

Proyecto Fin de Carrera

Detección automática de estructuras narrativas en
guiones cinematográficos

Autor

Juan Fco. López de la Osa

Directores

Manuel González Bedia

Miguel Aguilera Lizárraga

Departamento de Informática e Ingeniería de Sistemas

Escuela de Ingeniería y Arquitectura

Universidad de Zaragoza

Marzo 2016

A mi madre, por su apoyo incondicional.

Agradecimientos

Antes de nada quiero darle las gracias a Manolo por lanzarse en esta pequeña y rara aventura. Para mí ha sido un año fascinante, lleno de nuevas y estimulantes cosas por aprender. Y no puedo pensar en haber tenido mejor guía que él para esta travesía.

Gracias también a Miguel, por su paciencia y sus aportaciones.

Y a Goyo por compartir con nosotros su trabajo.

También a mis amigos de la carrera. Muy especialmente a Álvaro y Alba. Por todos los innumerables buenos ratos a lo largo de estos años y por haber estado ahí también en estos momentos previos a la entrega.

Y por supuesto a mis amigos de toda la vida, que siempre están cuando se les necesita.

Y finalmente, pero no por ello menos importante, a mi familia. Sobre todo a mis padres, a quienes nunca agradezco lo suficiente el apoyo del día a día.

A todos, gracias por estar ahí.

Resumen

En los últimos años las aplicaciones de narración automática han proliferado substancialmente tanto si observamos la cantidad de herramientas software disponibles como por el número de ámbitos de aplicación donde la generación automática de historias empieza a ser exhaustiva - redacción de noticias automáticas por bots, informes financieros automáticos con el fin de acelerar la toma de decisiones en gerencia, visualización de información combinada con procesos de "data analysis/Big data" para su rápida interpretación, etc.

Aún a pesar de las notables mejoras de las herramientas y su éxito a nivel operativo, lo cierto es que desde un punto de vista teórico es constatable que sus resultados dependen mucho del dominio específico generando soluciones poco generalizables ni escalables.

Superar estas deficiencias implica realizar análisis más teóricos sobre los mecanismos de funcionamiento de las narraciones. Existe un ámbito académico centrado en esta tarea que pretende, por un lado, profundizar en cuáles son las estructuras de los relatos en su sentido más amplio (escritos y/o visuales) en términos de ritmo, variabilidad, lenguaje metafórico, condiciones para la verosimilitud, carácter de personajes, tipo de tramas, etc. Por otra parte, pretende convertir toda esta información en mecanismos operativos para su posterior implementación.

Este proyecto se circunscribe en este ámbito de detección de características de estructuras narrativas, aplicado a guiones cinematográficos. La elección de este género narrativo tiene varias ventajas: (i) tiene una estructura estandarizada en secuencia, (ii) mantiene un nivel descriptivo con pocas figuras narrativas complejas (lo que simplificará el análisis), (iii) nos permite tener una dimensión temporal explícita (que es más fácil de extraer que en otros productos narrativos como la novela o los cuentos).

En concreto, en este trabajo se plantea:

- Realizar un estado del arte sobre las metodologías y tecnologías actuales para el desarrollo y análisis de estructuras narrativas generadas automáticamente.
- Análisis de las estructuras de "storytelling" desde un enfoque multidisciplinar (qué nos hace percibir y entender que estamos ante una narración y no ante un informe, o un documento expositivo).
- Definir índices que permitan caracterizar las estructuras detectadas en el punto anterior (organización de la historia en diferentes escalas como, por ejemplo, secuencia-escena-acto), despliegue temporal del relato (por ejemplo, la evolución de los temas) o estructuras y caracterización de relaciones entre los personajes de la misma.
- Utilizar esta batería de métricas para su aplicación en tareas de descripción y clasificación en el ámbito de los guiones cinematográficos.

Índice general

Agradecimientos	III
Resumen	IV
Índice general	V
1. Introducción y contexto	1
• 1.1. Objetivos	2
• 1.2. Tecnologías utilizadas	3
• 1.3. Contenido de la memoria	3
2. Estado del arte	5
• 2.1. Narratología y Narrativa computacional	5
• 2.2. Algunas soluciones disponibles	6
• 2.3. Enfoque de este proyecto	8
3. Métricas para la detección de estructura	9
• 3.1. Preprocesamiento	11
• 3.2. Índices sobre el género	11
• 3.3. Índices sobre personajes	11
• 3.4. Índices sobre los temas	13
- 3.4.1. Palabras relevantes	13
- 3.4.2. Diccionario	13
o 3.4.2.1. Ideas	14
o 3.4.2.2. Sentidos y palabras	18
- 3.4.3. Entropía	19
- 3.4.4. Entropía cruzada	20
• 3.5. Índices sobre las tramas	21
• 3.6. Índices narrativos	22
- 3.6.1. Detección de escenas	22
o 3.6.1.1. Medidas de temas para escenas	22
o 3.6.1.2. Medidas de personajes para escenas	23
o 3.6.1.3. División en escenas	23
- 3.6.2. Detección de actos	24
o 3.6.2.1. Medidas de temas para actos	24
o 3.6.2.2. Medidas de personajes para actos	25
o 3.6.2.3. Medidas de escenas para actos	25
o 3.6.2.4. División en actos	25
4. Pruebas experimentales	27
• 4.1. Índices sobre el género	28
• 4.2. Índices sobre los personajes	29
• 4.3. Índices sobre los temas	30
• 4.4. Índices sobre las tramas	31
• 4.5. Índices narrativos	32
5. Conclusión	33
• 5.1. Conclusiones técnicas	33
• 5.2. Conclusiones personales	34

• 5.3. Líneas abiertas	34
6. Carga y gestión del proyecto	37
• 6.1. Gestión de tiempos	37
• 6.2. Tiempo invertido por tareas	39
Referencias	41
Anexos	45
A. Preprocesado	47
• A.1. Aprendizaje del sangrado	49
• A.2. Parser	50
B. Comparativa de algoritmos para la obtención de temas	53
• B.1. Algoritmo de fuerza bruta	54
- B.1.1. Ejecución por secuencias	54
- B.1.2. Ejecución a nivel global	56
• B.2 Algoritmo genético	58
C. El guión literario cinematográfico	59
• C.1. Secuencia, escena y acto	59
• C.2. Cambiar de acto	60
• C.3. Una página, un minuto	60
D. Resultados de bajo nivel	61
E. Blade Runner	65
• E.1. Personajes	65
F. Cómo entrenar a tu dragón	67
• F.1. Personajes	67
G. Reservoir Dogs	69
• G.1. Personajes	69
H. The Matrix	71
• G.1. Personajes	71
I. Seven	73
• I.1. Personajes	73
J. Star Wars : Episodio V – El Retorno del jedi	75
• J.1. Personajes	75
K. Star Wars: Episodio I – El Retorno del jedi	77
• K.1. Personajes	77
L. El laberinto del fauno	79
• L.1. Personajes	79
• L.2. Temas	81
• L.3. Entropía	81
- L.3.1 Entropía temas	81
- L.3.2. Entropía personajes	82
• L.4. Grafos	83
• L.5. Escenas	87
• L.6. Actos	88
- L.6.1 Cambio Acto I-II	88
- L.6.2. Cambio Acto II-III	89
M. Tesis	91
• M.1. Personajes	91

• M.2. Temas	93
• M.3. Entropía	93
- M.3.1 Entropía temas	93
- M.3.2. Entropía personajes	94
• M.4. Grafos	95
• M.5. Escenas	99
• M.6. Actos	100
- M.6.1 Cambio Acto I-II	100
- M.6.2. Cambio Acto II-III	101
N. La isla mínima	103
• N.1. Personajes	103
• N.2. Temas	105
• N.3. Entropía	105
- N.3.1 Entropía temas	105
- N.3.2. Entropía personajes	106
• N.4. Grafos	107
Ñ. Actas, estructura original	111
• Ñ.1. Personajes	111
• Ñ.2. Temas	113
• Ñ.3. Entropía	113
• N.4. Grafos	133
O. Actas, estructura uniforme	115
• O.1. Personajes	115
• O.2. Temas	117
• O.3. Entropía	117
- 0.3.1 Entropía temas	117
- 0.3.2 Entropía personajes	118
• O.4. Grafos	119
P. Actas, estructura narrativa	120
• P.1. Personajes	120
• P.2. Temas	122
• P.3. Entropía	122
- P.3.1 Entropía temas	122
- P.3.2 Entropía personajes	123
• P.4. Grafos	124
Q. Para Elías	125
• Q.1. Personajes	125
• Q.2. Temas	127
• Q.3. Entropía	127
• Q.4. Grafos	127
• Q.5. Escenas	130
• Q.6. Actos	131
- Q.6.1 Cambio Acto I-II	131
- Q.6.2. Cambio Acto II-III	132
• Q.7. Guión literario	133

1

Introducción y contexto

¿Qué tienen de especial las narraciones?, ¿Qué hace que una historia, incluso aquellas que fueron escritas hace siglos y pensadas para medios tan distintos a los que ahora estamos acostumbrados, las identifiquemos como un relato?

El diseño de sistemas computacionales capaces de crear y/o interpretar el contenido narrativo de un texto es un área cada vez con mayor interés. Los campos de aplicación son numerosos (generación automática de resúmenes (Ouyang et al, 2014), de discursos políticos (Kassarnig, 2016), alternativa a la visualización en el tratamiento de grandes masas de datos (narrativescience.com), etc. En todos ellos, la dificultad está en determinar cómo encontrar y explotar aquellos elementos que hacen que una estructura textual adquiera "forma narrativa".

Las estructuras con "componente narrativo" no son simplemente estructuras secuenciales de hechos. Autores como (Bruner, 1980) advierten que los relatos implican la articulación de una realidad en base a ciertos puntos de interés de maneras muy particulares. Un flujo indiferenciado de estructuras causa-efecto se convierte en una historia solo si queda constituida en torno a ciertos "hitos" siguiendo unos patrones determinados (Dewey, 2007).

Existen numerosos métodos para la detección (extractiva) de significados en textos mediante la selección de frases, oraciones, etc. de la fuente original que se consideren relevantes (como modelos basados en n-gramas, modelos de "bag of words" o métodos de detección de secuencias de palabras frecuentes y maximales (Soto et al, 2009). Sin embargo, de manera general, estos mecanismos se plantean cómo detectar patrones textuales basados en análisis de reconocimiento lingüístico sin abordar niveles estructurales más profundos. Por esto se reconoce que, en la actualidad, no existe un modelo universalmente aceptado acerca de cómo un sistema computacional puede reconocer o utilizar narraciones que ha identificado o creado (el procesamiento de historias por humanos requiere la proyección de varios de nuestros intereses y capacidades cognitivas). Aunque no

conozcamos un canon respecto a qué patrones constituyen una historia, la realidad es que los humanos somos capaces de distinguir dónde se encuentra un buen relato. Nuestro objetivo en este proyecto es abordar el problema sobre si una máquina podría también ser capaz de detectarlo reconociendo las estructuras que configuran una historia en base a análisis textuales.

