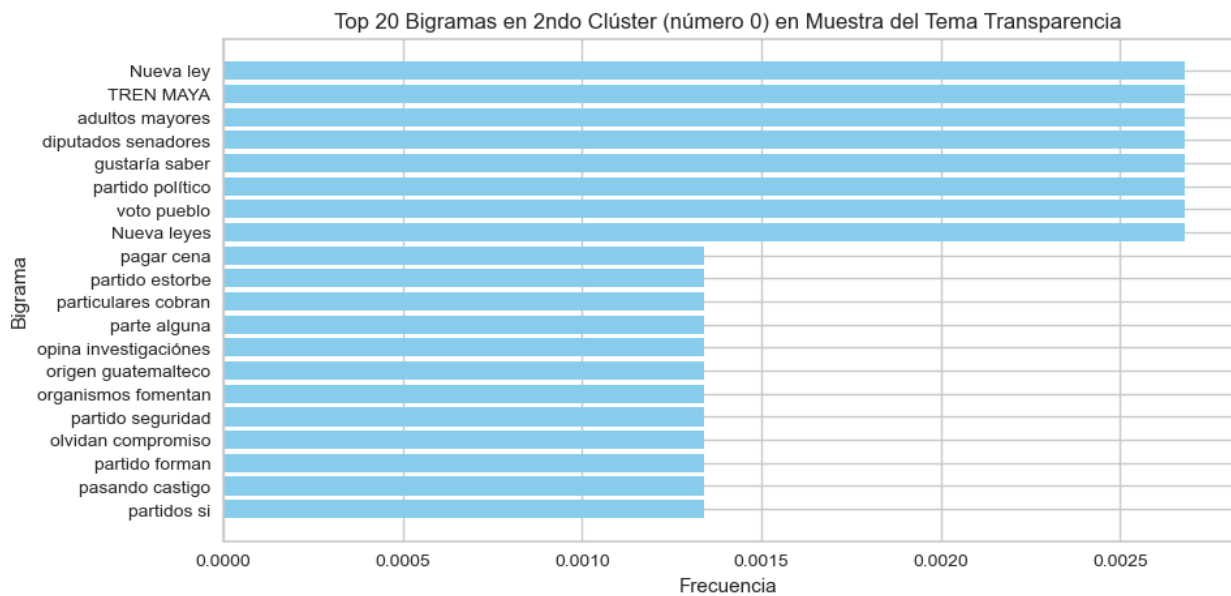
8. Top 20 trigramas en el primer cluster del tema Transparencia



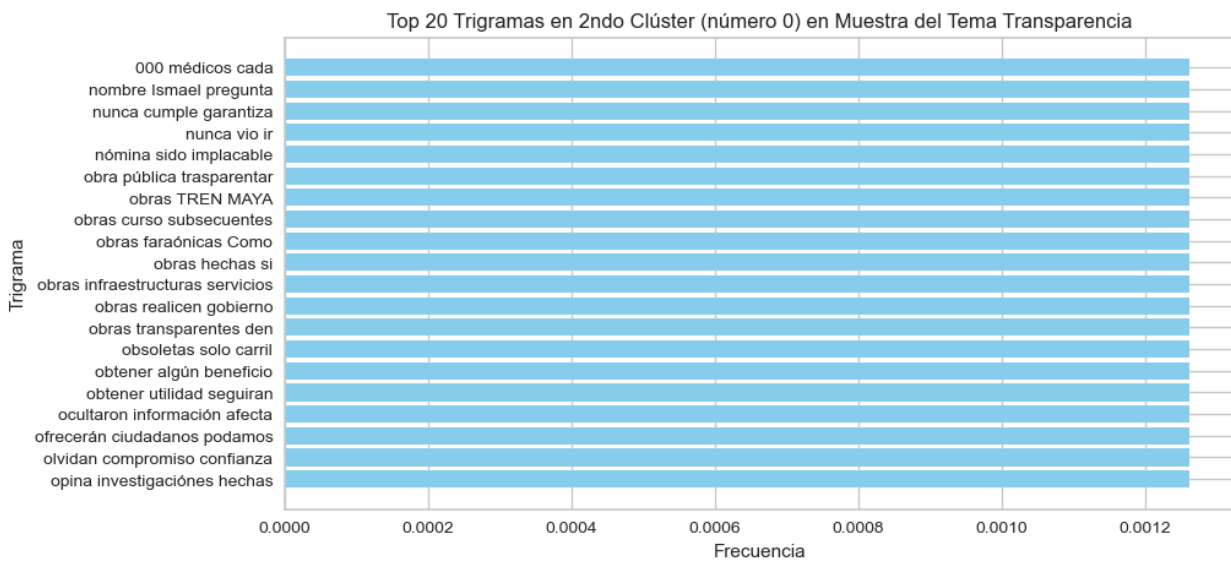
9. Top 20 términos en el segundo cluster del tema Transparencia



