**Bloque I: Expansión de cobertura**

**De acuerdo a la primera carencia que se vio en …**  
El objetivo de este bloque de estrategias es capturar nuevas tripletas que incrementen la cobertura global, incorporando relaciones que el extractor inicial no detectó.

4.2.1.1 Resultados por estrategia

## Estrategia 1: Normalización por Chunking y Re-Extracción

La idea de partida consistía en convertir cada sintagma nominal de los textos de entrada en un token único y luego volver a aplicar el extractor de tripletas, con el objetivo de aumentar el número de relaciones cuyos sujetos y objetos correspondan a entidades completas. Se esperaba así mejorar la cobertura del grafo sin sacrificar demasiado la precisión.

Al aplicar esta estrategia a cada uno de los textos de referencia, se observa un incremento notable en el número de tripletas extraídas:

* En Utilitarianism, el conteo pasó de 52 a 63 tripletas (11 tripletas más, +21,2 %).
* En Ellipse, aumentó de 48 a 60 tripletas (12 tripletas más, +25,0 %).
* En Homeostasis, subió de 34 a 58 tripletas (24 tripletas más, +70,6 %).
* En Inflation, creció de 101 a 120 tripletas (19 tripletas más, +18,8 %).

Muchas de estas nuevas relaciones corresponden a tripletas que antes estaban fragmentadas o habían sido omitidas porque el extractor de OpenIE no las había reconocido como entidades compuestas.

El siguiente paso es verificar si estas tripletas adicionales son realmente correctas y si aportan nueva información. Para ello se analiza las métricas de precisión, cobertura y similitud estructural frente al grafo gold.

Análisis por texto

La figura X muestra cómo la aplicación de la Estrategia 1 al grafo baseline modifica las métricas de comparación con el grafo gold.

[Poner imágenes B1U, B1E, B1H, B1I]

– Utilitarianism: En la gráfica se aprecia que todas las métricas mejoran, especialmente las coberturas: a cobertura semántica aumenta en un X % y la cobertura léxica en un X %. Estos resultados demuestran que al aplicar la Estrategia 1 al grafo baseline del Utilitarismo, se han extraído tripletas de alto valor, que mejoran tanto la precisión como la cobertura del grafo.

– Ellipse: En el caso de *Ellipse*, las coberturas también aumentan: la cobertura semántica se eleva en un X % y la cobertura léxica en un X %. aunque la precisión semántica mejora, la precisión léxica disminuye en un X %. Esto puede deberse a que la Estrategia 1 genera nuevas variantes para referirse a una misma entidad, y esas formas alternativas se alejan más de las expresiones utilizadas en el grafo gold. Aun así, la mejora en la precisión semántica indica que el significado de las nuevas tripletas sí que es preciso para el texto de la elipse. Esta caída en la precisión léxica se refleja también en una reducción de X % en la métrica léxica de similitud estructural.

– Homeostasis: En el texto de Homeostasis, la Estrategia 1 resulta menos efectiva. Aunque la cobertura semántica aumenta en un X %, la cobertura léxica permanece sin cambios y las precisiones caen. Esto indica que las nuevas entidades extraídas son realmente adicionales, por eso sube la cobertura, pero no coinciden suficientemente con las del grafo gold, lo que penaliza las precisiones.

Esta pérdida de precisión se refleja también en un ligero empeoramiento de la similitud estructural del grafo con el grafo gold, tal como muestran las métricas de Jaccard.

– Inflation: En *este caso*, la cobertura semántica se incrementa en un X % y la cobertura léxica en un X %, mientras que las precisiones se mantienen prácticamente constantes. La calidad del grafo generado con la Estrategia 1 es similar a la del baseline, pero incorpora nuevas tripletas valiosas, sobre todo desde el punto de vista semántico.

En todos los casos, la re-extracción a partir de chunks unificados mejora la cobertura del grafo sin sacrificar de forma significativa ni la precisión ni la similitud global con el grafo gold.

Evolución media de las métricas

Tras aplicar la Transformación 1 al grafo baseline de los cuatro textos, la evolución media de las métricas queda reflejada en la figura X.

[Poner imagen L1]

En la figura puede apreciarse que, en promedio, el aumento de las coberturas supera la ligera disminución de las precisiones.

Este comportamiento confirma que la estrategia logra elevar la cobertura de relaciones con un coste moderado en precisión. Además, la evolución de las métricas de Jaccard indica que la estructura global del grafo no se ve alterada de forma significativa.

## Estrategia 2: Extracción de tripletas de chunks con verbos

La hipótesis de partida era que, al reutilizar los chunks verbales como entradas para el extractor de tripletas de OpenIE, se detectarían relaciones que habían quedado fuera en la extracción inicial, aumentando la cobertura sin añadir apenas ruido.

Al aplicar esta segunda estrategia, se generaron nuevas tripletas para cada uno de los textos de referencia, sin contar las ya presentes en el baseline:

* Utilitarianism: 2 tripletas nuevas.
* Ellipse: 3 tripletas nuevas.
* Homeostasis: 5 tripletas nuevas.
* Inflation: 22 tripletas nuevas.