Puede decirse que los objetivos del proyecto entrarían dentro del ámbito de la “Narrativa Computacional”, una rama relativamente joven de una nueva disciplina conocida como “Humanidades digitales” en las que convergen los intereses tradicionales de las humanidades y la informática, procurando que los conocimientos de las nuevas tecnologías sirvan para abordar los problemas propios de las ciencias humanas pero mediante enfoques distintos y con acercamientos no convencionales. No se trata sólo de integrar nuevas herramientas (digitales) al campo de las humanidades, sino de abordar retos en la disciplina que no sean posibles lograr sin usar los avances en computación.

El interés por esta temática es patente tanto en el ámbito de la investigación como por recientes iniciativas provenientes de la industria. A modo de ejemplo, las líneas de investigación del grupo “Imagination, Computation, and Expression Laboratory” del MIT (groups.csail.mit.edu/icelab/) o empresas como “Narrative Science” (www.narrativescience.com/) o Yseop (yseop.com/). Los primeros declaran que sus intereses investigadores se centran en explorar modelos de narrativa computacional, de representación computacional de la identidad o de las interrelaciones entre computación y cultura. En el caso de las empresas mencionadas, el objetivo fundamental es poder generar documentos que representen la “historia” que esconden grandes volúmenes de información, mostrándola de una forma accesible y significativa, en formato narrativo. Estas empresas se centran hasta ahora principalmente en unos pocos nichos de mercado —deportes, finanzas o también en el campo de la información política— en los que lo que se muestra suele seguir una misma pauta, gira en torno a datos y estadísticas.

A parte de las líneas y objetivos presentados hasta ahora, el potencial de una mayor y mejor comprensión operativa (computacional) de la historias también puede tener aplicaciones directas y significativas en el mundo de los videojuegos, donde el desarrollo de argumentos profundos e inmersivos similares a los de las películas está cada vez más en boga y supone un claro nicho de negocio. Procesos de selección y cribado de guiones de cine —para los cuales las grandes productoras dedican muchos recursos— también podrían verse simplificados en cierta medida por este conocimiento.

1.1. Objetivos

En este trabajo nos planteamos definir y validar una batería de índices que permitan detectar “relatos” automáticamente. Se definirá y justificará cómo obtener información sobre

- El género.
- Los tipos de personaje
- El despliegue de temáticas
- Las relaciones entre personajes y situaciones que configuran la trama y, por último
- La estructura narrativa global.

El problema concreto que nos planteamos tiene que ver con la “estructura de sentido” que constituyen las historias. En concreto, nuestra aproximación al problema la realizaremos sobre historias plasmadas en guiones cinematográficos, que presentan unas características especiales que

(facilitan el trabajo y que) detallaremos más adelante. La validación vendrá a partir de una base de guiones de películas de distinta índole (thrillers, animación, ciencia ficción, acción, fantástica, policiaca, etc.). También se contrastarán los resultados con la aplicación de las mismas métricas en textos que no constituyen historias (en particular, unas actas del diario de sesiones parlamentarias) donde encontraremos estructuras textuales pero no aquellas que aparecen en el análisis de relatos.

El proyecto fin de carrera también buscará indirectamente los siguientes hitos:

- De manera general, conocer el marco teórico y el estado del arte en el campo de la Narrativa Computacional para garantizar que la propuesta fundamentada es sólida y a la vez novedosa.
- Desde un punto de vista teórico, caracterizar un conjunto de índices que sean capaces de identificar los rasgos y estructuras que tienen aquellos textos que identificamos como historias, lo que permitirá explorar una de nuestras capacidades más genuinas, “el razonamiento narrativo” (Bruner, 1980)¹.
- Desde un punto de vista aplicado, desarrollar el prototipo de un programa que sea capaz de extraer los elementos básicos de una estructura narrativa.

1.2. Tecnologías utilizadas

Básicamente el desarrollo se ha realizado usando Python, un lenguaje multiparadigma, ampliamente utilizado en la actualidad y que dispone de muchas librerías implementadas –como, por ejemplo, NLTK para Procesamiento de Lenguaje Natural o Networkx para grafos, ambos paquetes de interés para nuestro trabajo-. Independientemente de razones técnicas y características funcionales, la posibilidad de investigar un nuevo lenguaje de programación, ha constituido una de las motivaciones extra de este proyecto. En menor medida, también se han desarrollado scripts bash para Unix y algunos módulos desarrollados en código Java.

1.3. Contenido de la memoria

A continuación se presenta un resumen de los contenidos del trabajo:

- Capítulo 1. Motivación del interés por la cuestión tratada, objetivos que nos planteamos y presentación de la estructura de la presente memoria.
- Capítulo 2. Relación de conceptos básicos y descripción del estado del arte. Recursos existentes en el ámbito de la informática. Problemas y soluciones que detecta y dónde se ubica el proyecto.

¹ Esta propiedad, en contraposición al pensamiento deductivo lógico es algo que argumentó de manera pionera Jerome Bruner. Bruner fue uno de los pioneros del movimiento de Inteligencia Artificial aunque pronto se separa de la línea. Su publicación, en 1956, “A Study of Thinking” junto a Goodnow y Austin, marca – aunque la fecha es próxima al origen del movimiento de la Inteligencia Artificial- notables diferencias que separan los intereses iniciales de la investigación de Bruner de los de Newel y Simon, diferencias que se van haciendo más pronunciadas en escritos posteriores. Bruner identifica “dos modos de construir la realidad”, que “son irreducibles entre sí”, que denomina modos de razonamiento “narrativo” y “paradigmático o lógico-científico”. Este último, dice Bruner, verifica la verdad “mediante un llamamiento final a procedimientos para establecer pruebas formales y empíricas, mientras que los modos narrativos de razonamiento quizás no establezcan la verdad en sí, sino la verosimilitud”, es decir, “relatos creíbles (aunque no necesariamente ciertos)”. A través de la narrativa se pueden configurar comprensiones o interpretaciones que den explicaciones diferentes de tipo causa-efecto.

- Capítulo 3. Explicación de los índices desarrollados.
- Capítulo 4. Exposición de las pruebas realizadas: características, resultados y discusión.
- Capítulo 5. Resumen de las conclusiones tanto técnicas como personales, así como una lista de las líneas abiertas por el proyecto.
- Capítulo 6. Carga de trabajo.
- Referencias. Fuentes de información consultadas para realizar el proyecto.
- Anexos
 - Anexo A. Explicación de la pipeline de preprocesamiento de un guión. Detalles de la fase de aprendizaje del sangrado y el parser.
 - Anexo B. Comparativa algoritmos de desambiguación del sentido.
 - Anexo C. Estructura del guión cinematográfico.
 - Anexo D. Resultados sobre el género o de bajo nivel.
 - Anexo E. Resultados de los índices para la película “Blade Runner”.
 - Anexo F. Resultados de los índices para la película “Cómo entrenar a tu dragón 2”.
 - Anexo G. Resultados de los índices para la película “Reservoir Dogs”.
 - Anexo H. Resultados de los índices para la película “The Matrix”.
 - Anexo I. Resultados de los índices para la película “Seven”.
 - Anexo J. Resultados de los índices para la película “Star Wars: Episodio V – El Imperio Contraataca”.
 - Anexo K. Resultados de los índices para la película “Star Wars: Episodio VI – El Retorno del Jedi”.
 - Anexo L. Resultados de los índices para la película “El laberinto del fauno”.
 - Anexo M. Resultados de los índices para la película “Tesis”.
 - Anexo N. Resultados de los índices para la película “La isla mínima”
 - Anexo Ñ. Resultados de los índices para las actas del congreso manteniendo su estructura original.
 - Anexo O. Resultados de los índices para las actas del congreso con una estructura de secuencias uniforme.
 - Anexo P Resultados de los índices para las actas del congreso con la estructura narrativa de “La isla mínima”.

2

Estado del Arte

En esta sección pretendemos mostrar algunas de las más importantes herramientas lingüísticas recogidas en la literatura que implementan métodos de análisis lingüístico (a distintos niveles) a partir de un texto de entrada. Encontraremos desde analizadores sintácticos, etiquetadores gramaticales, sistemas para la división del texto en oraciones, etc. El objetivo es mostrar dónde se ubicaría nuestra propuesta. El desarrollo de estas técnicas y métodos se escapa del ámbito de este proyecto, por lo que se ha optado por evaluar sólo un conjunto de herramientas disponibles, seleccionando aquellas que nos han parecido más ilustrativas.

2.1. Narratología y Narrativa computacional

La Narratología es la ciencia dedicada al estudio de la estructura de las narraciones y de la manera en la cual los humanos las entendemos y las usamos. Este enfoque estructuralista establece la base de la Narratología moderna, aunque el estudio de las historias, los mitos y los relatos puede retrotraernos hasta tiempos antiquísimos -ver (Aristóteles, 2007)²- en los que ya se postula que construir historias pasa por seleccionar algunas unidades sobre los acontecimientos del mundo, determinar las más importantes entre las escogidas y ordenarlas en una trama. La Narratología moderna tiene sus orígenes en los análisis estructuralistas de los textos, basados en el trabajo de

² A pesar de haber pasado tantos siglos, las reflexiones de Aristóteles siguen siendo intuiciones muy inspiradoras en el estudio de los relatos y, por tanto, lugar obligado al que acudir para comprender el estado actual de las mismas-. De entre todos los planteamientos en esta obra destacamos tan sólo un par: la definición de los caracteres -los personajes- con respecto a su capacidad de tomar decisiones y ser activos en el desarrollo de los acontecimientos y también las formas mediante las cuales se pretende conmover al público. Éstas son la peripecia, la agnosis y el lance patético, que Aristóteles define respectivamente como el cambio de dirección de los hechos ocurridos con respecto a lo que se esperaba que ocurriera, el cambio sobre el conocimiento de un personaje acerca del mundo y las acciones de naturaleza destructiva (como asesinatos, torturas, heridas...).

Barthes o Todorov y otros (Barthes and Duisit, 1975, Todorov, 1977). Más recientemente, ha pasado a ser un área de interés para la ciencia cognitiva, considerando las narraciones como fenómenos psicológicos (Herman, 2000), y su estudio en base a las capacidades mentales que se necesitan para generarlas e identificarlas.