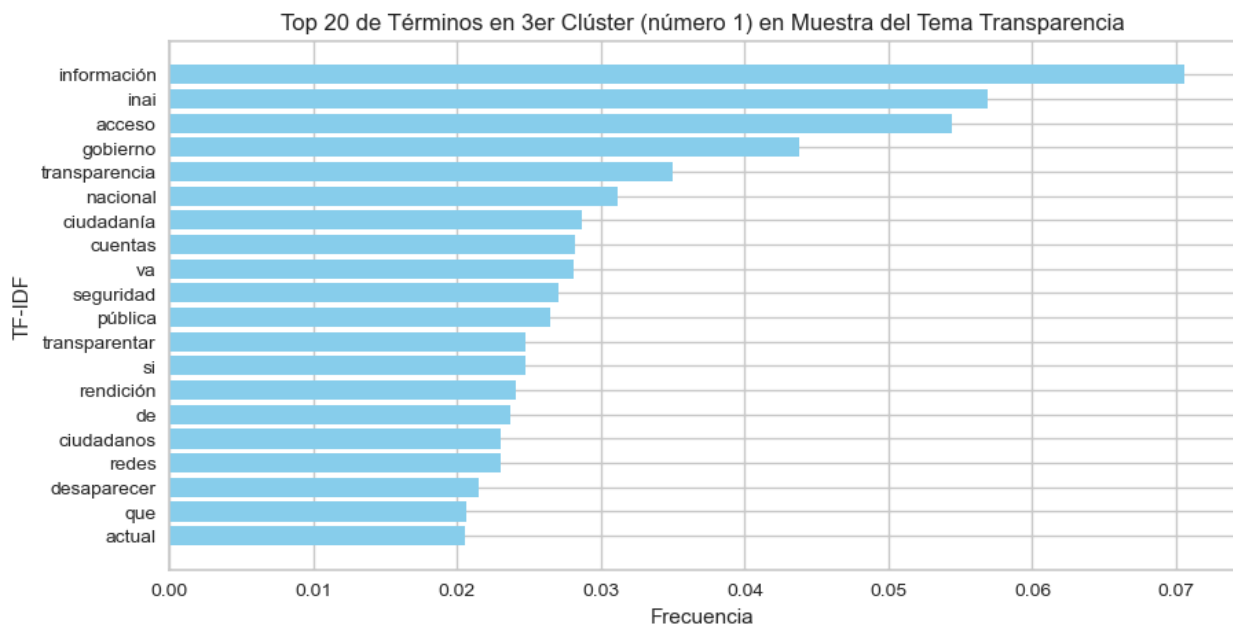
10. Top 20 bigramas en el segundo cluster del tema Transparencia



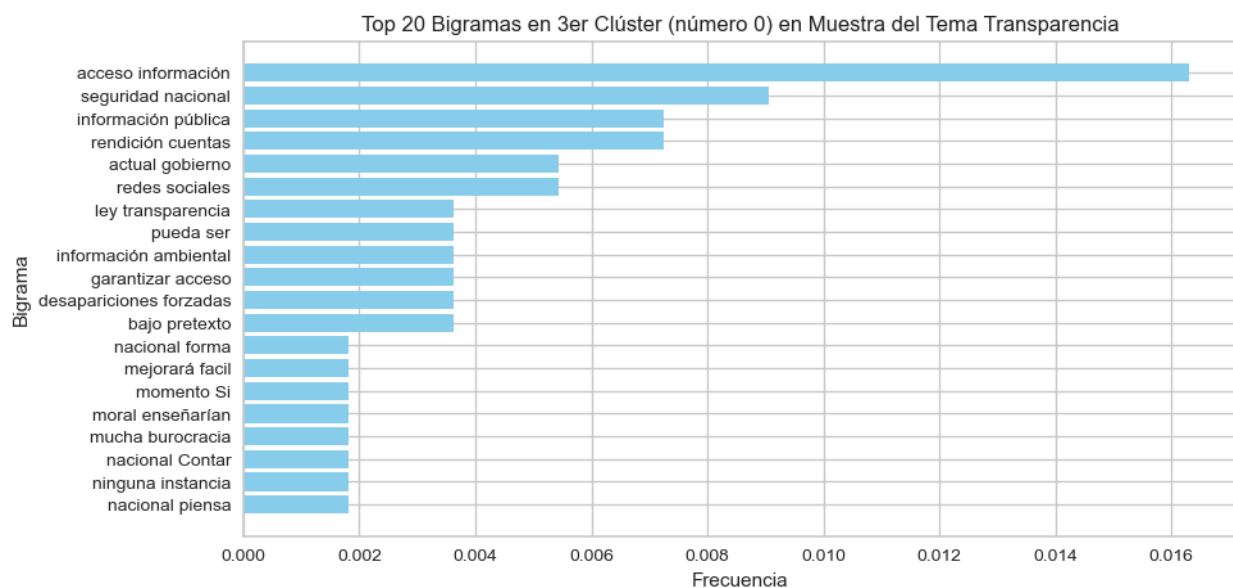
11. Top 20 trigramas en el segundo cluster del tema Transparencia



