

TÉCNICAS BASADAS EN GRAFOS APLICADOS AL PROCESAMIENTO DEL LENGUAJE

**Máster Universitario en Tecnologías
del Lenguaje.**

Universidad Nacional de Educación a Distancia.

Franly Iris Urbina Franco
furbina7@alumno.uned.es
71.997.418-N

Contenido

INTRODUCCIÓN	3
Parte I: Estudio de corpus.	¡Error! Marcador no definido.
Brown Corpus	¡Error! Marcador no definido.
Susanne Corpus	9
Penn Treebank Corpus	11
Análisis comparativo entre Brown Corpus y Susanne Corpus.....	13
Parte II: Comparativa de etiquetadores estadísticos.	14
Stanford Tagger.....	15
Análisis de las etiquetas de Stanford Tagger.	16
TreeTagger	18
Análisis de las etiquetas de TreeTagger.....	17
Análisis comparativo: TreeTagger vs Stanford POS Tagger.	19
CONCLUSIÓN	20
GLOSARIO	21
BIBLIOGRAFÍA.....	23

INTRODUCCIÓN

El objetivo de esta práctica es analizar y caracterizar diferentes tipos de redes mediante la herramienta Gephi, aplicando los conceptos fundamentales de la teoría de grafos estudiados en el Tema 1 y el Tema 2 de la asignatura. A través de la observación de sus métricas estructurales —número de nodos y aristas, tipo de grafo, longitud media de los caminos, coeficiente de clustering, diámetro, grado medio y número de componentes conexas— se busca comprender cómo la forma y la conectividad de una red reflejan su naturaleza y función.

Para ello se estudian cuatro grafos de distinta procedencia y propósito:

EuroSiSGenerale, que representa las relaciones entre asociaciones científicas europeas.

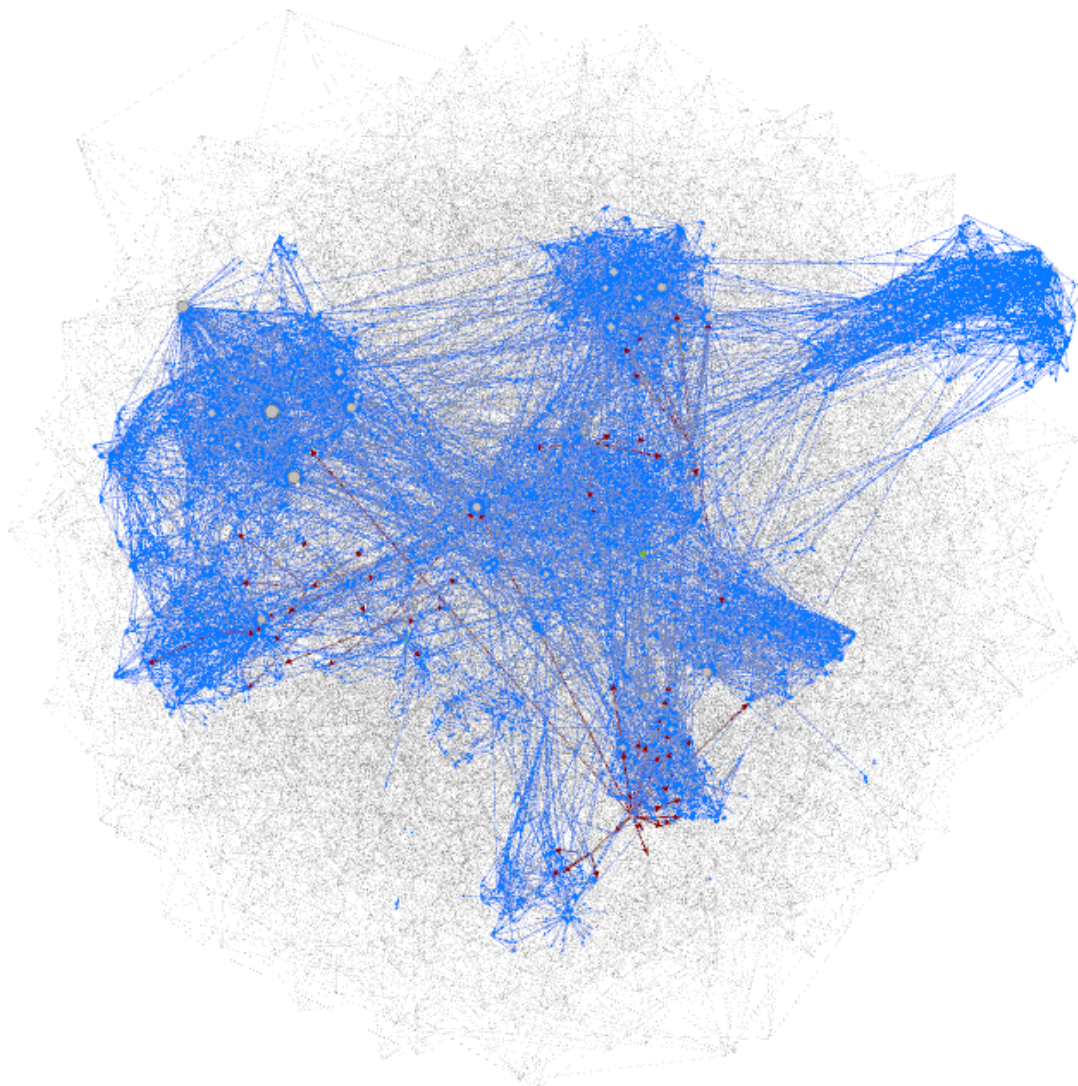
RetGraph, que muestra co-ocurrencias de palabras en frases dentro de un corpus documental.

LesMiserables, que modela la red de co-aparición de personajes de la novela de Victor Hugo.

BoundaryCountries, que representa las adyacencias geográficas entre países en un proyecto europeo.

El análisis de estos grafos permite identificar patrones característicos —como la distribución del grado, la presencia de comunidades, la existencia de nodos centrales o hubs, y las propiedades de pequeño mundo o de redes libres de escala— y compararlos con los modelos teóricos presentados en la documentación docente (Easley & Kleinberg, 2010; Araujo y Martínez-Romo, UNED). De este modo, la práctica ofrece una visión aplicada sobre cómo los conceptos matemáticos de la teoría de grafos se manifiestan en redes reales de naturaleza social, lingüística y geográfica.

EuroSisGenerale.gexf



Descripción.

El grafo EuroSiSGenerale representa la red de relaciones entre asociaciones científicas europeas generada por el proyecto European Science Information System, EuroSIS. Al realizar un primer acercamiento, podemos ver diferentes instituciones como La Universidad de Bielskok, La Red Europea de Centros de Ciencia y Museos, el Instituto de Chemii Przemyslowej, entre otros.