En este caso, al generar menos tripletas nuevas que con la estrategia 1, el impacto de la estrategia 2 será menor. Además, dado que algunas de estas tripletas coinciden con variantes ya presentes en el baseline, es imprescindible evaluar su relevancia frente al grafo “gold” utilizando las métricas correspondientes.

[Poner imágenes B2U, B2E, B2H, B2I]

Análisis por texto

* **Utilitarianism**: En la gráfica del utilitarismo (figura X) se observa que las coberturas semántica y léxica permanecen constantes, mientras que las precisiones disminuyen en torno al 1 %. Esto sugiere que las dos tripletas adicionales generadas no aportan información relevante al grafo, sino que introducen ruido. Con solo dos tripletas nuevas respecto al baseline, no hay cambios significativos.
* **Ellipse**: Ocurre lo mismo para el caso de la elipse. las coberturas semántica y léxica permanecen constantes, mientras que las precisiones disminuyen ligeramente. Esto sugiere que las tripletas extraídas de los chunks verbales tampoco aportan información relevante a este grafo.
* **Homeostasis**: Para este texto, se observa que la estrategia resultó más beneficiosa. Todas las métricas presentan un pequeño incremento. Sin embargo, al haberse incorporado tan pocas tripletas al baseline, el impacto global en las métricas tampoco es relevante.
* **Inflation**: Destaca la mejora de la cobertura semántica que crece en un X%. En este caso, las métricas semánticas mejoran más que las métricas léxicas, lo que indica que las tripletas añadidas con la estrategia 2 han aportado un valor semántico significativo al grafo baseline del texto sobre inflación. Sin embargo, salvo la cobertura semántica, el resto de las métricas se mantienen prácticamente constantes.

En conjunto, estos resultados muestran que la estrategia 2 es poco consistente. En algunos textos, aporta un ligero aumento de cobertura, mientras que en otros casos, no solo carece de efecto, sino que incluso introduce ruido en el grafo. Además, las diferencias entre aplicarla o no aplicarla son siempre muy leves.

Evolución media de métricas

[Poner imagenes L2]

En la figura X se observa que la cobertura aumenta también con esta estrategia, aunque en menor medida que conla estrategia 1. Concretamente:

La cobertura semántica mejora en un 5.1% y la léxica en un 1.5%. Por otro lado la precisión semantica empeora un 0.8% y la léxica un 2.7%

En cuanto a la similitud estructural el coeficiente de Jaccard semántico sube 1,0 %, pero el Jaccard léxico baja -1,0 %).

Estos resultados confirman que, en el plano semántico la mejora de la cobertura compensa a la empeora de la precisión, en el plano léxico no.

Por tanto, este método no constituye una mejora sólida y generalizable, sino que su eficacia depende del tipo de texto analizado.

En la figura X puede verse que la estrategia 2 también incrementa la cobertura, aunque en menor grado que la estrategia 1. La cobertura semántica aumenta un 5,1 % y la cobertura léxica un 1,5 %. Sin embargo, las precisión se ven perjudicadas: la precisión semántica cae un 0,8 % y la precisión léxica un 2,7 %. En cuanto a la similitud estructural, el coeficiente de Jaccard semántico sube un 1,0 %, pero el Jaccard léxico baja un en la misma medida. Estos resultados indican que, desde el punto de vista semántico, la ganancia en cobertura compensa la ligera pérdida de precisión, mientras que en el plano léxico esa compensación no se produce. Por ello, este método no ofrece una mejora sólida y generalizable, sino que su eficacia depende del tipo de texto analizado.

4.2.1.2 Evaluación del bloque

Este apartado tiene como objetivo determinar cuáles de las estrategias evaluadas en este bloque deben aplicarse al grafo baseline en las fases posteriores.

Por un lado, la Estrategia 1 (normalización mediante chunking y re-extracción) confirma de modo consistente la hipótesis inicial: en todos los casos, la re-extracción a partir de chunks unificados mejora la cobertura del grafo sin sacrificar de forma significativa ni la precisión ni la similitud global con el grafo “gold”. Esto demuestra que convertir los chunks en tokens únicos y volver a aplicar OpenIE es un método eficaz para capturar relaciones completas que el extractor original había pasado por alto.

Por otro lado, la Estrategia 2 (extracción de tripletas a partir de chunks verbales) solo aporta un ligero aumento de cobertura, muy variable según el texto, y en algunos casos incluso empeora las métricas más de lo que las mejora. Aunque en ocasiones añade relaciones nuevas y valiosas, su rendimiento no es sólido ni generalizable.

[Poner imagen evB1]

La figura X, que compara las medias de las métricas del grafo baseline, Transformación 1, Transformación 2 y la aplicación conjunta de ambas, muestra con claridad que la mejor opción es la Estrategia 1. Esta estrategia es la que más aumenta la cobertura de forma general sin afectar de manera significativa la precisión, lo que indica que no introduce ruido. Además, captura tripletas relevantes que la extracción inicial no había detectado. Por todo ello, se decide adoptar únicamente la Estrategia 1 como nueva base para las siguientes fases.