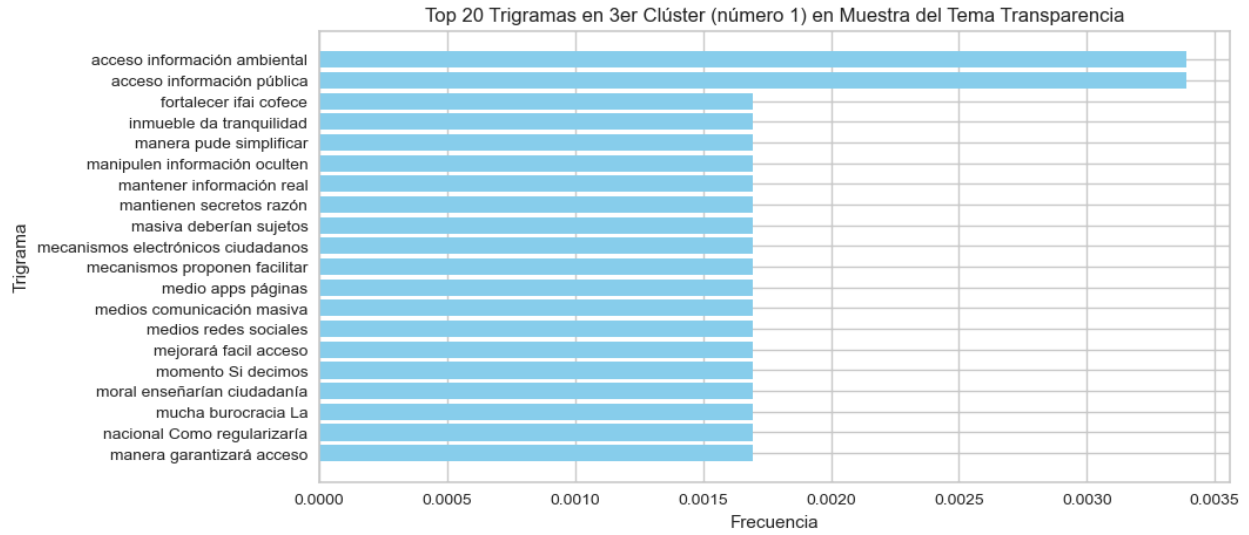
12. Top 20 términos en el tercer cluster del tema Transparencia



13. Top 20 bigramas en el tercer cluster del tema Transparencia



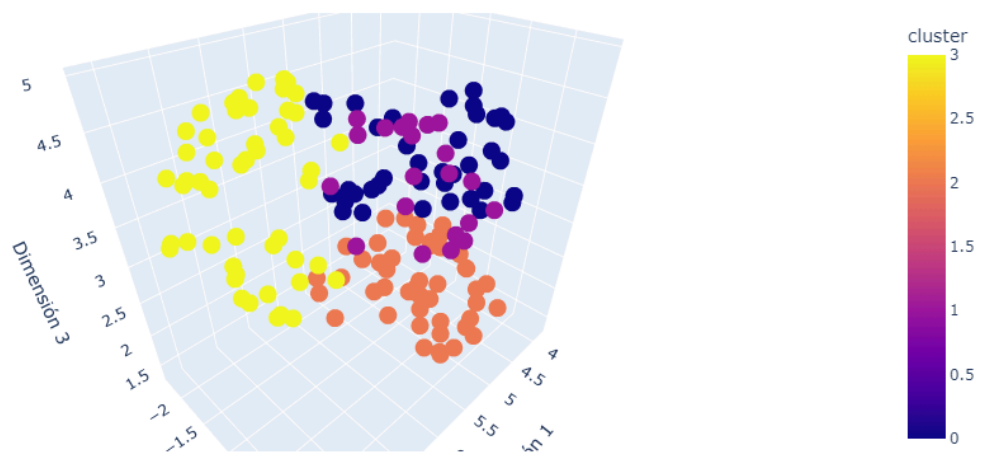
14. Top 20 trigramas en el tercer cluster del tema Transparencia