En particular, por lo que respecta al interés de desarrollar algoritmos para identificar/construir historias, el área de la Narrativa Computacional surge con la misión de avanzar en la representación y procesado de las narraciones por ordenadores. Uno de los subcampos más estudiados de la Narrativa Computacional es la generación de historias. Aparece en los años 70 del siglo XX por el interés de Roger Schank, pionero del estudio del impacto de la narrativa en humanos desde un punto de vista formal y computable. Schank comenzó interesándose por las capacidades humanas que permite la creación de relatos y postuló la existencia de mecanismos de memoria que no estarían basados en procesos que manipulan datos mentales, sino en procesos de recuerdo y adaptación de historias previas como en el Razonamiento Basado en Casos (que se origina en el trabajo de Schank) (Schank, 1982). Casi a la vez, los primeros resultados de estos estudios provocaron la aproximación inversa: en vez de intentar entender los métodos humanos con la ayuda de modelos computacionales, se empezaron a plantear estudios acerca de cómo las máquinas podrían emularlos. Sin embargo, tras esta etapa inicial de interés en el ámbito de la Inteligencia Artificial aplicada a la narrativa, fue una línea abandonada durante décadas posteriores (con las excepciones de algunos sistemas como (Turner, 1992).

Sin embargo, en las últimas décadas se ha renovado el interés por el estudio y la implementación de la Inteligencia Narrativa (Mateas and Sengers, 1999) y se buscan métodos que logren la formalización teórica de la narrativa, el desarrollo de sistemas de generación de historias, sistemas de extracción de conocimiento humano narrativo, etc.

El campo en el que nos centramos en esta memoria se conoce como la evaluación computacional de narrativa. La evaluación y la comprensión de la narrativa han sido estudiadas principalmente por la Psicología, pero con el uso de las computadoras, se intentan definir/encontrar métricas para comparar diferentes historias y su calidad como historias.

La definición formal de narración en este ámbito se centra en el “estudio formal de tramas” de historias, pero no en características de estilo o lingüísticas más complejas (Riedl and Young, 2006). No hay interés en estudios sobre la detección y clasificación de las diferencias en relatos a partir de figuras retóricas, estilo, o tipología sino en todo lo contrario: en la detección de aquella estructura mínima que podría caracterizar todas ellas y distinguirlas de un texto que no constituya un relato.

2.2. Algunas soluciones disponibles

El dominio de los programas para la detección o generación de historias es relativamente denso. En esta sección pasamos a mostrar solo algunos de los recursos más interesantes que tenemos a nuestra disposición en el ámbito de la Narrativa Computacional. Como regla general, las herramientas disponibles tanto en la extracción de historias, generadores de resúmenes, etc. recae en: el área de la “Generación de lenguaje natural” (o NLG por sus siglas en inglés, cuyo objetivo es producir texto a partir de una información de entrada) o en el campo del “Procesamiento de Lenguaje Natural” (o NLP por sus siglas en inglés, más preocupado del análisis del texto que en su producción). Algunas de las líneas vinculadas a estos campos en las que nos hemos interesado han sido el estudio de técnicas de “Extracción de Información” (IE, por ejemplo, a partir de la selección de palabras relevantes del texto o el apoyo en sistemas externos de representación estructurada del

conocimiento, como el uso de “Thesaurus”) (ver apartado 3.4.2. Diccionario) o el campo del “Language Modelling”. Algunos ámbitos que también han sido explorados finalmente no han sido explotados en la versión final del proyecto. Hablamos, por ejemplo, de técnicas de “Relation extraction” y “Sentiment analysis” para el diseño de la red de relaciones entre personajes o de Clasificadores de textos para agrupar los guiones por características comunes (ver apartado 5.3 Líneas abiertas).

De interés histórico, a continuación destacamos algunas de las herramientas más relevantes a lo largo de la historia de la disciplina,

- **Novel writer:** Esta herramienta (Klein et al., 1973) genera historias relacionadas con thrillers explotando mediante modelos probabilísticos cambios en la narrativa y en la secuencia de acontecimiento.
- **Grammar:** Rummelhart y Joseph (Rummelhart 1975) propusieron un sistema generador de historias basado en el uso de gramáticas generativas. De este modo se busca capturar propiedades estructurales de las narraciones en forma de reglas.
- **TaleSpin** de (Meehan 1976) crea historias sobre criaturas del bosque, y se centra en un enfoque por metas y eventos de los personajes.
- **Universe:** desarrollado en (Lebowitz, 1985) genera guiones para programas de televisión combinando historias que se superponen.
- **MINSTREL:** Centrado en historias en torno a los caballeros de la mesa redonda (Turner, 1992) el programa se dirige explícitamente a explotar la creatividad en la generación de historias.
- **Brutus:** Se generan automáticamente cuentos tradicionales (Bringsjord y Ferrucci, 1999).
- **Virtual Story Teller** (Theune et al., 2006) se basa en el enfoque de TaleSpin pero mediante un sistema multiagente en que los personajes son agentes de software. Los agentes tienen conocimiento sobre reglas propias del mundo e incluye un agente narrador, que cuenta la historia en lenguaje natural.
- **Fabulist:** En este trabajo (Riedl, 2004) se busca añadir credibilidad mediante el uso de planificadores de orden parcial.
- **ProtoPropp:** se incluyen ontologías (Peinado et al., 2008) y se explotan lógicas de predicados.

Desde un punto de vista más transversal, destacamos algunas de las herramientas lingüísticas más usadas y disponibles para la comunidad interesada en realizar análisis lingüísticos a distintos niveles dado un texto de entrada.

- **Stanford Parser:** El analizador sintáctico de dependencias Stanford Parser (Klein y Manning, 2003), es un analizador estadístico de código libre que combina el análisis sintáctico de un texto con el análisis de dependencias.
- La herramienta **Gate**, del inglés General Architecture for Text Engineering (Cunningham et al., 2002), es una arquitectura genérica para la realización, de una forma fácil, visual y transparente para el usuario, de análisis lingüísticos complejos y que surge en respuesta a la necesidad de realizar un proceso de estandarización para el desarrollo de herramientas de análisis lingüísticos.
- **WordNet** es una extensa base de datos léxica para el inglés desarrollada bajo la dirección de George A. Miller e iniciada en el año 1985, en cuyo proyecto han participado diferentes organizaciones públicas y privadas (Miller, 1995). WordNet nació con dos propósitos. Por un lado, la creación de una combinación entre léxico y tesoro fácil de usar y de

comprender; por otro lado y más importante, proporcionar una herramienta básica para el desarrollo de aplicaciones complejas de análisis textual y procesamiento del lenguaje natural. Está organizado en torno a cinco categorías léxicas básicas: nombres, verbos, adjetivos, adverbios y elementos funcionales, de las cuales actualmente solo existen las cuatro primeras.

2.3. Enfoque de este proyecto

Las técnicas arriba indicadas y, en general, cualquier modelo de generación o extracción de sentido automática, se encuentra con el problema del análisis de información cualitativa (semántica, de sentido), que en principio son aspectos que deberían ser identificados, interpretados y modelizados desde marcos teóricos cualitativos (psicológicos, psicosociales, culturales, lingüísticos, semiológicos o retóricos, etc.) pero que deben tratarse a partir del análisis estadístico a partir de una fuente textual proporcionada. Por tanto, el procedimiento tradicional de la cuantificación de datos cualitativos es la categorización, la codificación y la tabulación. En términos generales, el Análisis Estadístico de Datos Textuales (ADT) utiliza procedimientos que implican contar las ocurrencias de las unidades verbales básicas (generalmente palabras) y operar algún tipo de análisis estadístico a partir de los resultados de tales recuentos.

Este es también el dominio de nuestro proyecto, aunque la orientación es diferente a la mayor parte de enfoques mencionados. Normalmente, el objetivo de los investigadores que trabajan con estas técnicas consiste en analizar bases de textos, encontrar patrones, clasificarlos, etc. con un fin aplicado (redacción de resúmenes, organización de noticias, etc.). En nuestro caso, el enfoque y las decisiones de diseño tienen una orientación más teórica: encontrar un panel de medidas que pudieran explicar la estructura de *qué es una historia*.

3

Métricas para la detección de estructura

En este capítulo se mostrarán los diferentes índices que hemos utilizado para hacer nuestro estudio. Dividiremos el trabajo en una serie de niveles con métricas asociadas a cada uno de ellos. Hablaremos de índices de nivel bajo, medio o alto en función del origen de los datos con los que se elaboran las métricas y la naturaleza de sus resultados.

- Los índices que formarán parte del nivel más bajo corresponden a medidas directas sobre el guión como algunas de las ya planteadas en los scripts: número de secuencias, número de secuencias interiores y exteriores, número de líneas de diálogo, número de líneas de descripción...
- En un segundo nivel se definirán índices relacionados con personajes y palabras. Los cálculos se realizan también en base a datos extraídos del propio guión pero los resultados empiezan a ser diferentes a los del caso anterior, en el sentido que tienen un carácter más cualitativo: se seleccionan personajes y palabras por su importancia.
- En un tercer nivel distinguimos dos fases: (i) un índice temático (la solución adoptada en este nivel hará que nos decantemos por el español como idioma principal de los guiones a estudiar, con todas las restricciones que ello acarrea)³ y (ii) índices de extracción de estructuras narrativas de mayor nivel que las secuencias, esto es, las escenas y los actos (ver más sobre la estructura del guion cinematográfico en Apéndice C). La agrupación de secuencias en escenas se hace en base a personajes y temas. Sólo personajes en caso de ser un guión en inglés -de ahí que, aun produciendo un resultado, resulte éste menos fundamentado-. La división del guión en actos se hace en base a personajes, temas –si procede según el idioma-, escenas y una corrección teórica del resultado.

³ En el apartado “Diccionario conceptual” del punto 3.4.2 se explican con detenimiento las razones. En cualquier caso, es importante hacer notar que tan sólo esta métrica y los cálculos de entropía son dependientes del idioma, todas las demás no lo son. Por eso también presentamos resultados de guiones en inglés. No obstante, nos centramos en los resultados en español, pues esta fase tiene mucho que aportar al resultado final. La incidencia de los temas en las métricas construidas sobre ellos las vemos a continuación.

Por último, también dentro de los índices de alto nivel, se realizan unas medidas de entropía, que pretenden darnos una foto de la incertidumbre con la que los temas se despliegan a lo largo del guión, y de información mutua, que ponen de relieve las relaciones más importantes que se establecen entre algunos de los elementos anteriores.