Tabla de datos							
Nodos							
Aristas							
Configuración							
Añadir nodo							
Añadir arista							
Buscar/Reemplazar							
Importar hoja de cálculo							
Exportar tabla							
Más acciones							
Filtros							
Origen							
Destino							
Tipo							
Id							
Label							
Interval							
Weight							
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	1862 - Beskidzki Festiwal Nauki i Sztuki	Dirigida	8992			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2097 - Lahti University of Applied Sciences	Dirigida	8995			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2120 - Lappeenranta University of Technology	Dirigida	8996			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2267 - University of Oulu	Dirigida	8997			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2272 - Polish Academy of Sciences	Dirigida	9001			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2510 - Turku University of Applied Sciences	Dirigida	8999			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2511 - Tampere University of Technology	Dirigida	8998			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2530 - Université de Nancy	Dirigida	8994			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2540 - Sofia University 'St. Kliment Ohridski'	Dirigida	9000			2.0	
1581 - University of Bielsko-Biala, Institute of Textile Engineering and Polymer Materials	2556 - University of Debrecen	Dirigida	8993			2.0	
1768 - European Network of Science Centres and Museum	1578 - Association of Science and Technology centers	Dirigida	10108			2.0	
1768 - European Network of Science Centres and Museum	1697 - Cité de l'espace-Toulouse	Dirigida	10109			2.0	
1768 - European Network of Science Centres and Museum	1698 - Cité des Sciences et de l'Industrie	Dirigida	10110			2.0	
1768 - European Network of Science Centres and Museum	2193 - Museo Nazionale della scienza e della tecnologia	Dirigida	10107			2.0	
1976 - Instytut Chemii Przemysłowej	1518 - Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie	Dirigida	11183			2.0	
1976 - Instytut Chemii Przemysłowej	1969 - Polish Hydrogen and Fuel Cell Association Website	Dirigida	11180			2.0	
1976 - Instytut Chemii Przemysłowej	2014 - Instytut Podstawowych Problemów Techniki PAN	Dirigida	11179			2.0	
1976 - Instytut Chemii Przemysłowej	2289 - Gdansk University of Technology	Dirigida	11181			2.0	
1976 - Instytut Chemii Przemysłowej	2339 - Politechnika Poznańska	Dirigida	11182			2.0	
2031 - Istituto Veneto	1501 - Accademia Nazionale delle Scienze	Dirigida	11388			2.0	
2031 - Istituto Veneto	1824 - European Science Foundation	Dirigida	11386			2.0	
2031 - Istituto Veneto	1853 - EXPLORA-La tv della scienza	Dirigida	11382			2.0	
2031 - Istituto Veneto	1882 - Fishbase	Dirigida	11389			2.0	
2031 - Istituto Veneto	2004 - Institut national du patrimoine	Dirigida	11387			2.0	
2031 - Istituto Veneto	2106 - Accademia Nazionale dei Lincei	Dirigida	11385			2.0	
2031 - Istituto Veneto	2172 - Molecular Lab	Dirigida	11383			2.0	
2031 - Istituto Veneto	2297 - Piatra	Dirigida	11384			2.0	
2182 - Institute for Political Sciences	1571 - Armenian Sociological Association	Dirigida	12309			2.0	
2182 - Institute for Political Sciences	1680 - Central European University	Dirigida	12300			2.0	
2182 - Institute for Political Sciences	1766 - HAS Institute for Economy	Dirigida	12302			2.0	
2182 - Institute for Political Sciences	1805 - University of Debrecen - Centre of Arts, Humanities and Sciences	Dirigida	12306			2.0	
2182 - Institute for Political Sciences	1824 - European Science Foundation	Dirigida	12304			2.0	
2182 - Institute for Political Sciences	1839 - Institute for Political Sciences	Dirigida	12292			2.0	
2182 - Institute for Political Sciences	1904 - Work and technology research centre	Dirigida	12291			2.0	
2182 - Institute for Political Sciences	2140 - Median Market Research Co.	Dirigida	12297			2.0	

Imagen 1.- Tabla de datos de EuroSisGenerale.gexf

Se trata de un grafo dirigido y ponderado, donde los nodos corresponden a organizaciones y las aristas reflejan posibles vínculos de colaboración o dependencia entre ellas.

Una cualidad destacable de esta red es su estructura jerárquica, la presencia de unos pocos nodos con un número muy elevado de conexiones sugiere la existencia de instituciones que actúan como puntos intermedios de conexión. Esta configuración favorece una difusión eficiente de información en toda la red, aunque implica dependencia estructural de estos nodos principales.

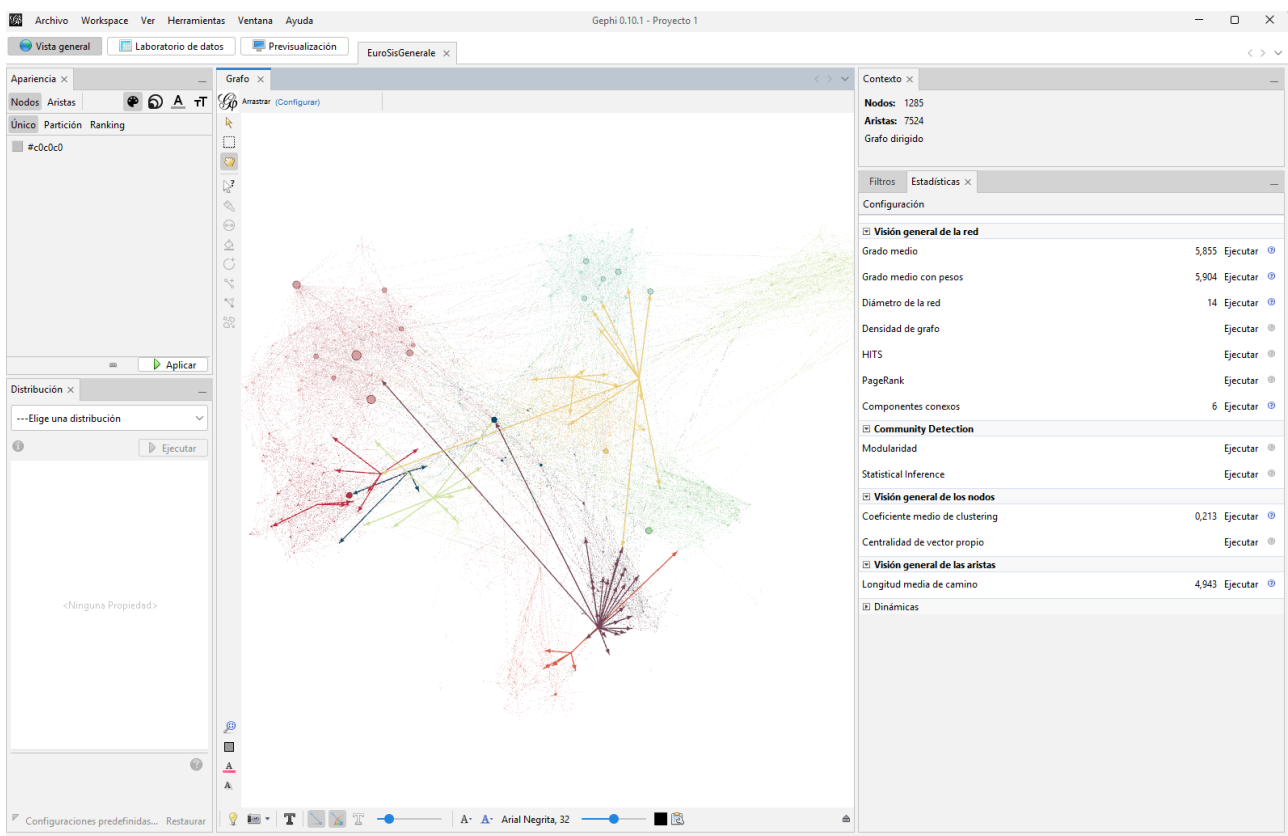


Imagen 2. Datos Estadísticos de EuroSisGenerale.gexf

Connected Components Report

Parameters: Network Interpretation: directed

Results:

Number of Weakly Connected Components: 6
Number of Strongly Connected Components: 511

Size Distribution

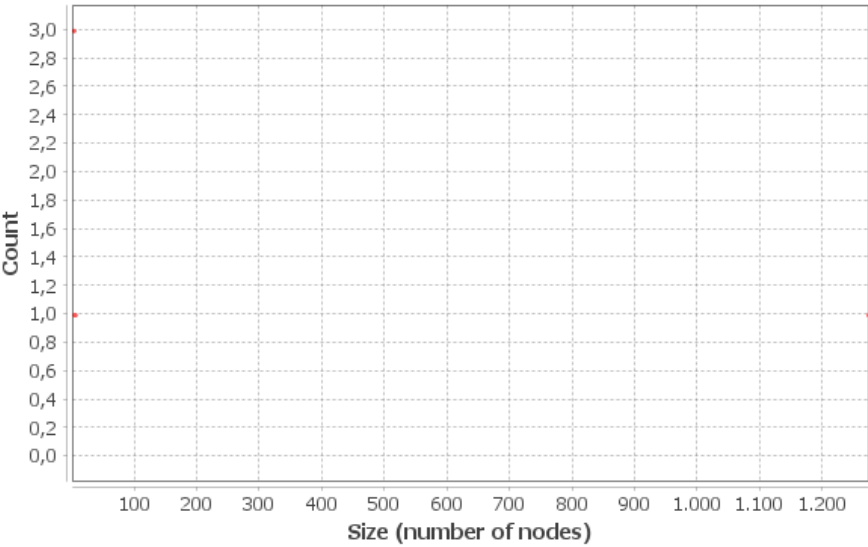


Imagen 3. Datos Estadísticos de EuroSisGenerale.gexf – Número de componentes fuertemente conexas SCC y débilmente conexas WCC.

Extraemos los datos ejecutando en el apartado de Estadísticas los valores que nos interesan, con ellos, conseguimos generar la siguiente tabla resumen.