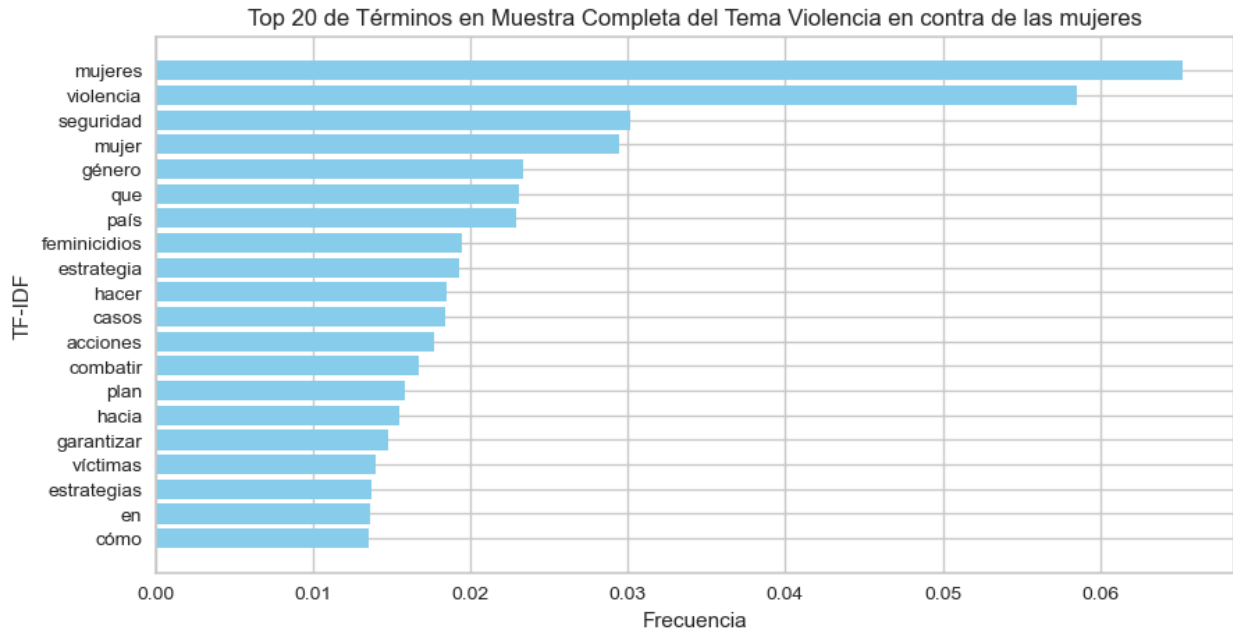
VIOLENCIA EN CONTRA DE LAS MUJERES

1. Visualización de clusters semánticos de Violencia en Contra de las Mujeres

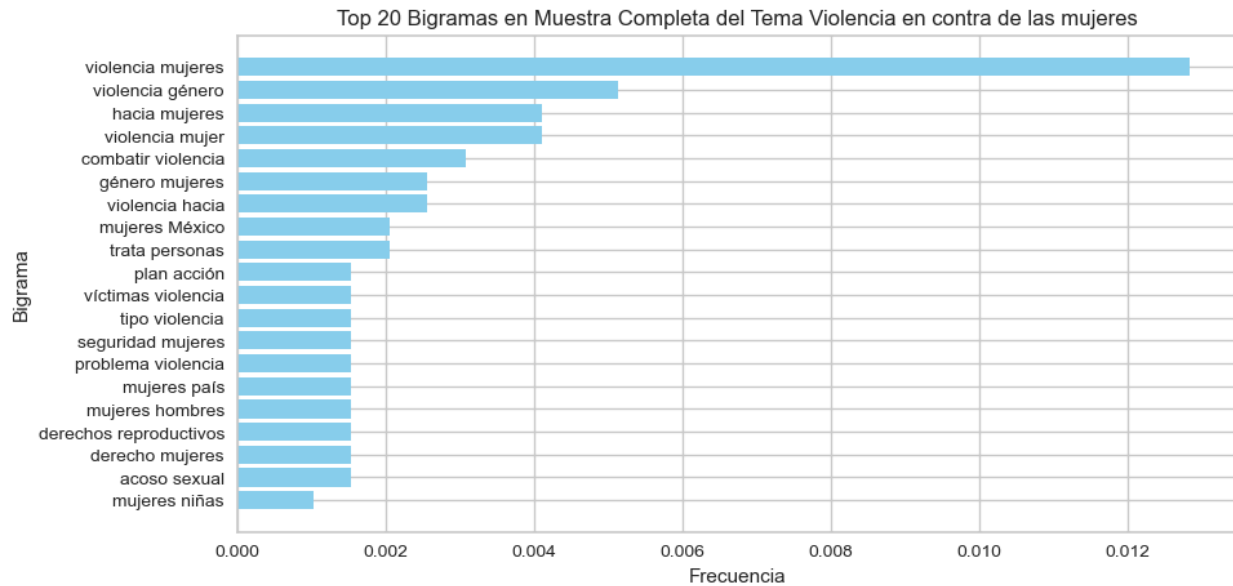
Visualización de clústers semánticos de Violencia en contra de las mujeres



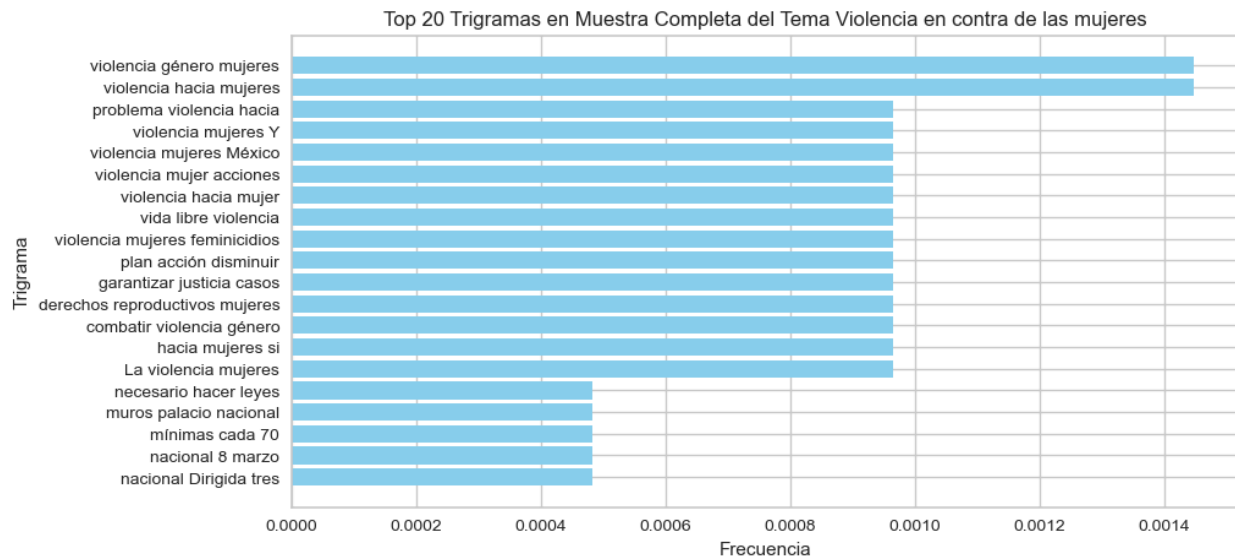
2. Top 20 términos en la muestra completa del tema Violencia en Contra de las Mujeres