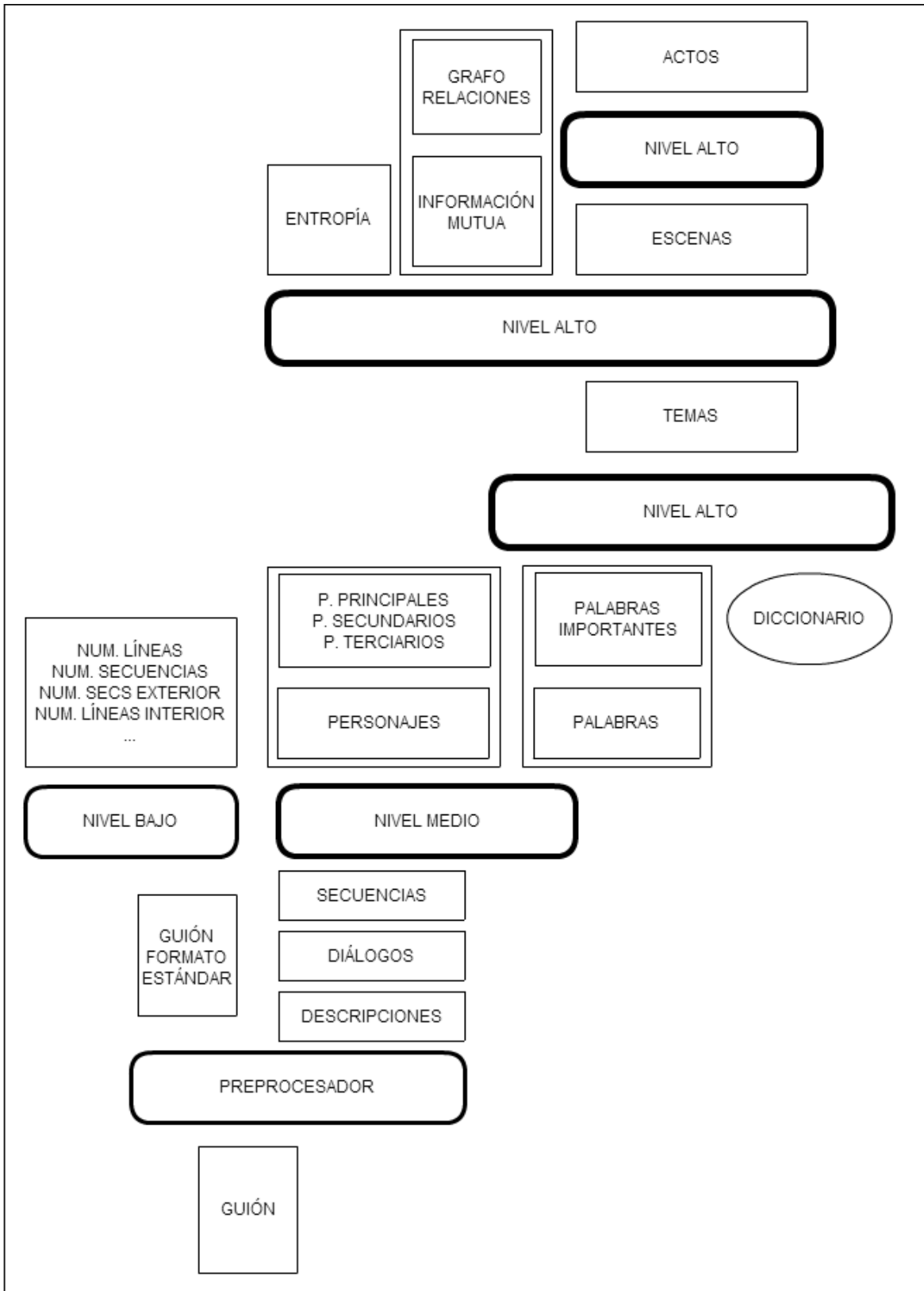


Figura 3.1: Esquema de la estructura del programa

3.1 Preprocesamiento

Antes de someter el guión a la batería de índices que se van a introducir, es importante tanto comprobar que se cumple con el formato requerido como realizar las correcciones pertinentes que garanticen la normalización de su longitud.

Comprobar la adecuación del guión con respecto al estándar es condición indispensable para extraer correctamente su estructura básica. En caso de verificarse el formato, se asigna una etiqueta a cada línea, que la caracteriza como título de secuencia, nombre de personaje, descripción... A partir de esta información se podrán seguir elaborando el resto de índices.

Por su parte, la normalización de la longitud del guión, que conlleva entre otros la eliminación de líneas vacías repetidas, es fundamental a la hora de poder comparar resultados entre guiones. Además, la extensión de un guión tiene también una lectura temporal expresada en la convención de que una página se corresponde a un minuto de película, ver apartado C.3.

La explicación de la pipeline de preprocesamiento se desarrolla en el Anexo A.

3.2 Índices sobre el género

La idea es la de encontrar unos índices que permitan categorizar los guiones en primera instancia, sin una gran elaboración previa, de la misma manera que cualquiera reaccionamos catalogando el tipo de historia a la que nos enfrentamos nada más ver los elementos que ésta pone en juego. En este caso, dichos elementos tienen que ver básicamente con la estructura sintáctica y el formato. Son medidas directas y cuantitativas que realizamos sobre el guión. A saber:

- Número de líneas.
- Número de secuencias.
- Número de líneas por secuencia.
- Número de secuencias interiores.
- Número de secuencias exteriores.
- Número de líneas de descripción.
- Número de líneas de diálogo.
- Ratio de líneas de descripción-diálogo.

Los denominamos “índices de género” por la información que puede obtenerse a partir del conocimiento del número de líneas, de las citas entre personajes, etc. A título de ejemplo, la relación entre texto referente a escenas exteriores o interiores nos da una pauta para identificar, por ejemplo, historias de acción (con persecuciones) o dramas más intimistas (un solo escenario y más centrado en relaciones o diálogos). La extensión también puede indicar la pauta de complejidad del filme (no será fácil que tramas muy enrevesadas puedan desplegarse en textos pequeños).

En realidad, tan sólo el número de líneas y de secuencias se utilizan más adelante. El resto quedan como una suerte de estadísticas que caracterizan el guión de forma general, y que podrían ser seguramente más explotadas en un futuro en la línea de un clasificador de guiones como se propone en el apartado 5.3. Los resultados de estos índices pueden verse en el Anexo D.

3.3 Índices sobre personajes

En este caso la idea es la de caracterizar conceptos asociados típicamente a los personajes como, por ejemplo, el protagonismo, lo mucho que hablan y actúan o los momentos en los que son presentados. Obtendremos así pues varios índices a partir de los personajes y su información asociada en el guión. Antes de entrar a explicar con más detalle el proceso, hay un caso especial a

tener en cuenta que añade un paso previo al del resto: los personajes que no hablan (Artoo⁴, Chewbacca, Toothless⁵...) se deben incluir manualmente, a razón de uno por línea, en un fichero de personajes. El programa carga primero dicho fichero y añade a continuación los que detecte automáticamente en la fuente textual.

La extracción de personajes se apoya en dos funciones. La primera itera a lo largo de las líneas del guión hasta dar con aquellas que hayan sido etiquetadas como nombre de personaje. Entonces, la segunda función se encargará de extraer el nombre limpiando la línea de espacios y tabuladores y eventuales anotaciones de voz en off (V.O.) u off screen (O.S.). También aporta información acerca de si el diálogo es continuación de un bloque de diálogo anterior (CONT) o no.

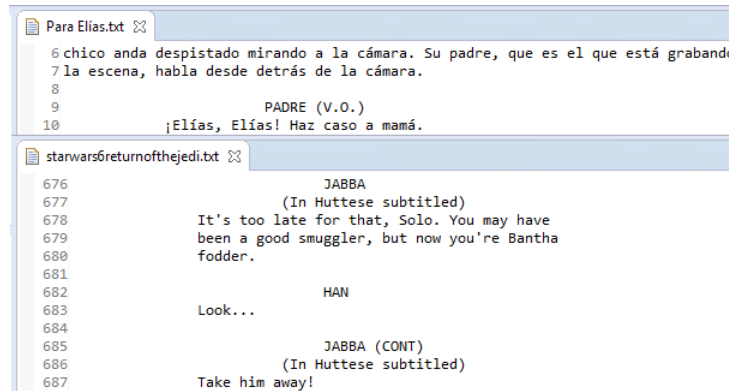


Figura 3.2: Añadidos que pueden acompañar al nombre de un personaje

La información buscada para cada personaje y cada secuencia es:

- Número de veces que habla, o número de bloques de diálogo.
- Número de líneas de diálogo.
- Número de referencias, o número de veces que aparece nombrado en las descripciones.

Y los índices que se calculan son:

- Número de personajes por secuencia.
- Número de líneas de diálogo por personaje en cada secuencia.
- Número de veces que es referenciado un personaje por secuencia.
- Ratio entre líneas de diálogo y referencias por personajes en cada secuencia.

Por último, también se realiza una clasificación de los personajes para determinar si son principales, secundarios o terciarios. Las características tomadas como relevantes para esta clasificación son el número de líneas de diálogo por personaje en cada secuencia y el número de referencias de cada personaje por secuencia, tomadas como valores en el espacio:

$$Protagonismo(p) = \{líneas_diálogo_p, referencias_p\}$$

Mediante un simple algoritmo de clusterización⁶ estos puntos son agrupados en tres conjuntos. Los grupos resultantes se ordenan de mayor a menor importancia en función de las coordenadas de su centro.

Valga como ejemplo la primera Figura de cualquiera de los Anexos E - Q, donde pueden verse los resultados obtenidos por este índice para cada guión.

⁴ Forma de referirse al personaje R2D2 en los guiones de “Star Wars: Episodio V - El imperio contraataca” y “Star Wars: Episodio VI - El retorno del Jedi”.

⁵ Nombre en inglés del personaje Desdentao de “Cómo entrenar a tu dragón 2”.

⁶ Algoritmo *k-means*, ya implementado en el módulo `sklearn.cluster`.

Otras medidas llevadas a cabo en este ámbito son las de la contabilización del número de personajes –principales, secundarios y terciarios- presentados hasta una secuencia dada y el de la aparición o no aparición de los personajes a lo largo de todas las secuencias. De nuevo, los ejemplos de estas métricas pueden verse correspondientemente en las Figuras 2 y 3 de los Anexos mencionados en anterior párrafo.

Para una presentación y discusión de los resultados de estos índices ver apartado 4.3

3.4 Índices sobre los temas

Los índices que se exponen a continuación pretenden no sólo extraer los temas de los que trata la historia sino también captar la dinámica con la que se despliegan a lo largo del tiempo. La extracción de temas se divide en dos etapas. (i) Primero, en detectar las palabras relevantes de las secuencias (Chalmers et al; 2010) y (ii) segundo, en encontrar unas ideas de carácter más abstracto en cuyo significado se engloben dichas palabras. Para la caracterización temporal se recurre a conceptos de Teoría de la Información como, fundamentalmente, diferentes índices relacionados con la entropía.

3.4.1. Palabras relevantes

Para enfrentar este problema se plantean dos funciones principales. La primera se dedica, para cada escena, a contar el número total de palabras y el número de repeticiones de cada palabra. La segunda trabaja en base a estos datos para calcular la métrica TFIDF.

En general, la métrica TFIDF trata de ordenar las palabras w de un texto d perteneciente a un corpus D en función de lo mucho que aparecen éstas en dicho texto, $TF(w, d)$, y lo poco que aparecen en el resto del corpus, $IDF_D(w, d)$. El producto de estos dos valores da como resultado un valor que permite cuantificar la relevancia de las palabras de un documento en el contexto de un corpus concreto. Expresado en términos formales:

$$TF(w, d) = \frac{|\{w' \in d : w' = w\}|}{|\{w \in d\}|}$$

$$IDF_D(w, d) = \log \frac{|D|}{|\{d \in D : w \in d\}|}$$

$$TFIDF_D(w, d) = TF(w, d) * IDF_D(w, d)$$

A la hora de implementarlo en el contexto que nos ocupa, cada secuencia del guión se considerará como un texto independiente d dentro del corpus total D –que queda conformado entonces por todas las secuencias del guión–.