Datos	Valor
Nodos	1285
Aristas	7524
Grafo dirigido	Mixto
Grafo con pesos	Sí
Longitud media de los caminos	4.94
Coefficiente de clustering medio	0.213
Diámetro de la red	14u
Grado medio	5.855
Grado medio (ponderado)	5.904
Componentes conexas fuertes SCC	511
Componentes conexas débiles WCC	6

Tabla 1. Datos de EuroSisGenerale.gexf

Distribución del grado

la probabilidad de que un nodo tomado al azar tenga grado k , y se calcula como:

$$P(k) = \frac{N_k}{N}$$

donde N_k es el número de nodos con grado k .

En **EuroSiSGenerale**, el **grado medio es 5.855**, lo que indica que la mayoría de los nodos están conectados con alrededor de seis enlaces. Aunque la red tiene un número total de **1285 nodos y 7524 aristas**, sugiere cierta heterogeneidad. El coeficiente de clustering, la longitud media de camino y el diámetro indica que hay caminos cortos que acortan la red y los clustering locales muestran como grupos vecinos también se conectan entre sí.

El grafo posee una gran cantidad de relaciones fuertemente conexas entre entidades pequeñas y solo unas pocas sugieren tener relaciones fuertemente conexas con algunas grandes, lo que hace ver una asimetría entre instituciones centrales que reciben comunicación con entidades pequeñas, pero no siempre hay reciprocidad, tal como se puede esperar de entidades institucionales grandes.

Clasificación del grafo

Gracias a las métricas obtenidas por el programa Gephi, podemos ver varias características que posee el grafo y que no son excluyentes entre sí.

Red Small-World.

La red EuroSiSGenerale, puede considerarse de tipo mundo pequeño porque combina un alto valor medio de clustering local de 0.213 con caminos cortos entre nodos distantes de 4.93.¹ En este tipo de redes, aun cuando cada nodo se conecta con un número pequeño de vecino, se puede alcanzar a otro nodo en pocos pasos a través de intermediarios.

La existencia de nodos altamente conectados como grandes agencias científicas y universidades, permite que asociaciones o centros más pequeños queden indirectamente conectados.

Este patrón cumple la propiedad característica del pequeño mundo descrita por Easley y Kleinberg:

$$L \approx L_{\text{rand}} \text{ y } C \gg C_{\text{rand}},$$

EuroSiSGenerale es un ejemplo donde se produce el mismo efecto observado en la red de colaboraciones científicas que inspiró el concepto de seis grados de separación, es por ello por lo que puede considerarse una red de mundo pequeño ya que combina una longitud media de los caminos pequeña con un coeficiente de clustering alto en comparación con una red aleatoria.

Red libre de escala (scale-free)

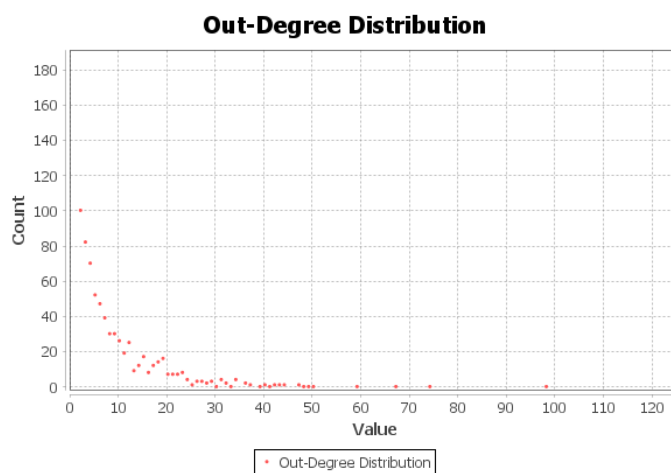
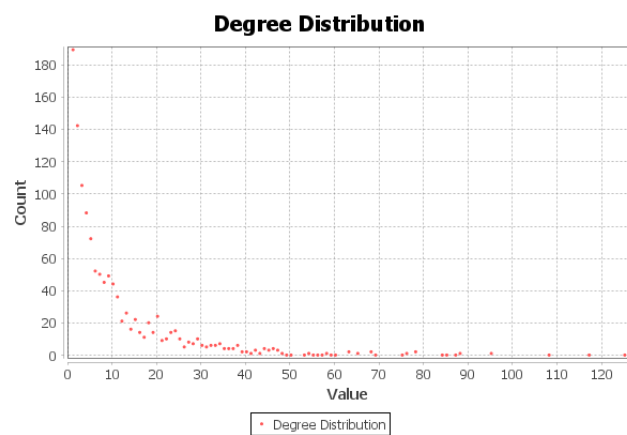
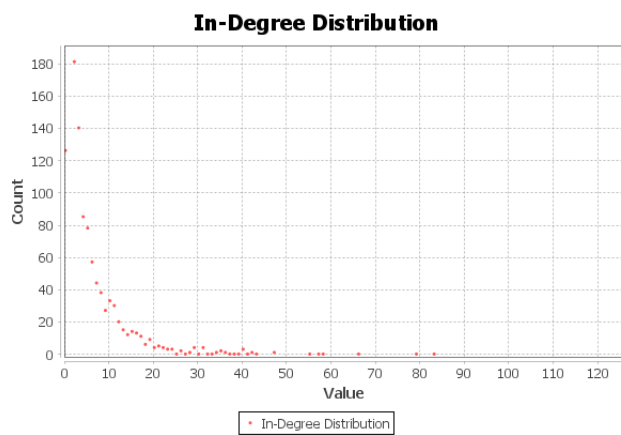
Se define una **red libre de escala** como aquella donde la distribución del grado sigue una ley de potencias:

$$P(k) = k^{-c}$$

con c entre aproximadamente 2 y 3.

Se puede observar que la red no muestra una distribución uniforme o aleatoria de conexiones, sino una disminución rápida y continua del número de nodos al aumentar el grado, con cola larga hacia la derecha. Este Patrón es coherente con una distribución de tipo potencia.

Average Weighted Degree: 5,904



Susanne Corpus

El Susanne Corpus, del acrónimo Surface and Underlying Structural Analyses of Natural English, fue desarrollado por Geoffrey Sampson durante la década de 1990 en la Universidad de Sussex en Falmer, Inglaterra. El objetivo era crear un corpus que ofreciera una anotación sintáctica detallada basada en textos del Brown Corpus. De hecho, se considera una versión anotada del Brown, centrada en proporcionar una representación más profunda de la estructura gramatical del inglés.

La primera versión se creó en 1992, y la última, la Release 5, finalizó el 11 de agosto del 2000. Fue creado como parte del proceso de desarrollo de una taxonomía orientada al lenguaje de ingeniería y de un esquema de anotación para la gramática del inglés tanto británico como americano. El esquema utilizado por SUSANNE es exhaustivo, cubre todas las características de la gramática inglesa que pueden ser anotadas formalmente, incluyendo los fenómenos que ocurren en la práctica del inglés moderno dado a que ha sido desarrollado y depurado por numerosos investigadores y aplicado a una gran cantidad de ejemplos de inglés británico y americano. Se puede decir, por tanto, que es consistente.

El corpus contiene aproximadamente 130.000 palabras y está diseñado para servir como referencia para el análisis sintáctico y semántico del lenguaje natural. Cada oración está analizada de forma jerárquica de árbol, con las etiquetas de los nodos no terminales insertados entre la apertura y el cierre, representando explícitamente las relaciones entre sus constituyentes, lo que lo convierte en un corpus esencial para el estudio de la estructura oracional y el desarrollo de parsers.

Frase: *The boy that Mary saw yesterday was running in the park.*

```
(S
  (NP-SBJ
    (DT The)
    (NN boy)
    (SBAR-REL # oración de relativo
      (WHNP (WDT that))
      (S
        (NP-SBJ (NP (NP Mary))) # sujeto de la relativa
        (VP
          (VBD saw)
          (NP-OBJ (NN yesterday))))))
    (VP
      (VBD was)
      (VP
        (VBG running)
        (PP
          (IN in)
          (NP
            (DT the)
            (NN park))))))
```

Figura 1. Estructura jerárquica de frase y etiquetado sintáctico en el *Susanne Corpus*

Donde:

- **S**: Oración completa.
- **NP**: Sintagma nominal.
- **VP**: Sintagma verbal.
- **PP**: Sintagma preposicional.
- **SBAR**: Oración subordinada.
- **WHNP**: wh-sintagma¹.
- **-SBJ / -OBJ / -REL**: función de sujeto, objeto, relativa.
- **DT**: determinante
- **JJ**: adjetivo
- **NN**: sustantivo
- **NNP**: propio
- **VBZ**: verbo en tercera persona del singular.
- **VBD**: verbo en pasado
- **VBG**: verbo en gerundio
- **IN**: preposición.
- **WDT**: relativo.