3. Top 20 bigramas en la muestra completa del tema Violencia en Contra de las Mujeres



4. Top 20 trigramas en la muestra completa del tema Violencia en Contra de las Mujeres



5. Análisis semántico por clusters (conteo de preguntas por cluster)

Análisis Semántico por Clústers. Frecuencia de términos y relaciones entre palabras por clúster:

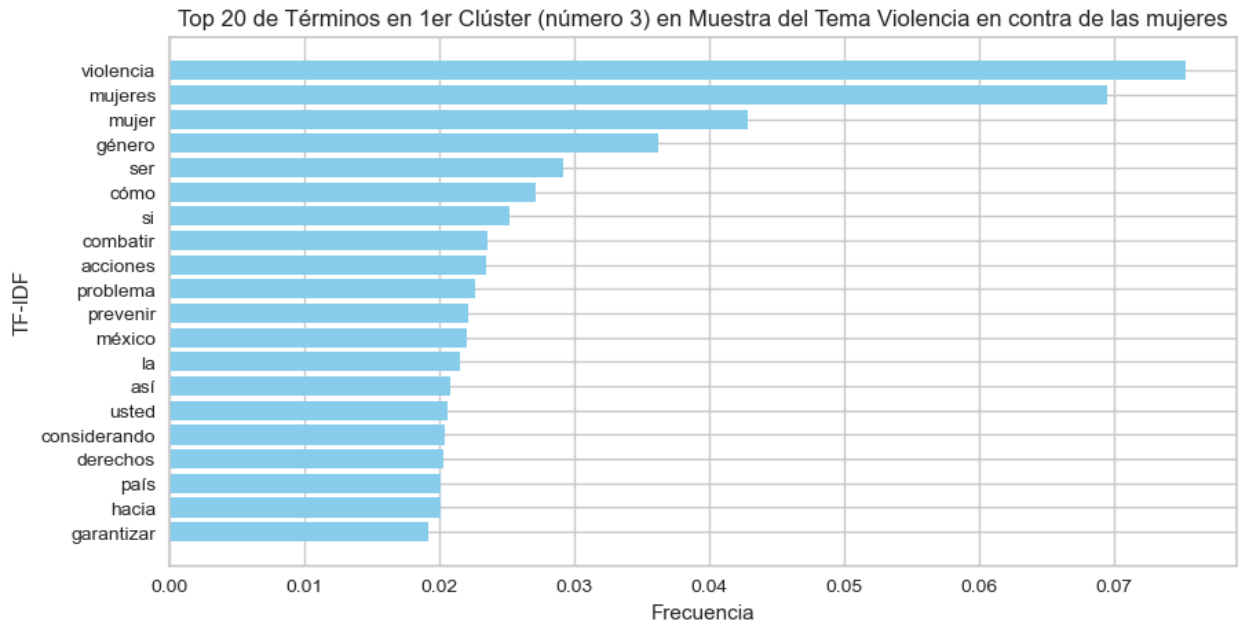
Identificar clústers más grandes en la muestra del tema (mayor frecuencia de preguntas por similitud semántica):

```
In [44]: # Identificar y enlistar clusters con mayor cantidad de preguntas (más frecuentes) del tema
for cluster, freq in dfClusterizado['cluster'].value_counts(ascending=False).head(10).items():
    print(f"Cluster {cluster} - {freq} preguntas")

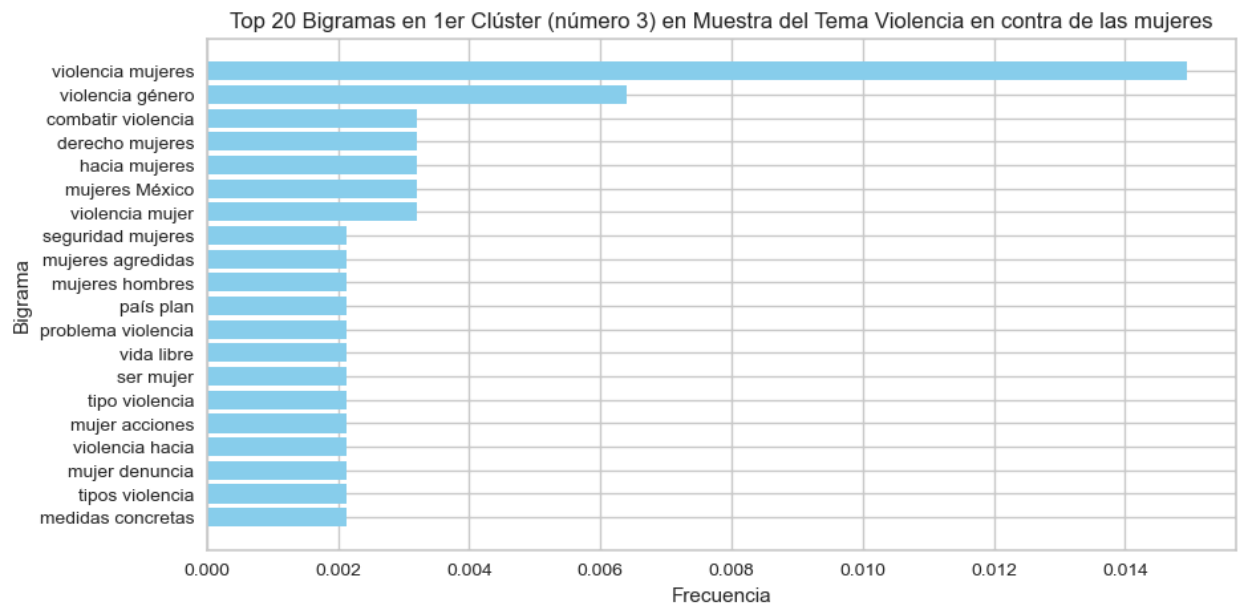
print()
```

```
Cluster 3 - 48 preguntas
Cluster 2 - 46 preguntas
Cluster 0 - 37 preguntas
Cluster 1 - 25 preguntas
```