El resultado al que llegamos es el de una lista de las palabras de cada secuencia ordenada “por relevancia”.

3.4.2. Diccionario

La entrada para este bloque son las propias palabras del texto tras su paso por la métrica TFIDF y por la conexión con un diccionario analógico conceptual. El resultado obtenido son los temas del guión, aunque con un cierto matiz: no tanto los temas en sí de los que se habla, sino las líneas temáticas por las que discurre el texto. No se puede decir realmente que se sepa cuáles son los temas del guión, pero sí que hay unas ciertas unidades temáticas. El problema es que las etiquetas o nombres asignados -como temas- a cada una de estas unidades temáticas, no son muy informativos. A veces hace falta saber qué palabras son las responsables de que salga uno u otro tema, e incluso la

secuencia de la que proceden, para saber exactamente a qué se refiere una etiqueta. Por ejemplo, podemos tener un resultado que proporcionase dos secuencias en las que se hablase del tema “Artes plásticas”. En una de ellas, algunas de las palabras que se engloban bajo este tema son “fotografías”, “imagen”, “fundido” o “iluminado”. Mientras que en la otra secuencia son palabras como “teclado”, “melodía” o “piano”. Ambas tratan sobre artes plásticas, aunque parece que una se decanta por el sentido de artes visuales y la otra por el de artes musicales. No es por tanto un problema debido a una mala identificación de los temas. Si bien la denominación que engloba las palabras puede ser ambigua o inexacta, la agrupación de éstas bajo la misma unidad temática es correcta.

El hecho de que los temas no sean muy informativos es debido a que son denominaciones muy abstractas. Lo cual es esperable que ocurra, ya que el algoritmo empleado en la tarea de englobar palabras bajo temas se queda siempre con aquellos temas que son compartidos por más palabras, es decir, los más generales. Cuanto más concreto seleccionemos un tema –y por tanto, como veremos, más descendamos en el árbol de ideas relacionadas con las palabras- menos seguros estaremos de la validez general del tema, por ser compartido por menos palabras. Y al revés, cuantas más palabras se quieran relacionar bajo un tema, más abstracto será éste. Este es un problema inherente al lenguaje natural de muy difícil solución. Por eso optamos en este caso por una aproximación conservadora, que extraiga temas muy abstractos pero que sean compartidos por cuantas más palabras mejor. Queda para el futuro plantear otras soluciones más sofisticadas que permitan afinar la calidad informativa de las respuestas. No obstante, los resultados alcanzados con el planteamiento actual son, como veremos, satisfactorios.

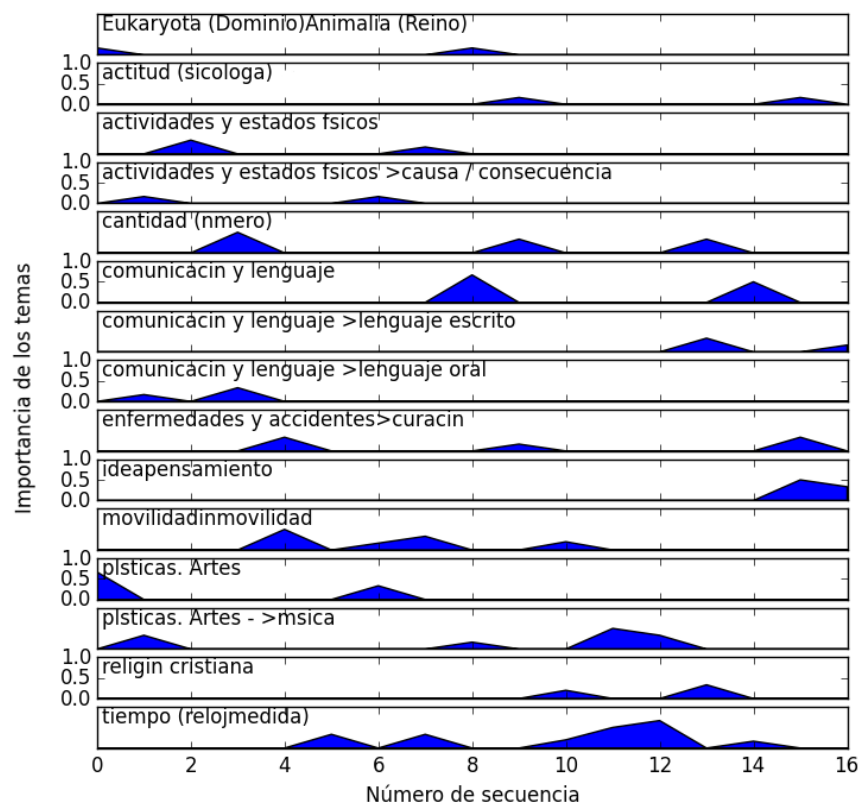


Figura 3.3: Distribución de temas en “Para Elías” (ver apartado Q.7)

3.4.2.1 Ideas

Señalábamos al principio del apartado que una segunda etapa del proceso de extracción de temas era encontrar unas ideas de carácter abstracto en las que se englobaran las palabras seleccionadas previamente como relevantes. Este problema a su vez podemos dividirlo en dos pasos: por un lado, encontrar la manera de obtener esas ideas, por otro, la manera de relacionar palabras e ideas.

El elemento clave para solucionar todo esto va a ser el uso de un diccionario especial.

Soluciones disponibles

Este diccionario especial es un diccionario analógico conceptual. Analógico quiere decir que es un diccionario en el que las palabras se agrupan por campos semánticos (es decir, porque son análogas) frente a un diccionario alfabético, como por ejemplo el de la RAE, en el que encontramos las palabras agrupadas alfabéticamente. La diferencia entre un diccionario analógico y uno de sinónimos o antónimos es que en la relación de palabras sugeridas del analógico se incluyen también referencias a otros campos semánticos que puedan estar relacionados. Si el caso de uso de un diccionario alfabético es el de comprender lo que se está leyendo -comprender la idea a la que se refiere una palabra-, el de uno analógico sería dar con la palabra exacta de la idea que tenemos en mente -pensando más en un contexto de escritura que de lectura-. Conceptual quiere decir que la búsqueda se hace en base a conceptos, no de palabras.

Para esta tarea nos basaremos en la plataforma Zirano (www.zirano.com). La Plataforma Zirano es una Plataforma Multilenguaje de diccionarios analógico-conceptuales que se construyen y se consultan por medio de un interfaz web. En español, permite acceder al diccionario de Julio Casares⁷, en la que la búsqueda se hace a través de una palabra inicial devolviendo como output “el concepto” al que pertenece. Podemos así buscar por “madera elástica” y llegar a “arco” o por “noche cerrada” y llegar a “preocupación”.

Zirano funciona de la siguiente manera: (i) el usuario introduce una idea, (ii) el sistema le sugiere una lista de ideas relacionadas, (iii) el usuario selecciona una de ellas y (iv) el sistema le muestra finalmente una lista de palabras asociadas (organizadas en grupos de sustantivos, artículos, adjetivos, verbos, adverbios...). En este proyecto se (re)define el proceso del siguiente modo: (i) se introduce una *palabra*, (ii) se sugieren posibles *sentidos*, (iii) se selecciona uno de ellos y (iv) se devuelve un conjunto de *ideas* relacionadas. La extracción de los temas en otros dominios (Latasa, 2011) nos ha servido como punto inicial. En nuestro caso hay que garantizar que la respuesta incluya también las ideas que el diccionario relaciona con todos esos sentidos.

No podemos usar otras herramientas disponibles, como por ejemplo, WordNet, que nos habíamos planteado en un principio para esta tarea y cuyas limitaciones son expuestas más adelante. La principal y crucial consecuencia de seguir el desarrollo con Zirano es que sólo podremos obtener los temas de aquellos guiones que estén escritos en español (WordNet nos permitiría trabajar con datos en inglés).

Bien es cierto que en inglés existe el tesoro de Roget, que puede incluirse en esta línea de organizar palabras en función de la similitud semántica de sus ideas asociadas (Jarmasz, 2003). Pero las herramientas encontradas no son comparables con las de Zirano o WordNet en cuanto a disponer una plataforma de fácil acceso y documentada, por lo que, si bien es una posible vía más a explorar en el futuro, queda descartado como posible solución en el ámbito de este proyecto.

WordNet distingue sustantivos, verbos, adjetivos y adverbios; el resto, como las preposiciones, no aparecen. En cuanto a los sustantivos, agrupa las palabras en conjuntos de sinónimos, llamados synsets. Las relaciones –semánticas- que se establecen entre dichos synsets son básicamente de hiponimia⁸ e hiperonimia⁹. Estas relaciones de tipo IS-A muestran que WordNet actúa no sólo como un tesoro -que organiza palabras en función de un criterio específico, por ejemplo, similitud de significado- sino que puede entenderse también como una ontología -ya que establece unos conceptos generales sobre los que se asienta cualquier relación entre palabras-.

⁷ Julio Casares (1877-1964) fue un académico e importante lingüista español autor del “Diccionario ideológico de la lengua española”. También conocido por el apellido del propio autor, este diccionario es en el que se basa la plataforma Zirano que aquí se utiliza.

⁸ Una palabra se considera hipónima de otra cuando el significado de la primera está incluido en el de la segunda; p. ej., gorrión respecto a pájaro.

⁹ Una palabra se considera hiperónima de otra(s) cuando el significado de la primera incluye al de otra u otras; p. ej., pájaro respecto a jilguero y gorrión.

FreeLing es una librería –como por ejemplo NLTK- que ofrece herramientas para analizar el lenguaje y, entre los muchos idiomas que soporta, se incluyen el español y el inglés. Según se explica en la documentación, posee un par de módulos que tratan con el sentido de las palabras en ambos idiomas. Sin embargo, se centran en el problema de desambiguación de sentido, no en relacionar palabras con ideas. Y, pese a ofrecer también recursos como WordNet, esto tampoco soluciona nuestro problema. Valga el siguiente ejemplo para aclarar la razón que nos lleva a decantarnos por el diccionario conceptual y la diferencia entre las soluciones que brindan una y otra herramienta.

A la vista de las palabras más importantes de la primera secuencia del guión de “Para Elías” (ver apartado Q.7), se realiza una búsqueda de las mismas en el diccionario Zirano. Es importante recalcar que, al hacerse manualmente, el sentido que se selecciona para cada palabra es aquel con el que se usa en el guión. Las palabras son: “cámara”, “piano”, “notas”, “tono” y “melodía”.