Las anotaciones del Susanne Corpus van mucho más allá del etiquetado léxico tradicional. Además de incluir etiquetas gramaticales (POS tagging), incorpora árboles sintácticos completos, marcadores de función gramatical como sujeto u objeto y etiquetas morfológicas. En total, emplea más de 350 tipos de etiquetas estructurales, organizadas jerárquicamente, que permiten describir con precisión tanto la forma superficial de las frases como su estructura.

El Susanne ha tenido un papel clave en el desarrollo de herramientas de análisis sintáctico, evaluación de parsers y sistemas de chunking lingüístico. Gracias a su nivel de detalle y coherencia estructural, permite el desarrollo de modelos capaces de aprender dependencias sintácticas complejas, evaluar la corrección de análisis automáticos y entrenar sistemas que combinan información léxica y estructural en tareas avanzadas de procesamiento del lenguaje natural.

²Un "wh-sintagma" se refiere a un sintagma (o grupo de palabras) que comienza con una palabra interrogativa (generalmente las que en inglés empiezan por "wh-", como what, where, when, why) y se utiliza para formular preguntas. En lingüística, el término "palabra qu-" se usa en español por analogía con el inglés.

Penn Treebank Corpus

El Penn Treebank es uno de los corpus más influyentes en la historia del procesamiento del lenguaje natural desarrollado por la Universidad de Pensilvania en Filadelfia, (EE. UU) en el Laboratorio del Departamento de Computación y de Ciencias de la Información (LINC) a finales de los años 80 y principios de los 90 bajo la dirección de Mitchell Marcus.

El objetivo principal fue el de proporcionar un corpus extensamente anotado a nivel léxico y sintáctico del inglés americano. Su primera versión se publicó en 1993, y desde entonces se ha convertido en un estándar de referencia en la evaluación de modelos lingüísticos y parsers.

El corpus se basa en diferentes fuentes textuales, incluyendo artículos periodísticos del Wall Street Journal, así como textos del Brown Corpus, del Switchboard Corpus³ y del ATIS (Air Travel Information System)⁴. Posee más de 4,5 millones de palabras en inglés americano, distribuidas en distintos subcorpus adaptados a tareas específicas de análisis del lenguaje.

El Penn Treebank fue etiquetado de forma manual y semiautomática, utilizando el etiquetador CLAWS⁵ como punto de partida y revisiones posteriores por lingüistas. Las anotaciones se dividen en dos niveles principales:

1. Etiquetado léxico o POS tagging, con un conjunto reducido de unas 45 etiquetas gramaticales, que se convirtió en el estándar de facto para muchos sistemas posteriores.
2. Anotación sintáctica jerárquica, en la que cada oración está representada mediante un árbol de constituyentes, siguiendo la gramática de constituyentes del inglés (similares a las del Susanne, pero más normalizadas).

Las etiquetas del Penn Treebank están basadas en el Corpus Brown, pero al tener una orientación computacional y estadística, se priorizó la consistencia y manejabilidad de las etiquetas, por ende, se redujo el número de etiquetas mediante la eliminación de la redundancia léxica y sintáctica. En consecuencia, este corpus consta de 36 etiquetas POS y otras 12 etiquetas para signos de puntuación y símbolos de monedas, lo que hace un total de 48 etiquetas frente a las aproximadamente 87 del Brown Corpus.

El valor del Penn Treebank reside en su enorme influencia en la creación y evaluación de modelos estadísticos del lenguaje y parsers probabilísticos. Los Modelos de Gramáticas de Estructura de Frases Probabilísticas (PCFG), así como en la formación de conjuntos de datos de referencia (benchmarks) para tareas de POS tagging, parsing, chunking y posteriormente, entrenamiento de modelos neuronales, se vieron beneficiados por esta reducción del etiquetado.

Su estructura y coherencia lo hacen ideal para el desarrollo de modelos capaces de aprender dependencias gramaticales complejas y realizar análisis sintáctico probabilístico, convirtiéndolo en uno de los corpus más citados en la lingüística computacional moderna.

³El Switchboard Telephone Speech Corpus es un corpus de lengua inglesa hablada que constaba de casi 260 horas de habla. Fue creado en 1990 por Texas Instruments a través de una subvención DARPA y lanzado en 1992 por NIST. El corpus contiene 2.400 conversaciones telefónicas entre 543 hablantes estadounidenses.

⁴The Airline Travel Information System dataset (ATIS) es un corpus para la comprensión lingüística en sistemas de diálogo orientados a tareas. Contiene 4978 enunciados de entrenamiento de los datos.

⁵CLAWS es un etiquetador gramatical automático.

	Brown	SUSANNE	Penn Treebank
Tipo de etiquetado	POS tagging	Anotación sintáctica (árboles de constituyentes) + POS + funciones	Anotación sintáctica (árboles de constituyentes) + POS
Tamaño	≈1.014.312 palabras, unos 500 textos con aproximadamente 2000 palabras cada uno para su primera versión en 1961.	≈130.000 palabras (muestras del Brown)	>4,5 millones de palabras (fases 1989–1992)
Número de etiquetas (POS)	≈87 etiquetas POS clásicas	≈353 “wordtags” (morfosintácticas) + etiquetas estructurales.	48 POS (36 POS + 12 de puntuación/símbolos)
Temática	15 categorías: prensa (reportaje, editorial, reseñas), religión, “skills & hobbies”, “popular lore”, belles-lettres/biografía/memorias/ensayo, miscelánea, learned (ciencia/medicina/tecnología...), ficción (general, misterio, ciencia, western, romance), humor.	64 textos del corpus Brown.	Principalmente Wall Street Journal; además materiales de Brown, Switchboard y ATIS.
Idioma	Inglés americano escrito, EE. UU., 1961.	Inglés americano escrito, EE. UU., 1961 (Subconjunto de Brown)	Inglés americano escrito, Wall Street Journal y habla transcrita.

Tabla 2. Comparativa de Brown, SUSANNE y Penn Treebank

Análisis comparativo entre Brown Corpus y Susanne Corpus.

Breve análisis (5 líneas como máximo) de que corpus, el de Susanne o el de Brown, es más apropiado en función de sus características para extraer información estadística significativa referente a cuáles son las etiquetas léxicas y las parejas de etiquetas léxicas consecutivas que aparecen más frecuentemente en los textos.

Dado el análisis anterior, el Brown Corpus es más apropiado porque contiene alrededor de un millón de palabras de distintos géneros. Esto permite obtener frecuencias léxicas y combinaciones de etiquetas estadísticamente significativas. Susanne Corpus, sin embargo, como un subconjunto, está enfocado al análisis estructural del texto y no en su frecuencia. Brown ofrece mayor representatividad, tamaño y variedad léxica para este tipo de estudio.

Parte II: Comparativa de etiquetadores estadísticos.

En esta segunda parte se realiza un estudio comparativo entre dos etiquetadores estadísticos para el procesamiento del lenguaje natural: **Stanford POS Tagger** y **TreeTagger**. Ambos sistemas permiten asignar categorías gramaticales a las palabras de un texto en base al tagset del Penn Treebank, pero se diferencian en los modelos probabilísticos que emplean y en la granularidad de sus etiquetas.

Stanford Tagger, fue desarrollado en Java por un grupo de investigación perteneciente a la Universidad de Stanford, California (EE. UU) y se basa en modelos de Máxima Entropía entrenados con el Penn Treebank. Este etiquetador posee una alta precisión contextual, la cual, se aproxima al 97% para determinar la categoría más probable.

TreeTagger, fue desarrollado en C por Helmut Schmid en la Universidad de Stuttgart (1995), utiliza árboles de decisión con un modelo probabilístico del contexto. Su precisión se aproxima al 95% y su velocidad de ejecución, facilita la ejecución para grandes textos, diferentes idiomas y necesidades en el campo estadístico.

Ambos fueron elegidos por su entrenamiento a base del inglés americano escrito, con lo que podemos realizar una comparativa justa en base al siguiente texto.