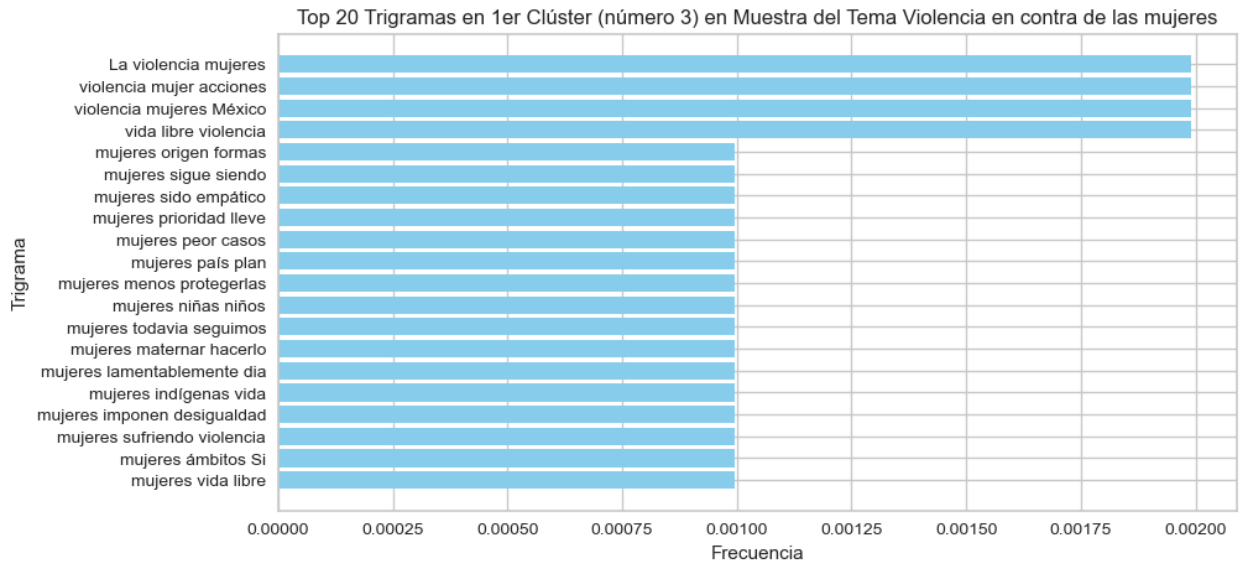
6. Top 20 términos en el primer cluster del tema Violencia en Contra de las Mujeres



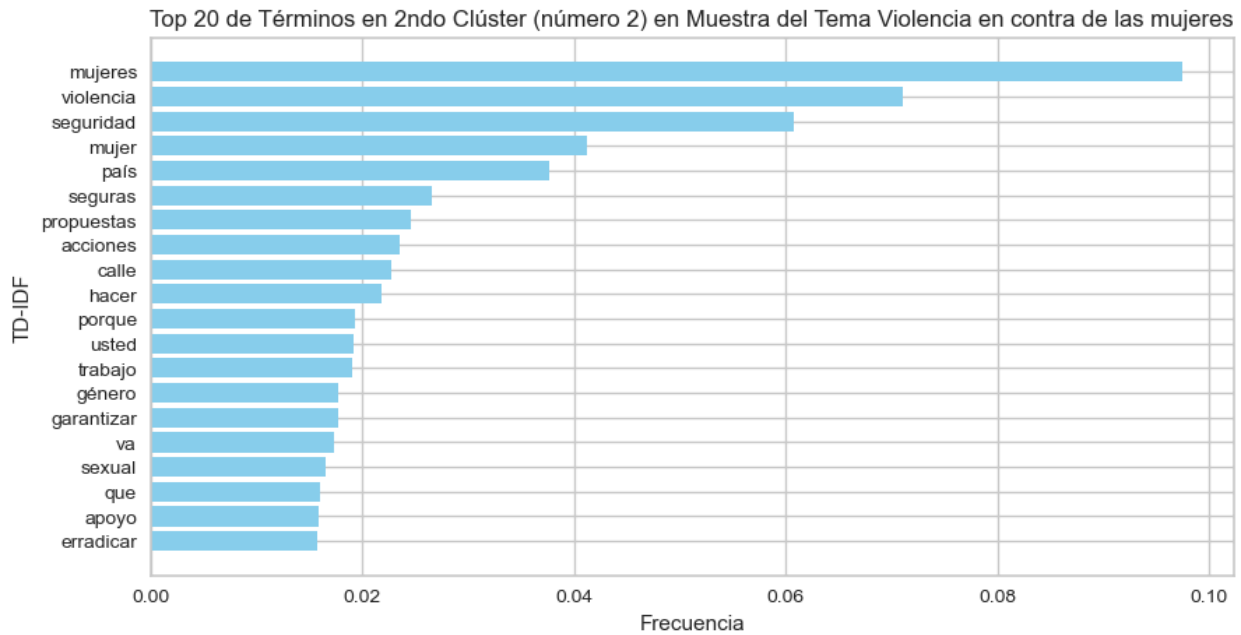
7. Top 20 bigramas en el primer cluster del tema Violencia en Contra de las Mujeres