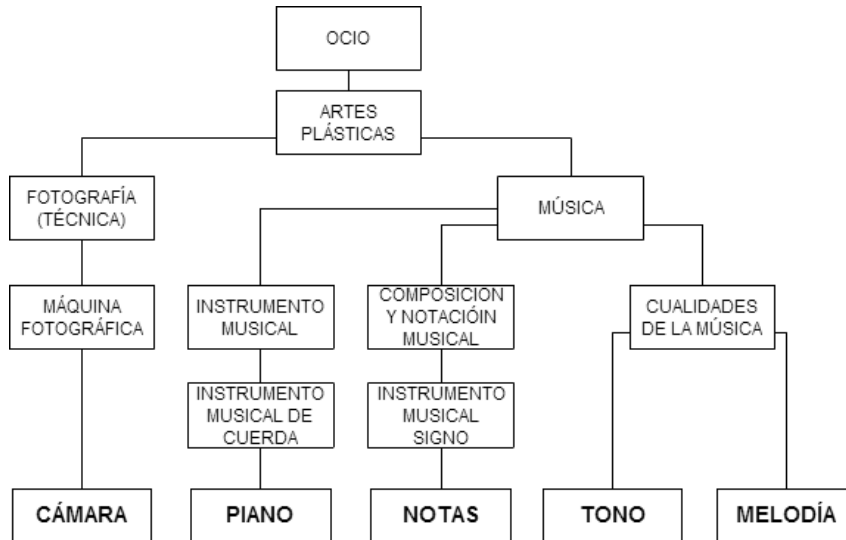


Figura 3.4: Relaciones entre palabras según Zirano

Las hojas son las palabras que hemos buscado, mientras que el resto de nodos son las ideas de las que éstas derivan –más abstractas conforme más subimos en el árbol-. Si nos fijamos en aquellos nodos comunes a varias palabras podremos hacernos una imagen de las ideas que resumen la secuencia. Las ideas que mejor resuman las palabras de una secuencia serán consideradas los temas de la misma.

En la práctica surgen dos problemas. Que, como se ha dicho ya previamente, la idea no sea muy informativa y no quede claro a qué se refiere y también que, a la hora de realizar este proceso automáticamente, seamos capaces de seleccionar en qué sentido se están usando las palabras. Respecto a lo primero, como ya se ha dicho, adoptamos una solución conservadora que se quede con los comunes más abstractos, primando *recuerdo* sobre *precisión*. Y respecto a lo segundo, la solución la vemos más adelante y también más ampliada en el Anexo B. De momento, veamos qué puede aportar WordNet a la solución del problema teórico que venimos planteando.

Tomamos las mismas diez palabras pero traducidas al inglés: “camera”, “piano”, “notes”, “tone” y “melody”.

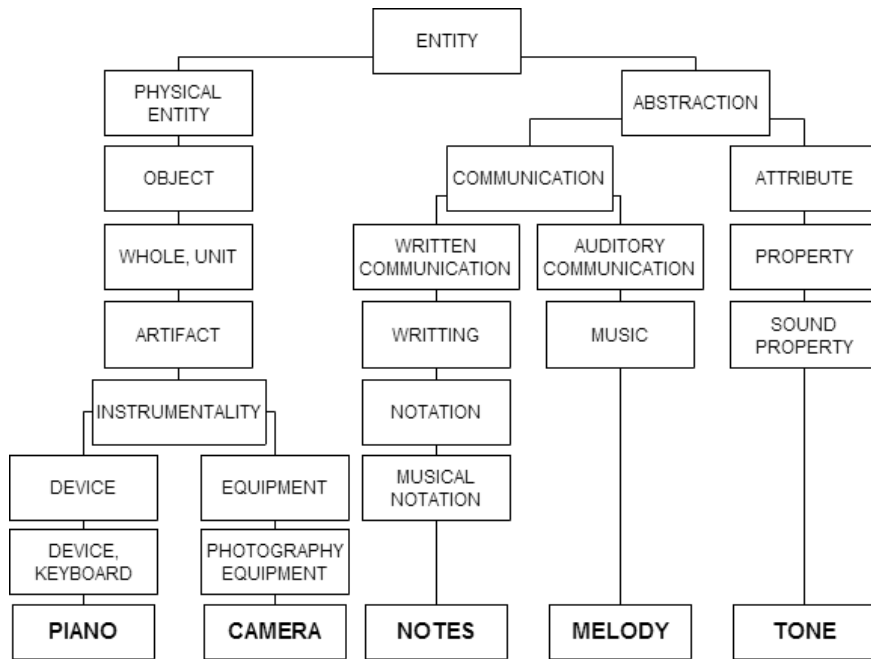


Figura 3.5: Relaciones entre palabras según WordNet

Como puede observarse en la Figura 3.5, tomar el contenido de los nodos comunes como los temas de la secuencia, ofrece una respuesta sustancialmente distinta a la anterior. No es lo mismo decir que una secuencia trata sobre “ocio” y “artes plásticas” que sobre “entidades físicas” y “abstracciones”. Por mucho que el primer grupo pueda ser matizado, la información aportada va mucho más en la línea de lo buscado. El problema con WordNet es que las relaciones IS-A que establece nos informan acerca de cosas que las palabras de inicio son pero de un modo más abstracto. Los temas son una generalización relacionada con las palabras, pero en un sentido amplio. Relacionar “piano” y “música” resulta natural para cualquier persona, pero un piano no es en ningún sentido música. Se puede decir que es una entidad física, un objeto, un todo o unidad, un artefacto, un instrumento, un dispositivo, un teclado... pero no *es* música. Por eso, porque nunca va a poder relacionarnos “piano” con “música”, desechamos la opción de WordNet.

Implementación

Una vez establecida la conexión con la página web, que permite enviar la petición de búsqueda de la palabra, recibe un html y lo procesa en busca de la lista de sentidos -mediante expresiones regulares que aprovechan las etiquetas del documento- todavía hay que añadir una etapa de búsqueda más. A partir de la lista de sentidos, habrá que pedir la página que contiene el árbol de ideas asociado a éstos. El resultado se devuelve en una larga cadena de texto que ya será tarea de Python traducir a la correspondiente estructura de datos.

Tras la puesta a punto de la conexión con el diccionario surge un problema con los tiempos de respuesta. Ciertas palabras tardan mucho en obtener respuesta del diccionario. Frente a las unidades o pocas decenas de segundos que suele tardar en darse una respuesta, algunas pueden llegar a tardar cientos de segundos. Un análisis en profundidad de estos casos revela que cuantos más sentidos tiene una palabra, más se demora la respuesta. Por eso, y teniendo en cuenta además que seleccionar un sentido entre los cientos que pueden llegar a tener algunas palabras no va a aportar mucho, se decide crear una lista de palabras semánticamente no relevantes en la que se irán añadiendo aquellas que tengan más de un número de sentidos fijado. Antes de lanzar la búsqueda de una palabra al diccionario se comprobará que no pertenece ya a dicha lista. Típicamente, las palabras que se seleccionan son preposiciones, pronombres, interjecciones, conectores, diminutivos... A la vista de los resultados que se van obteniendo, el máximo que se fija para no rechazar una palabra es de veinte sentidos.

Una vez ya extraída y organizada toda la información hay que decidir en qué sentido se utiliza cada palabra y cuál es el tema de cada secuencia.

3.4.2.2 Sentidos y palabras

A modo de síntesis, hasta ahora:

- Se han extraído las palabras importantes para cada secuencia.
- Se han obtenido diferentes sentidos posibles con los que se podrían utilizar dichas palabras.
- Se ha obtenido también una lista de ideas asociadas a cada sentido de cada palabra, organizada de mayor a menor grado de abstracción.

La estrategia para lograr llegar a extraer los temas del guión va a ser encontrar cuáles son aquellas ideas de mayor nivel de abstracción que se relacionan con más palabras. De manera que, expresándolo en los términos en que venimos haciéndolo, los temas toman su nombre de las ideas más abstractas que el diccionario conceptual es capaz de devolver. Pero en esta cadena que se ha establecido -guión, secuencias, palabras, sentidos, ideas, temas-, todavía queda un paso por aclarar: ¿en qué sentido se utilizan las palabras en el guión?

Planteamos que el sentido que será aquel compartido mayoritariamente por todas las palabras de la secuencia. Dicho de otra manera, que maximice el número de palabras que lo comparten.

Implementación

Se plantean dos aproximaciones de implementación para este problema de optimización: una de fuerza bruta y otra basada en un algoritmo genético.

Para el algoritmo de fuerza bruta se valora cuántos sentidos de palabras comparten una misma idea -en concreto, la idea de mayor nivel de abstracción-. Aquella que es compartida por más sentidos se selecciona como el primer tema del que se habla en la secuencia. Después se vuelve a repetir el conteo, previa eliminación de aquellas palabras que tuvieran algún sentido relacionado con la idea que se acaba de seleccionar. Y así hasta que no queden más palabras. También se calcula para cada una de las ideas seleccionadas cuántas palabras han contribuido a seleccionarla, dando así una noción de en qué proporción se habla de unos y otros temas.

Es importante hacer notar que la distribución de temas se hace en este caso a nivel de las secuencias. Otra opción sería agrupar globalmente todas las palabras y sus sentidos e ideas asociadas. Ejecutar el algoritmo de extracción de temas por fuerza bruta nos daría en este caso una distribución de los temas a nivel global de todo el guión. Ver ejemplo del Anexo B para un mayor detalle.

El algoritmo genético selecciona inicialmente ocho individuos: ocho distribuciones aleatorias de sentidos para todas las palabras. De todas las ideas asociadas a cada sentido sólo se tendrá en cuenta la más global. A esta solución parcial se le otorga un valor que aumenta cuantos menos temas se seleccionen globalmente, lo que quiere decir que premia que se agrupen bajo un mismo tema cuantas más palabras mejor. En función de este valor se determinan cuáles son los dos mejores individuos, que pasan directamente a la siguiente generación. Y en función de estos dos se generan el resto de individuos, mediante combinaciones lineales aleatorias, hasta completar la población de ocho. Además, con probabilidad 0.1 se produce una mutación, y el individuo no es generado como combinación lineal de los mejores de la anterior generación sino aleatoriamente. En cualquier caso, al mantener los mejores individuos de la generación pasada, se asegura que la población no empeorará en ningún caso. Esto es lo que se llama un algoritmo genético elitista. En el caso peor la solución se quedará estancada, momento en el que se termina de ejecutar el algoritmo.

Como se ha indicado a raíz del criterio valorativo para las soluciones parciales, y al contrario que en la aproximación anterior, en este caso la función de fitness premia soluciones que tengan el menor número de temas posible globalmente. Incorporando otros parámetros a la función se podría tratar de encontrar soluciones más equilibradas, que tengan en cuenta tanto la distribución por secuencias como a nivel general. (Ver Anexo B).

3.4.3 Entropía

Las siguientes métricas son planteadas con la intención de entender el despliegue de los temas que se acaban de extraer. La idea es capturar los momentos más inesperados de la misma, entre los que podrían encontrarse los puntos de giro de la historia.