On a warm afternoon, the old man decided to watch the river flow while a group of children laughed and played near the bank. The light breeze carried the scent of wet grass, and a few leaves began to dance lazily on the surface of the water. He had planned to record the song of the birds, but the steady hum of traffic from the nearby road made it almost impossible. With a sigh, he leaned back on the bench, closing his eyes for a moment just to feel the warmth of the sun.

Across the bridge, a couple argued quietly outside a small hotel. "Did you remember to book the room?" she asked, and he smiled weakly, holding up a crumpled receipt. The owner, a thin man with kind eyes, opened the door and waved them in. "You're lucky!," he said, "there's only one room left." The old man watched them disappear inside.

Later, a group of workers arrived from the factory, carrying boxes too heavy to bear for long. One of them tried to park the truck under an old oak tree, while another knelt to repair the flat tire. The sound of their tools echoed through the park like a short melody of iron and effort. The old man smiled; life, he thought, always found a rhythm—even in the smallest moments.

As the sun began to set, the light turned golden, painting the water in shades of fire. He packed his things slowly, then walked home along the narrow path beside the river. The town was quiet now, except for a street musician playing an old tune on his guitar. The man stopped, listening. For a second, he felt young again—free of time, weight, and worry. When the song ended, he clapped softly, whispered "thank you," and kept walking. Above him, the first stars had already appeared, clear and sharp against the fading blue. He smiled, tightened his coat, and disappeared into the night.

Stanford Tagger

On_IN a_DT warm_JJ afternoon_NN ,_, the_DT old_JJ man_NN decided_VBD to_TO watch_VB the_DT river_NN flow_NN while_IN a_DT group_NN of_IN children_NNS laughed_VBD and_CC played_VBD near_IN the_DT bank_NN ._.

The_DT light_NN breeze_NN carried_VBD the_DT scent_NN of_IN wet_JJ grass_NN ,_, and_CC a_DT few_JJ leaves_NNS began_VBD to_TO dance_VB lazily_RB on_IN the_DT surface_NN of_IN the_DT water_NN ._.

He_PRP had_VBD planned_VBN to_TO record_VB the_DT song_NN of_IN the_DT birds_NNS ,_, but_CC the_DT steady_JJ hum_NN of_IN traffic_NN from_IN the_DT nearby_JJ road_NN made_VBD it_PRP almost_RB impossible_JJ ._.

With_IN a_DT sigh_NN ,_, he_PRP leaned_VBD back_RB on_IN the_DT bench_NN ,_, closing_VBG his_PRP\$ eyes_NNS for_IN a_DT moment_NN just_RB to_TO feel_VB the_DT warmth_NN of_IN the_DT sun_NN ._.

Across_IN the_DT bridge_NN ,_, a_DT couple_NN argued_VBD quietly_RB outside_IN a_DT small_JJ hotel_NN ._.

"_`` Did_VBD you_PRP remember_VB to_TO book_VB the_DT room_NN ?_.

"_" she_PRP asked_VBD ,_, and_CC he_PRP smiled_VBD weakly_RB ,_, holding_VBG up_RP a_DT crumpled_VBN receipt_NN ._.

The_DT owner_NN ,_, a_DT thin_JJ man_NN with_IN kind_JJ eyes_NNS ,_, opened_VBD the_DT door_NN and_CC waved_VBD them_PRP in_IN ._.

"_`` You_PRP 're_VBP lucky_JJ !_.

,_, "_" he_PRP said_VBD ,_, "_`` there_EX 's_POS only_RB one_CD room_NN left_VBN ._.

"_" The_DT old_JJ man_NN watched_VBD them_PRP disappear_VB inside_RB ._.

Later_RB ,_, a_DT group_NN of_IN workers_NNS arrived_VBD from_IN the_DT factory_NN ,_, carrying_VBG boxes_NNS too_RB heavy_JJ to_TO bear_VB for_IN long_JJ ._.

One_CD of_IN them_PRP tried_VBD to_TO park_VB the_DT truck_NN under_IN an_DT old_JJ oak_NN tree_NN ,_, while_IN another_DT knelt_VBN to_TO repair_VB the_DT flat_JJ tire_NN ._.

The_DT sound_NN of_IN their_PRP\$ tools_NNS echoed_VBD through_IN the_DT park_NN like_IN a_DT short_JJ melody_NN of_IN iron_NN and_CC effort_NN ._.

The_DT old_JJ man_NN smiled_VBD ;_: life_NN ,_, he_PRP thought_VBD ,_, always_RB found_VBD a_DT rhythm_NN —_HYPH even_RB in_IN the_DT smallest_JJS moments_NNS ._.

As_IN the_DT sun_NN began_VBD to_TO set_VB ,_, the_DT light_NN turned_VBD golden_JJ ,_, painting_VBG the_DT water_NN in_IN shades_NNS of_IN fire_NN ._.

He_PRP packed_VBD his_PRP\$ things_NNS slowly_RB ,_, then_RB walked_VBD home_RB along_IN the_DT narrow_JJ path_NN beside_IN the_DT river_NN ._.

The_DT town_NN was_VBD quiet_JJ now_RB ,_, except_IN for_IN a_DT street_NN musician_NN playing_VBG an_DT old_JJ tune_NN on_IN his_PRP\$ guitar_NN ._.

The_DT man_NN stopped_VBD ,_, listening_VBG ._.

For_IN a_DT second_JJ ,_, he_PRP felt_VBD young_JJ again_RB —_HYPH free_RB of_IN time_NN ,_, weight_NN ,_, and_CC worry_NN ._.

When_WRB the_DT song_NN ended_VBD ,_, he_PRP clapped_VBD softly_RB ,_, whispered_VBD "_`` thank_VBP you_PRP ,_, "_" and_CC kept_VBD walking_VBG ._.

Above_IN him_PRP ,_, the_DT first_JJ stars_NNS had_VBD already_RB appeared_VBN ,_, clear_JJ and_CC sharp_JJ against_IN the_DT fading_VBG blue_JJ ._.

He_PRP smiled_VBD ,_, tightened_VBD his_PRP\$ coat_NN ,_, and_CC disappeared_VBD into_IN the_DT night_NN ._.

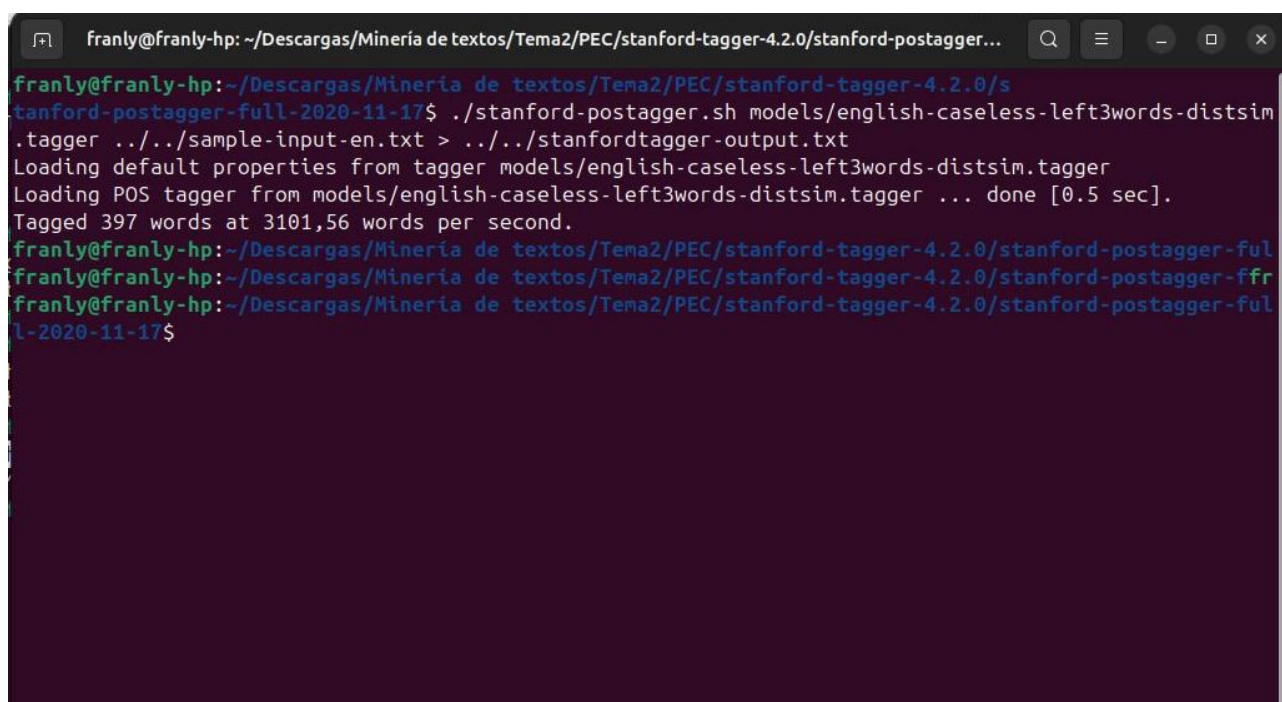
Análisis de las etiquetas de Stanford Tagger.