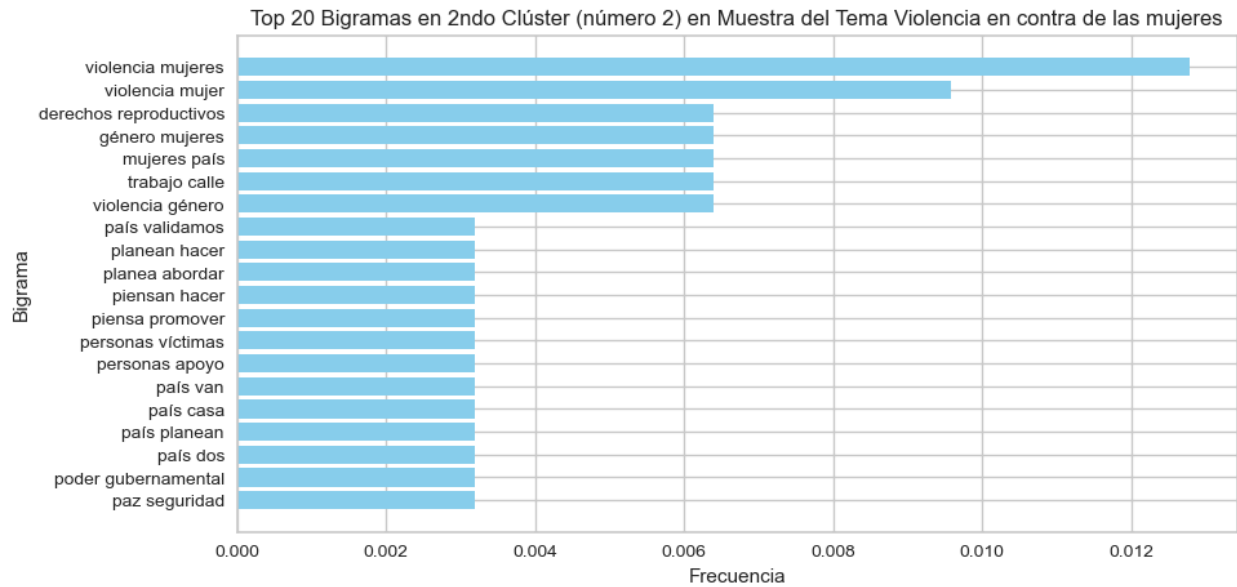
8. Top 20 trigramas en el primer cluster del tema Violencia en Contra de las Mujeres



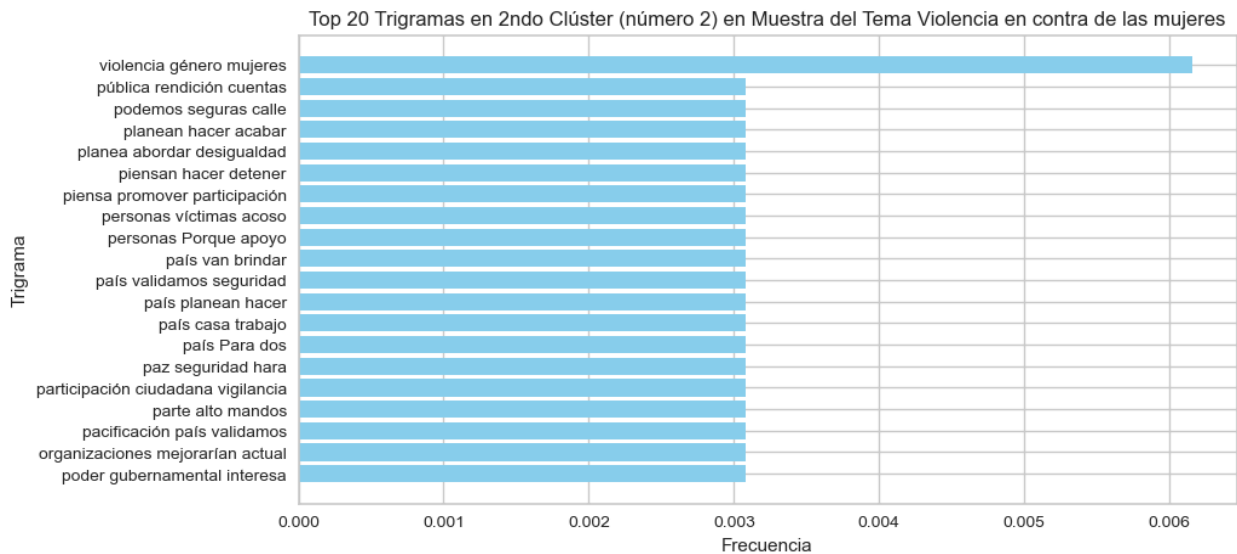
9. Top 20 términos en el segundo cluster del tema Violencia en Contra de las Mujeres