La entropía cuantifica la cantidad de información promedio asociada a una variable aleatoria

$$H(X) = - \sum_{i=1}^{N_X} p_x(X_i) * \log p_x(X_i)$$

y sus características principales son:

- $H(X) \geq 0$ La entropía de una variable aleatoria discreta es una función no negativa.
- $H(X) \leq \log(N_X)$ El valor máximo que puede tomar la entropía es el logaritmo del número de elementos, que se produce cuando la distribución es equiprobable para todos los casos: $p_x(X_i) = \frac{1}{N_X}$

Aplicando esta medida a la distribución de los temas a lo largo del guión podemos saber qué temas son más entrópicos. Pero lo que buscamos realmente es saber cuáles son las secuencias más informativas. Para ello, se hace la suma de la entropía de todos los temas a lo largo de una ventana deslizante de x secuencias. El tamaño óptimo x de ventana depende de la longitud total del guión. En este trabajo, para los guiones de largometraje, se ha establecido en diez secuencias.

La gráfica obtenida hay que leerla no como aquella secuencia que resulta más sorprendente sino como aquel conjunto de x secuencias que presentan una distribución temática menos uniforme, dando una idea de los temas que se presentan simultáneamente. Por tanto, si se despliegan muchos temas a la vez, se supone que en la historia están ocurriendo cosas. Y al revés, un valor bajo de entropía significaría que más o menos se está hablando de lo mismo todo el rato (no que se hable de cosas “poco interesantes”).

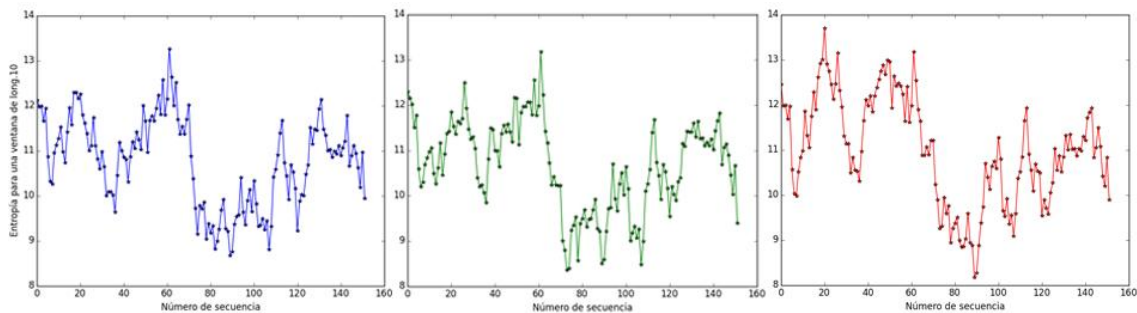


Figura 3.6: Evolución de la entropía para tres ejecuciones diferentes.

Hay que hacer notar que, como puede observarse en la Figura 3.6, la entropía cambia notablemente de una ejecución a otra -debido a que la distribución de los temas cambia-. Ahora bien, un análisis detallado muestra que hay puntos que siguen siendo invariablemente más o menos entrópicos. Se sigue un patrón. Esto se hace más evidente cuanto más observemos la gráfica desde un punto de vista general. Así pues, en la primera mitad se observan con más facilidad un par de ciclos de subida y bajada de entropía y en la segunda mitad tres ciclos, si bien aquí hay mucho más movimiento.

Estos ciclos de alta y baja entropía capturan momentos de mayor o menor dispersión en los temas y personajes, asociando momentos de mayor desorden y de cruce de tramas diferentes a alta entropía, y momentos de enfoque en unas pocas tramas (por ejemplo en momentos de mayor tensión narrativa) en momentos de baja entropía. Este tipo de ciclos podrían estar capturando estructuras de

presentación-nudo-desenlace a diferentes escalas. Esta es una hipótesis muy interesante que habría que seguir investigando.

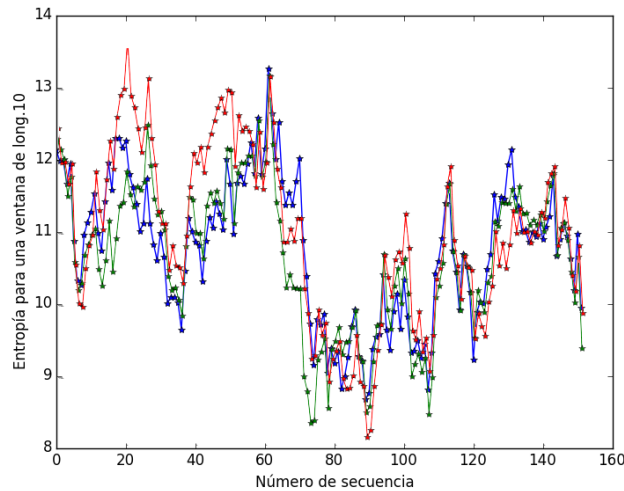


Figura 3.7: Superposición de las distribuciones de entropía.

Para comprobar que los datos siguen efectivamente la estructura observada y no responden a un mero ruido estadístico se lleva a cabo el siguiente experimento. Los valores de importancia posibles que pueden tener los temas pasan a ser 0 ó 1. Si el tema no aparece en la secuencia tendrá valor 0 y, si aparece, sea en la media que sea, tendrá valor 1. Los resultados obtenidos para estos datos presentan la misma estructura que en el caso anterior, por lo que los damos como fiables.

3.4.4 Entropía cruzada

Se lleva también a cabo otra medida sobre la entropía, pero no en términos absolutos sino relativos a la entropía general. Esto quiere decir que, por ejemplo, una ventana de secuencias en la que hubiera poca entropía -su distribución es “poco sorprendente”- puede diferir en cambio mucho de la distribución total, resultando muy informativa con respecto a la distribución general.

$$H(p, q) = - \sum_X p(X) * \log q(X)$$

El resultado obtenido demuestra que las ventanas con poca entropía son también poco entrópicas en relación con el conjunto y al revés, las muy entrópicas son también más entrópicas que el resto respecto del conjunto. Lo que pone de manifiesto que la distribución de la información es homogénea a lo largo del guión.

Para comparar más fácilmente las distribuciones, la Figura 3.8 relaciona en la misma gráfica las Figuras L.3.1.1 y L.3.1.2:

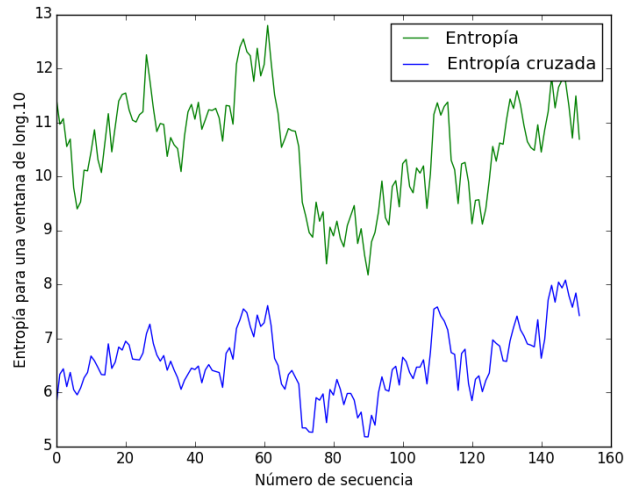


Figura 3.8: Superposición de las distribuciones de entropía en “El laberinto del fauno”

3.5 Índices sobre las tramas

Si hablamos de la trama como la unión o el tejido que se establece entre los elementos de la historia, el intento de representación de la misma mediante una red o un grafo no parece desacertado. Los elementos básicos de todos los índices y medidas que venimos presentando hasta el momento son los temas y los personajes, que son justamente lo que vamos a relacionar en estos índices.

La información mutua es un concepto complementario al de la noción de entropía. Se calcula sobre dos variables aleatorias y representa la información que aporta una sobre el conocimiento de la otra.

$$I(X, Y) = - \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} p_{X,Y}(x_i, y_j) * \log \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i) * p_Y(y_j)}$$

Y sus características principales son:

- $I(X, Y) = I(Y, X)$ La información mutua entre dos variables aleatorias discretas es una función conmutativa.
- $I(X, Y) \geq 0$ La información mutua es siempre no negativa.
- $I(X, Y) \leq \min(H(X), H(Y))$ El valor máximo que puede tomar la información mutua está acotado por el mínimo entre la entropía cada una de las variables aleatorias.

Nos planteamos buscar las relaciones entre temas y personajes: ¿Qué temas se relacionan más entre sí? ¿Con qué temas se relacionan más los personajes? Las respuestas a estas preguntas se pueden medir con el cálculo de la información mutua entre los temas y entre los temas y los personajes. Se seleccionarán aquellas parejas que presenten una mayor relación y se procederá a la representación gráfica de los resultados en forma de un grafo.

En el grafo, los nodos son temas o personajes, y las aristas indican una relación entre nodos. Para las parejas de temas se seleccionan aquellas que se encuentran por encima del percentil 80, y para las de temas y personajes del percentil 60.

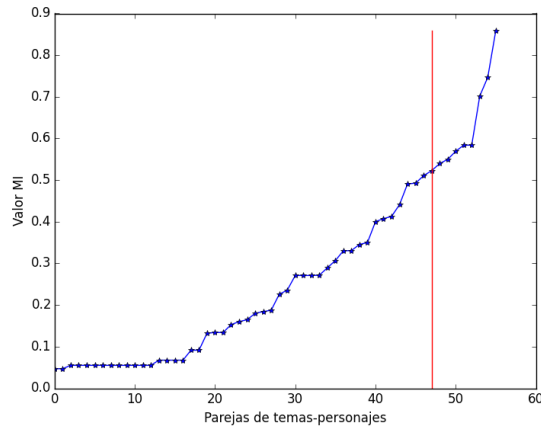


Figura 3.9: Valores de MI para parejas de temas y personajes en “Para Elías” y posición del percentil 60

Normalmente obtenemos un grafo no conexo en el que resulta muy interesante observar qué personajes -y a través de qué temas- se encuentran agrupados en diferentes subgrafos. Los nodos de un mismo subgrafo están pintados del mismo color para evitar ambigüedades. Para clarificar el estudio se introduce una leyenda que mapea letras con temas y números con personajes. Valgan las Figuras L.4.1 y ;.4.1 a modo de ejemplo.

De nuevo, diferentes ejecuciones del programa arrojan una salida diferente. Resulta muy interesante no obstante ver cómo sigue habiendo relaciones temáticas y entre personajes que siempre se mantienen (pueden compararse las Figuras L.4.2 y M.4.2 con las otras dos que se acaban de mencionar. Ver el Anexo B para más información al respecto de las posibles diferencias de una ejecución a otra).

3.6 Índices narrativos

La idea de la división de una historia en escenas (por ejemplo, la típica escena de persecución, la del enfrentamiento final en un western, etc.) y actos –presentación, nudo y desenlace- es ampliamente conocida y utilizada. Estas métricas vienen a intentar encontrar estas estructuras narrativas de mayor nivel que las secuencias y que no se encuentran señaladas explícitamente en un guión cinematográfico.