Token	Etiqueta dada	Etiqueta esperada	Explicación	Contexto
flow	NN	VB	Verbo en infinitivo tras “watch” (“to watch the river flow”); no es un sustantivo aquí.	On a warm afternoon, the old man decided to watch the river flow while a group of children laughed...
light	NN	JJ	Modificador adjetival de un nombre (“light breeze”); no funciona como sustantivo sino como adjetivo.	The light breeze carried the scent of wet grass, and a few leaves began to dance lazily on the surface...
's	POS	VBZ	Contracción de “is” (“there ’s only one room”); no es el clítico posesivo POS, es la tercera persona del singular del verbo 'to be'.	“You ’re lucky!,” he said, “there’s only one room left.”
another	DT	PRP	another funciona como pronombre (“another [one]”) sin sustantivo explícito, no como determinante.	...while another knelt to repair the flat tire.
left	VBN	JJ	Adjetivo predicativo (“one room left”); no es participio pasivo.	“You ’re lucky!,” he said, “there’s only one room left.”
long	JJ	RB	Uso adverbial de duración tras “for” (“for long”); no es adjetivo.	...carrying boxes too heavy to bear for long.
knelt	VBN	VBD	Pasado simple irregular (“another knelt ...”); actúa como verbo principal, no es participio aquí.	...while another knelt to repair the flat tire.
second	JJ	NN	Tras un determinante (a), second es un nombre de tiempo (“a second”), no un adjetivo (JJ).	For a second, he felt young again — free of time, weight, and worry.
thank	VBP	VB	En la interjección/imperativo “thank you”, thank va en forma base (VB), no presente (VBP).	...he clapped softly, whispered “thank you,” and kept walking.
blue	JJ	NN	blue se usa como nombre de color (“the fading blue”), no como adjetivo.	...appeared, clear and sharp against the fading blue.

Tabla 3. Análisis sintáctico de las etiquetas generadas por Stanford Tagger.

Tras pasar el texto etiquetado por un analizador⁶, descubrimos lo siguiente: El total de 329 palabras se desglosó con un total de 397 tokens, ya que se analizan signos de puntuación y comillas. De todos estos Token los cuales 9 son incorrectos (según evaluación lingüística) y el resto, 388 fueron etiquetados correctamente, generando una precisión estimada del 97.73%. Este hecho es revelador porque se ajusta a lo que promete la descripción.

En la [Tabla 3. Análisis sintáctico de las etiquetas generadas por Stanford Tagger](#) se puede observar en detalle que etiquetas asignó el etiquetador, cuales debían ser las esperadas, el motivo y la frase donde la palabra aparece y a pesar de contener errores, este número de aciertos implica un profundo desarrollo en el análisis del etiquetado.



```

franly@franly-hp: ~/Descargas/Minería de textos/Tema2/PEC/stanford-tagger-4.2.0/stanford-postagger...
franly@franly-hp:~/Descargas/Minería de textos/Tema2/PEC/stanford-tagger-4.2.0/s
tanford-postagger-full-2020-11-17$ ./stanford-postagger.sh models/english-caseless-left3words-distsim
.tagger ../../sample-input-en.txt > ../../stanfordtagger-output.txt
Loading default properties from tagger models/english-caseless-left3words-distsim.tagger
Loading POS tagger from models/english-caseless-left3words-distsim.tagger ... done [0.5 sec].
Tagged 397 words at 3101,56 words per second.
franly@franly-hp:~/Descargas/Minería de textos/Tema2/PEC/stanford-tagger-4.2.0/stanford-postagger-ful
franly@franly-hp:~/Descargas/Minería de textos/Tema2/PEC/stanford-tagger-4.2.0/stanford-postagger-ffr
franly@franly-hp:~/Descargas/Minería de textos/Tema2/PEC/stanford-tagger-4.2.0/stanford-postagger-ful
l-2020-11-17$

```

Imagen 1. Salida por consola de la ejecución del etiquetador Stanford Tagger.

⁶Realmente, se tuvo que generar muchos análisis, aproximadamente 15, con un programa en Python y tras ajustarlo varias se pudo conseguir un análisis bastante preciso de lo que Stanford Tagger generó.

TreeTagger

On	IN	on	.	SENT	.	of	IN	of	kind	NN	kind
a	DT	a	He	PP	he	the	DT	the	eyes	NNS	eye
warm	JJ		had	VBD	have	sun	NN	sun	,	,	,
	warm		planned	VTB	plan	.	SENT	.	opened	VBD	open
afternoon		NN	to	TO	to	Across	IN		the	DT	the
	afternoon		record	VB			across		door	NN	door
,	,	,		record		the	DT	the	and	CC	and
the	DT	the	the	DT	the	bridge	NN		waved	VBD	wave
old	JJ	old	song	NN	song		bridge		them	PP	them
man	NN	man	of	IN	of	,	,	,	in	RP	in
decided	VBD		the	DT	the	a	DT	a	.	SENT	.
	decide		birds	NNS	bird	couple	NN		"	``	"
to	TO	to	,	,	,		couple		You	PP	you
watch	VB		but	CC	but	argued	VBD		're	VBP	be
	watch		the	DT	the		argue		lucky	JJ	lucky
the	DT	the	steady	JJ		quietly	RB		!	SENT	!
river	NN	river		steady			quietly		,	,	,
flow	NN	flow	hum	NN	hum	outside	IN		"	"	"
while	IN	while	of	IN	of		outside		he	PP	he
a	DT	a	traffic	NN		a	DT	a	said	VBD	say
group	NN			traffic		small	JJ	small	,	,	,
	group		from	IN	from	hotel	NN	hotel	"	``	"
of	IN	of	the	DT	the	.	SENT	.	there	EX	there
children	NNS	child	nearby	JJ		"	``	"	's	VBZ	be
laughed	VBN	laugh		nearby		Did	VBD	do	only	RB	only
and	CC	and	road	NN	road	you	PP	you	one	CD	one
played	VBN	play	made	VBD		remember		VBP	room	NN	room
near	IN	near		make			remember		left	VBD	leave
the	DT	the	it	PP	it	to	TO	to	.	SENT	.
bank	NN	bank	almost	RB		book	VB	book	"	"	"
.	SENT	.		almost		the	DT	the	The	DT	the
The	DT	the	impossible	JJ		room	NN	room	old	JJ	old
light	JJ	light		impossible		?	SENT	?	man	NN	man
breeze	NN		.	SENT	.	"	"	"	watched	VBD	
	breeze		With	IN	with	she	PP	she		watch	
carried	VBD	carry	a	DT	a	asked	VBD	ask	them	PP	them
the	DT	the	sigh	NN	sigh	,	,	,	disappear		VBP
scent	NN	scent	,	,	,	and	CC	and		disappear	
of	IN	of	he	PP	he	he	PP	he	inside	RB	
wet	JJ	wet	leaned	VBD	lean	smiled	VBD	smile		inside	
grass	NN	grass	back	RB	back	weakly	RB		.	SENT	.
,	,	,	on	IN	on		weakly		Later	RBR	later
and	CC	and	the	DT	the	,	,	,	,	,	,
a	DT	a	bench	NN		holding	VBG	hold	a	DT	a
few	JJ	few		bench		up	RP	up	group	NN	
leaves	NNS	leaf	,	,	,	a	DT	a		group	
began	VBD	begin	closing	VBG	close	crumpled	JJ		of	IN	of
to	TO	to	his	PP\$	his		crumpled		workers	NNS	
dance	VB		eyes	NNS	eye	receipt	NN			worker	
	dance		for	IN	for		receipt		arrived	VBN	
lazily	RB	lazily	a	DT	a	.	SENT	.		arrive	
on	IN	on	moment	NN		The	DT	the	from	IN	from
the	DT	the		moment		owner	NN		the	DT	the
surface	NN		just	RB	just		owner		factory	NN	
	surface		to	TO	to	,	,	,		factory	
of	IN	of	feel	VB	feel	a	DT	a	,	,	,
the	DT	the	the	DT	the	thin	JJ	thin	carrying	VBG	carry
water	NN		warmth	NN		man	NN	man	boxes	NNS	box
	water		warmth			with	IN	with	too	RB	too

