**Bloque III: Mejora de precisión**

En este último bloque, se aborda el problema señalado en el estado del arte de los sistemas OpenIE relacionado con la extracción redundante de tripletas. El objetivo es eliminar aquellas tripletas redundantes o de baja confianza, con el fin de incrementar la precisión y reducir el ruido estructural.

## Estrategia 6: Filtrado de tripletas por confianza

La hipótesis de partida para esta estrategia era que las tripletas con puntuaciones de confianza bajas suelen corresponder a relaciones ambiguas, incorrectas o mal formadas. Su eliminación incrementaría la precisión global del grafo sin comprometer sustancialmente la cobertura.

Dado que el porcentaje de tripletas con una confianza inferior a 1 en el conjunto de referencia es muy bajo, se espera que el impacto de esta estrategia sea reducido. Es necesario analizar las métricas para confirmarlo.

**Análisis por texto**

[Poner imágenes B6U, B6E, B6H, B6I]

**Utilitarianism**: En el caso del Utilitarismo, como muestra la figura X, la eliminación de las tripletas con nivel de confianza inferior a 1 no ha supuesto ninguna mejora en las métricas del grafo. Esto indica que esas tripletas, además de ser precisas, contenían conceptos relevantes que también aparecen en el grafo de gold.

**Ellipse**: En el caso de la Elipse sucede lo mismo que en con el texto del Utilitarismo. Las coberturas semántica y léxica empeoran especialmente entre un X % y un X %, lo que indica que las tripletas eliminadas contenían entidades y relaciones relevantes que sí aparecen en el grafo gold. Aunque la precisión semántica mejora ligeramente, esta ganancia no compensa la pérdida en cobertura.

**Homeostasis**: En este caso, casi todas las métricas se mantienen constantes, a excepción de las precisiones, que mejoran muy ligeramente (menos de un X %). Esto puede explicarse porque la información contenida en las tripletas eliminadas también está presente en otras tripletas. Al retirarlas, la precisión aumenta un poco, pero la calidad general del grafo permanece prácticamente igual.

**Inflation**: En el caso del texto sobre inflación, la Estrategia 6 consigue una ligera mejora en la precisión léxica, lo que sugiere que elimina variantes de expresión alejadas del grafo gold. Sin embargo, el resto de las métricas o bien se mantienen constantes o empeoran.

En conjunto, el filtrado de tripletas por nivel de confianza no ejerce un impacto significativo en los textos de referencia. Las escasas mejoras que aporta resultan poco consistentes.

**Evolución media de métricas**

[Poner imagen L6]

La figura X confirma lo anterior: la precisión semántica y léxica apenas suben un 1 %, mientras que el resto de métricas retroceden, lo que indica que los pequeños avances en precisión no logran equilibrar las pérdidas en cobertura y coherencia estructural.

## Estrategia 7: Eliminación de tripletas similares

La hipótesis de la Estrategia 7 plantea que depurar las tripletas redundantes incrementará la precisión del grafo al evitar la sobre-representación de las mismas relaciones. Para ello se sigue el proceso descrito en la metodología, que no solo elimina las redundancias, sino que también enriquece relaciones para hacerlas más completas.

**Análisis por texto**

[Poner imágenes B7U, B7E, B7H, B6I]

**Utilitarianism**: En el caso del Utilitarismo, la aplicación de la Estrategia 7 produce un incremento notable de la precisión. Laprecisión semántica sube un **X %** y la léxica un **Y %**. Asimismo, la similitud estructural con el grafo gold mejora en **Z %** a nivel semántico y en **W %** a nivel léxico. Que la cobertura permanezca constante indica que no se han eliminado nodos ni aristas relevantes.

**Ellipse:** En el caso de Elipse, aunque aumentan tanto las precisiones como los coeficientes de Jaccard, se aprecia un claro empeoramiento de la cobertura semántica. Esto indica que, durante la depuración de tripletas, se han eliminado algunas relaciones que aportaban un significado importante y que sí estaban presentes en el grafo gold.

**Homeostasis:** En Homeostasis ocurre lo mismo que en Utilitarianism. Las coberturas (tanto semántica como léxica) permanecen constantes, mientras que las precisiones y las similitudes estructurales mejoran. Destaca un aumento del X % en la precisión semántica, lo que indica que se han eliminado muchas tripletas redundantes sin perder información relevante.

**Inflation:** En el caso de *Inflation*, las coberturas semántica y léxica disminuyen ligeramente, pero la ganancia en precisión es más notable, sobre todo en la precisión semántica, lo que indica que, a pesar de perder algo de amplitud en la captura de relaciones, las tripletas retenidas son más exactas.

Estos resultados muestran que la eliminación de tripletas redundantes mejora de forma general la precisión y la calidad del grafo, con un coste muy bajo en la cobertura en algunos casos.

**Evolución media de métricas**

[Poner imagen L7]

En la figura X se aprecia que, al aplicar la Estrategia 7, tanto la precisión como la similitud estructural mejoran de forma notable, con incrementos que oscilan entre el 18 % y el 29 %. Por otro lado, la cobertura semántica apenas se ve afectada, empeorando solo un 5,2 %, y la cobertura léxica disminuye únicamente un 0,9 %. Por tanto, la Estrategia 7 cumple su objetivo de aumentar la precisión y la coherencia estructural sin alterar de manera significativa la cobertura.

Evaluación del bloque

El objetivo de este último bloque consistía en mejorar la precisión del grafo obtenido tras aplicar las transformaciones seleccionadas en los bloques anteriores.

Por un lado, la Estrategia 6 no logra mejoras constantes. En general reduce las métricas y apenas incrementa la precisión, de modo que los pocos avances obtenidos resultan insuficientes frente a las pérdidas en cobertura y similitud estructural.

Por otro lado, la estrategia 7 de eliminación de tripletas similares, sí que muestra un aumento en la precisión y la coherencia estructural sin alterar de manera significativa la cobertura.

En la figura X se aprecia que la Estrategia 6 apenas aporta mejoras, mientras que la Estrategia 7 mejora todas las métricas salvo la cobertura, que sufre una ligera disminución respecto al grafo de inicio de este bloque. Por tanto, se selecciona la Estrategia 7, pues es la que mejor cumple el objetivo de conservar únicamente las tripletas más precisas.

[Poner imagen evB3]