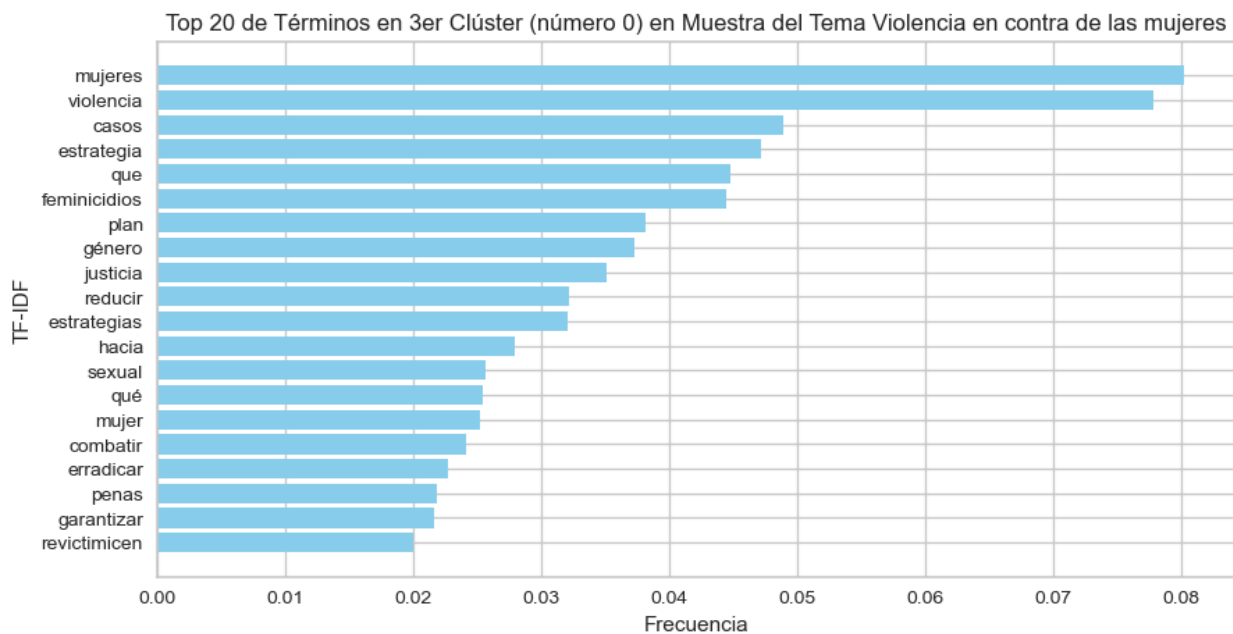
10. Top 20 bigramas en el segundo cluster del tema Violencia en Contra de las Mujeres



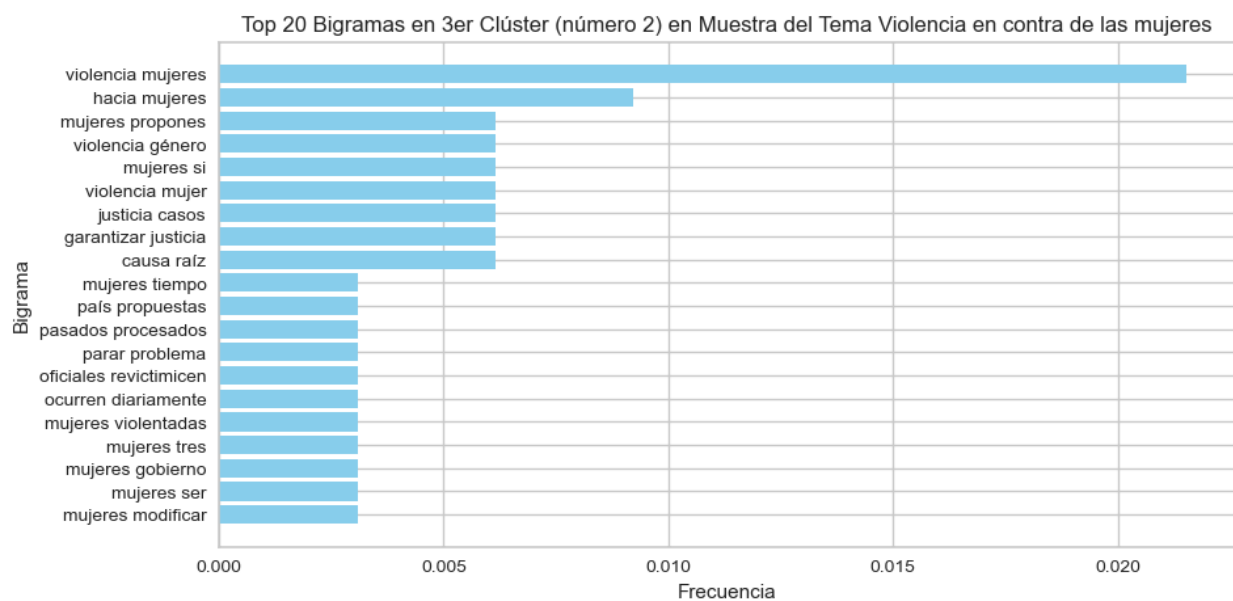
11. Top 20 trigramas en el segundo cluster del tema Violencia en Contra de las Mujeres



12. Top 20 términos en el tercer cluster del tema Violencia en Contra de las Mujeres



13. Top 20 bigramas en el tercer cluster del tema Violencia en Contra de las Mujeres



14. Top 20 trigramas en el tercer cluster del tema Violencia en Contra de las Mujeres

Top 20 Trigramas en 3er Clúster (número 0) en Muestra del Tema Violencia en contra de las mujeres