heavy	JJ		;	:	;	beside	IN		,	,	,
	heavy		life	NN	life		beside		he	PP	he
to	TO	to	,	,	,	the	DT	the	clapped	VBD	clap
bear	VB	bear	he	PP	he	river	NN	river	softly	RB	softly
for	IN	for	thought	VBD	think	.	SENT	.	,	,	,
long	RB	long	,	,	,	The	DT	the	whispered		VBD
.	SENT	.	always	RB		town	NN	town		whisper	
One	CD	one		always		was	VBD	be	"	``	"
of	IN	of	found	VBD	find	quiet	JJ	quiet	thank	VB	
them	PP	them	a	DT	a	now	RB	now		thank	
tried	VBD	try	rhythm—even	NN		,	,	,	you	PP	you
to	TO	to		<unknown>		except	IN		,	,	,
park	VB	park	in	IN	in		except		"	"	"
the	DT	the	the	DT	the	for	IN	for	and	CC	and
truck	NN	truck	smallest	JJS	small	a	DT	a	kept	VBD	keep
under	IN		moments	NNS		street	NN		walking	VBG	walk
	under			moment			street		.	SENT	.
an	DT	an	.	SENT	.	musician	NN		Above	IN	
old	JJ	old	As	IN	as		musician			above	
oak	NN	oak	the	DT	the	playing	VBG	play	him	PP	him
tree	NN	tree	sun	NN	sun	an	DT	an	,	,	,
,	,	,	began	VBD	begin	old	JJ	old	the	DT	the
while	IN	while	to	TO	to	tune	NN	tune	first	JJ	first
another	DT		set	VB	set	on	IN	on	stars	NNS	star
	another		,	,	,	his	PP\$	his	had	VBD	have
knelt	VB	kneel	the	DT	the	guitar	NN		already	RB	
to	TO	to	light	NN	light		guitar			already	
repair	VB		turned	VBD	turn	.	SENT	.	appeared		VBN
	repair		golden	JJ		The	DT	the		appear	
the	DT	the		golden		man	NN	man	,	,	,
flat	JJ	flat	,	,	,	stopped	VBD	stop	clear	JJ	clear
tire	NN	tire	painting	VBG	paint	,	,	,	and	CC	and
.	SENT	.	the	DT	the	listening	VBG	listen	sharp	JJ	
The	DT	the	water	NN		.	SENT	.		sharp	
sound	NN			water		For	IN	for	against	IN	
	sound		in	IN	in	a	DT	a		against	
of	IN	of	shades	NNS		second	JJ		the	DT	the
their	PP\$	their		shade			second		fading	JJ	
tools	NNS	tool	of	IN	of	,	,	,		fading	
echoed	VB	echo	fire	NN	fire	he	PP	he	blue	NN	blue
through	IN		.	SENT	.	felt	VBD	feel	.	SENT	.
	through		He	PP	he	young	JJ		He	PP	he
the	DT	the	packed	VBD	pack		young		smiled	VBD	smile
park	NN	park	his	PP\$	his	again—free	NN		,	,	,
like	IN	like	things	NNS	thing		<unknown>		tightened	VBD	
a	DT	a	slowly	RB		of	IN	of		tighten	
short	JJ	short		slowly		time	NN	time	his	PP\$	his
melody	NN		,	,	,	,	,	,	coat	NN	coat
	melody		then	RB	then	weight	NN		,	,	,
of	IN	of	walked	VBD	walk		weight		and	CC	and
iron	NN	iron	home	NN		,	,	,	disappeared		VBD
and	CC	and		home		and	CC	and		disappear	
effort	NN	effort	along	IN	along	worry	VBP	worry	into	IN	into
.	SENT	.	the	DT	the	.	SENT	.	the	DT	the
The	DT	the	narrow	JJ		When	WRB	when	night	NN	night
old	JJ	old		narrow		the	DT	the	.	SENT	.
man	NN	man	path	NN	path	song	NN	song			
smiled	VBD	smile				ended	VBD	end			

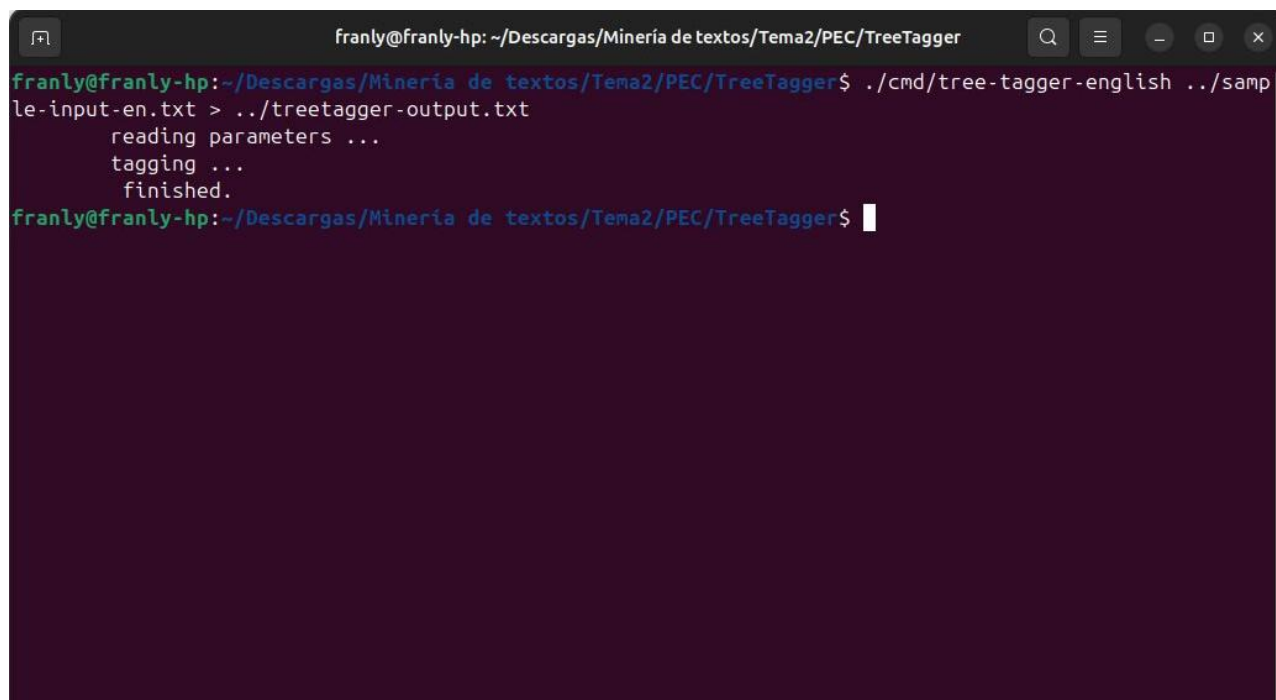
Análisis de las etiquetas de TreeTagger.

Token	Etiqueta dada	Etiqueta esperada	Explicación	Contexto
flow	NN	VB	Verbo base después de watch (“to watch the river flow”); no es un sustantivo.	...decided to watch the river flow while a group of children laughed and played...
laughed	VDN	VBD	Verbo principal en pasado; no participio.	...the river flow while a group of children laughed and played near the bank.
played	VDN	VBD	Verbo principal en pasado; no participio.	...children laughed and played near the bank.
kind	NN	JJ	Modificador de “eyes” (“kind eyes”); no sustantivo.	...a thin man with kind eyes...
left	VBD	JJ	Adjetivo predicativo (“one room left”); no forma verbal.	...there’s only one room left.
disappear	VBP	VB	Verbo en infinitivo tras watched; no forma finita.	...the old man watched them disappear inside.
arrived	VDN	VBD	Verbo principal en pasado; no participio.	...a group of workers arrived from the factory...
knelt	VDN	VBD	Pasado simple irregular; no participio.	...another knelt to repair the flat tire.
echoed	VDN	VBD	Verbo principal en pasado; no participio.	...their tools echoed through the park...
home	NN	RB	Complemento adverbial de dirección (“walked home”); no sustantivo.	...then walked home along the narrow path...
second	JJ	NN	Nombre de tiempo (“for a second”); no adjetivo.	For a second, he felt young again...
worry	VBP	NN	Coordinado con sustantivos (“time, weight, and worry”); no verbo.	...free of time, weight, and worry.

Tabla 4. Análisis sintáctico de las etiquetas generadas por TreeTagger.

Para el análisis de TreeTagger aplicamos la misma técnica anterior. Tras pasar por un analizador de Python y examinar su salida, descubrimos un total de 393 tokens, de los cuales 12 son errores y el resto, 381, son correctos. Su precisión lingüística estimada es del 97%.

TreeTagger mostró un comportamiento estable y eficiente durante el proceso, generando una salida clara con la estructura esperada de tres columnas: token, categoría gramatical y lema. Su velocidad de procesamiento y su bajo consumo de recursos lo hacen especialmente adecuado para corpus extensos o tareas de minería de textos donde la lematización es un requisito. Los errores detectados se concentran principalmente en la distinción entre tiempos verbales irregulares (VBN/VBD) y en ciertos usos adjetivales posnominales, lo que refleja las limitaciones propias de su modelo estadístico basado en árboles de decisión. Aun así, la consistencia del etiquetado confirma que TreeTagger mantiene un rendimiento alto y fiable para textos en inglés estándar.



```
franly@franly-hp: ~/Descargas/Minería de textos/Tema2/PEC/TreeTagger
franly@franly-hp:~/Descargas/Minería de textos/Tema2/PEC/TreeTagger$ ./cmd/tree-tagger-english ../sample-input-en.txt > ../treetagger-output.txt
    reading parameters ...
    tagging ...
    finished.
franly@franly-hp:~/Descargas/Minería de textos/Tema2/PEC/TreeTagger$
```

Imagen 2. Salida por consola de la ejecución del etiquetador Tree Tagger.

Análisis comparativo: TreeTagger vs Stanford POS Tagger.

Tras utilizar el mismo texto de entrada en inglés, un fragmento narrativo de 329 palabras, se ejecutaron ambos etiquetadores desde la terminal de Linux (Ubuntu), redirigiendo la salida a un archivo (*.txt) para su posterior análisis con Python.

En ambos casos, el etiquetado fue exitoso y no se detectaron fallos en la ejecución ni pérdida de datos.

Stanford POS Tagger demostró una mayor sensibilidad al contexto sintáctico, especialmente en la diferenciación de tiempos verbales y en la identificación precisa de categorías como verbos en pasado simple (VBD) frente a participios pasivos (VBN), así como en la correcta asignación de etiquetas a adverbios, preposiciones y determinantes según su posición en la oración. Su modelo de Máxima Entropía, basado en inferencia estadística y entrenado sobre el corpus Penn Treebank, le permite aprovechar el contexto de palabras adyacentes para resolver ambigüedades, mostrando un comportamiento consistente con las reglas del inglés estándar contemporáneo.

TreeTagger, por su parte, destacó por su rapidez y estabilidad. Su arquitectura basada en árboles de decisión y trigramas probabilísticos produce una salida ligera y fácilmente interpretable, incluyendo la lematización de cada término, lo que facilita el análisis posterior en tareas de minería de textos y conteo de frecuencias léxicas. No obstante, mostró cierta confusión entre las etiquetas VBN y VBD en formas verbales irregulares como knelt o arrived, además de asignaciones menos precisas en adjetivos posnominales y nombres usados con función adjetiva. Estas diferencias son esperables dado que su modelo, aunque eficiente, utiliza un conjunto más reducido de rasgos contextuales que el de Stanford.

Ambos alcanzan una precisión cercana al 97%, lo que sugiere que el texto de entrada es afín a estructuras sintácticas estándar del inglés y que ambos modelos están bien entrenados para la variedad americana. La similitud en los resultados demuestra que, pese a sus diferencias metodológicas, ambos etiquetadores mantienen un nivel de rendimiento elevado y consistente. El Stanford Tagger presenta una ligera ventaja en precisión contextual y en la correcta clasificación de tiempos verbales, mientras que TreeTagger ofrece una ejecución más rápida, menor consumo de recursos y la ventaja añadida de proporcionar lemas junto con las etiquetas.

CONCLUSIÓN

El análisis de los distintos grafos ha permitido comprobar cómo las propiedades estructurales básicas de una red —grado, clustering, diámetro, longitud media de los caminos y número de componentes conexas— proporcionan información clave sobre su naturaleza y comportamiento global.

En el caso de EuroSiSGenerale, se observó una red dirigida y ponderada, con hubs claramente definidos y una estructura de tipo pequeño-mundo con tendencia libre de escala, propia de las redes sociales e institucionales.

El grafo RetGraph mostró una distribución del grado altamente sesgada y clustering bajo, reflejando la naturaleza dispersa y asimétrica de las co-ocurrencias léxicas.

LesMiserables presentó una red no dirigida y ponderada, con clustering elevado y caminos cortos, característica de las redes sociales narrativas, donde los protagonistas funcionan como nodos centrales.

Por último, BoundaryCountries representó una red espacial y casi planar, con módulos regionales bien definidos y un comportamiento no aleatorio pero no libre de escala, dominado por la restricción geográfica.

En conjunto, la práctica ha permitido contrastar los modelos teóricos —aleatorio, pequeño-mundo y libre de escala— con ejemplos reales de redes complejas, mostrando que la mayoría de las redes naturales y sociales no son puramente aleatorias, sino que presentan estructura, jerarquía y modularidad.

El ejercicio con Gephi ha sido una oportunidad valiosa para visualizar, medir y comprender la diversidad estructural de las redes, y para afianzar los conceptos fundamentales de la teoría de grafos aplicados al procesamiento del lenguaje y al análisis de sistemas complejos.

GLOSARIO

Grafo:

Representación matemática de una red formada por un conjunto de nodos (vértices) y aristas (enlaces) que indican relaciones entre ellos. Puede ser dirigido o no dirigido, y con o sin pesos.

Nodo (vértice):

Entidad individual dentro del grafo (por ejemplo, una persona, palabra, país o personaje).

Arista (enlace):

Conexión entre dos nodos que representa una relación o interacción. Puede tener dirección (A \rightarrow B) y/o peso (intensidad o frecuencia).

Grado (degree):

Número de aristas que inciden en un nodo. En grafos dirigidos se distingue entre grado de entrada (indegree) y grado de salida (outdegree).

Distribución del grado:

Función que describe la probabilidad de que un nodo tenga un determinado grado. Permite identificar si la red es aleatoria, de tipo small-world o scale-free.

Coeficiente de clustering:

Medida que indica la densidad de conexiones locales; representa la probabilidad de que los vecinos de un nodo también estén conectados entre sí.

Longitud media del camino:

Promedio de las distancias más cortas entre todos los pares de nodos. Mide la eficiencia global de la red.

Diámetro de la red:

Distancia máxima (camino más corto) entre dos nodos cualesquiera. Indica el “tamaño efectivo” de la red.

Componente conexa:

Conjunto de nodos en el que todos están conectados por algún camino.

En grafos dirigidos se distingue entre componentes fuertemente conexas (SCC) y débilmente conexas.

Red de pequeño mundo (small-world):

Red con caminos cortos entre nodos y alto clustering. Suele aparecer en redes sociales o biológicas.

Red libre de escala (scale-free):

Red cuya distribución de grados sigue una ley de potencias, es decir, pocos nodos concentran la mayoría de las conexiones (hubs).

Hub:

Nodo con un grado significativamente mayor al promedio; actúa como punto de conexión central en la red.

Componente gigante:

La componente conexa más grande del grafo, que contiene la mayoría de los nodos y domina la estructura global de la red.

BIBLIOGRAFÍA

Radev, D., & Mihalcea, R. (2011). *Graph-based Natural Language Processing and Information Retrieval*. Cambridge University Press.

Araujo, L., & Martínez-Romo, J. (2024). *Tema 2: Redes y grafos, definiciones y conceptos* [Material docente]. Universidad Nacional de Educación a Distancia (UNED), Máster en Inteligencia Artificial.

Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.

Radev, D., & Mihalcea, R. (2011). *Graph-based Natural Language Processing and Information Retrieval*. Cambridge University Press.

Lo primero que llama la atención es que las relaciones no dirigidas apuntan a autoenlaces entre ellas. Los pesos para esas relaciones son bajos, lo que sugiere una relación débil que podría sugerir actividades internas, subprogramas, proyectos internos o colaboración entre departamentos.