3.6.1 Detección de escenas

Se realiza en base a dos medidas: de temas y de personajes. En el caso de guiones en inglés sólo se tienen en cuenta los personajes.

3.6.1.1 Medidas de temas para escenas

Se quiere tener una idea de en qué medida los temas seleccionados se comportan como bloques. Para ello se buscan aquellos temas que se repiten de una secuencia a la siguiente, premiándoles más cuantas más secuencias seguidas tenga un bloque, cuanto mayor sea el porcentaje medio de importancia del tema en dichas secuencias y cuanto menos fragmentado se encuentre el tema en general.

$$BM(t) = \frac{1}{N} * \frac{1}{n_B} * \left(\sum_{i=1}^{n_B} \bar{b}_i * l_i \right)$$

- n_B : número de bloques del tema t
- \bar{b}_i : valor medio del bloque i -ésimo
- l_i : longitud del bloque i -ésimo

La normalización de la medida se lleva a cabo dividiendo por el número de secuencias N , que es el valor máximo que podría llegar a tener la longitud de un bloque.

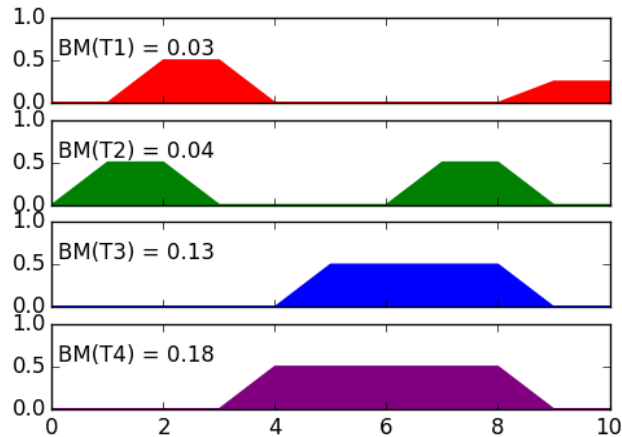


Figura 3.10: Ejemplo de la métrica de bloque

A la hora de agrupar secuencias en escenas, se eliminan primero aquellos temas que han tenido una métrica de bloque nula. Luego, para el resto de temas, ordenados de mayor a menor según acabamos de calcular, se van comprobando los bloques que haya en ese tema. Si ninguna secuencia de ese bloque ha sido ya asignada a una escena, entonces se agrupan como una escena. Quedarán grupos de secuencias que han sido asignadas a una escena y grupos entre medio que no.



Figura 3.11: Ejemplo de asignación de secuencias a escenas según temas.

3.6.1.2 Medidas de personajes para escenas

La idea es saber, para cada pareja de una secuencia y su siguiente, cuál es el porcentaje de personajes de la primera que aparecen también en la segunda. Habiendo hecho eso para todo el guión, se calcula la media de repetición de personajes en la historia y se vuelven a repasar los porcentajes obtenidos entre los pares sucesivos de secuencias: si superan la media del guión, entonces son considerados como la misma escena.

La agrupación en secuencias es más fácil en este caso. Todas las parejas sucesivas de secuencias que se consideren escenas entre sí conforman una escena.

3.6.1.3 División en escenas

En caso de no haber calculado la métrica de los temas -porque el guión estaba en inglés-, se divide en escenas tomando directamente la agrupación según personajes explicada en el punto anterior. Para el caso en el que se han llevado a cabo las dos métricas se procede en cambio de otro modo.

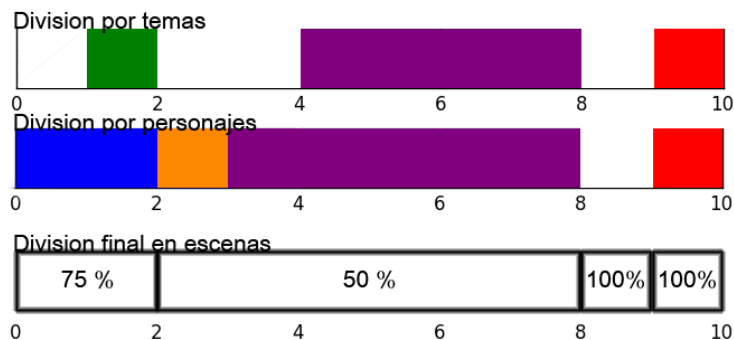


Figura 3.12 Ejemplo de asignación de secuencias a escenas.

Se mira a ver qué secuencias son límite entre escenas a la vez en ambas métricas: esas secuencias – de consenso- serán las que marcarán definitivamente el cambio entre escenas. Adicionalmente, comprobamos cómo de conformes estaban cada una de las soluciones intermedias con este mínimo común final, fijándonos para ello en la cantidad de escenas que estas soluciones habían encontrado en el intervalo seleccionado finalmente.

3.6.2 Detección de actos

Se realiza en base a tres medidas: de temas, de personajes y de escenas. En el caso de guiones en inglés sólo se tienen en cuenta personajes y escenas.

3.6.2.1 Medidas de temas para actos

La idea es saber en qué medida los temas resurgen al cabo de varias secuencias. Se buscan aquellos temas que vuelven a surgir unas secuencias más adelante, premiando más cuantas más secuencias separen una aparición de otra y cuanto mayor sea el porcentaje de importancia del tema en dichas apariciones.

$$EM(t) = \frac{1}{2 * (N - 2)} * \left(\sum_{i=1}^{n_E} l_i * (Imp_{pre} + Imp_{post}) \right)$$

- n_E : número de emergencias del tema t a lo largo del guión.
- Imp_{pre} : porcentaje de importancia del tema en la última secuencia antes de desaparecer.
- Imp_{post} : porcentaje de importancia del tema en la primera secuencia que vuelve a emerger.
- l_i : longitud del i-ésimo espacio entre apariciones

La normalización de la medida se hace dividiendo por el valor máximo que puede resultar de la suma de los porcentajes de importancia de los temas -que es 2- y por el número de secuencias total $N - 2$, que sería la longitud máxima de emergencia para un tema en el caso de aparecer en la primera secuencia y no volver a aparecer hasta la última.

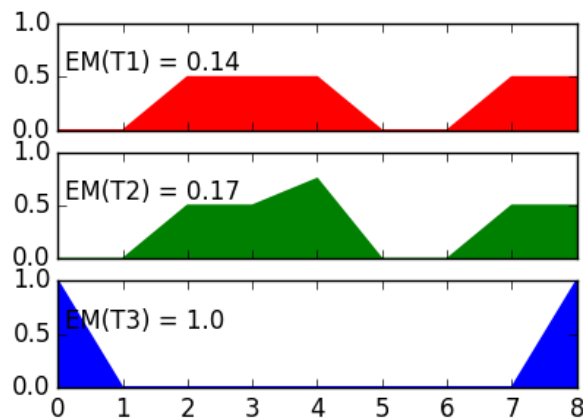


Figura 3.13: Ejemplo de la métrica de emergencia

La idea que inspira esta métrica, al igual que ocurría con la entropía –ver apartado 3.5.1-, sería la de captar los puntos de giro los puntos de la historia. Ambas métricas persiguen el mismo objetivo teórico –encontrar el “cambio de la acción en sentido contrario”, lo que Aristóteles llamaba la peripecia (Aristóteles, 2007) -pero presentan resultados bien distintos, pues capturan distintos comportamientos. Una prima importancia distanciada y la otra incertidumbre en la distribución. Por ejemplo: 0, 1, 0, 0, 0, 0, 0, 0, 0, 1 tendrá mucha emergencia pero poca entropía.

3.6.2.2 Medidas de personajes para actos

Se obtiene la serie de personajes que ya han sido presentados hasta el momento, es decir, cuántos de los personajes principales, secundarios y terciarios han aparecido ya hasta una escena dada.

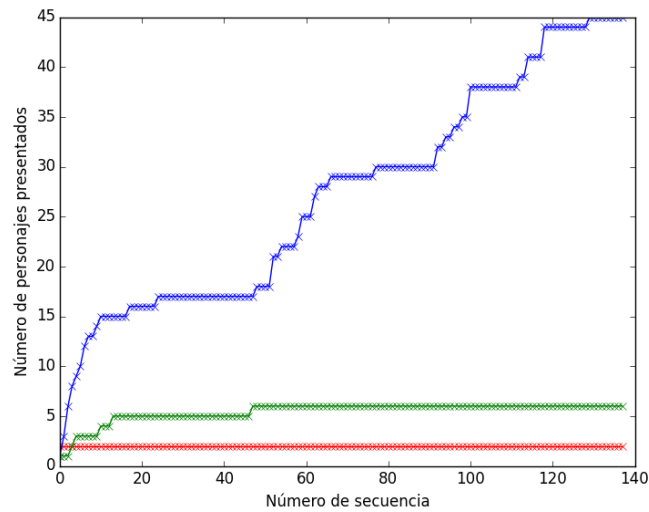


Figura 3.14: Acumulación de personajes presentados hasta una secuencia dada para "Star Wars Episodio VI". Personajes principales en rojo, secundarios en verde y terciarios en azul.

Resultará especialmente interesante saber las secuencias de inicio y fin de aquellos intervalos en los que se estabiliza el plantel de personajes presentado.

3.6.2.3 Medidas de escenas para actos

Agrupación de secuencias en actos tal y como se acaba de explicar en el apartado anterior 3.5.1.

3.6.2.4 División en actos

Una vez está el guión dividido en intervalos (según resurgimiento de temas, de estabilización de número de personajes presentados y de escenas), la atención pasa a centrarse en ver qué secuencias son esas en las que se cambia de un intervalo a otro y cuál es la probabilidad teórica de que ocurra ahí el cambio; por un lado, del primer al segundo acto y, por el otro, del segundo al tercer acto. Para el caso del cambio del I al II acto se consideran las secuencias de inicio de los intervalos, por entenderse que a partir de ahí se produce un periodo de estabilización de la historia. Es un periodo de secuencias durante las cuales hay tema concreto del que no se habla, o no se presenta ningún personaje nuevo o se está hablando más o menos de los mismos temas. Para el cambio del II al III acto se consideran las secuencias del final de los intervalos, por entenderse en este caso que es a partir de ahí cuando empiezan a producirse cambios en algunas de estas líneas.

La probabilidad teórica se calcula sobre una página, por lo que primero debemos hacer una traducción de números de secuencia a números de página –los guiones que procesamos están en texto plano, donde sólo hay información sobre el número de línea o de secuencia-. Para ello, se tiene en cuenta el número de línea en que empieza y acaba una secuencia dada y el número de líneas que caben por página según las convenciones explicadas en el Anexo C. Se asigna como página de la secuencia la media entre la página de inicio y de final, y se calcula su probabilidad teórica. En una película de 100 minutos –que correspondería aproximadamente a un guión de 100 páginas- se supone que el cambio entre el I y II acto tendría lugar alrededor del minuto veinte, y entre el II y III acto alrededor del minuto ochenta y cinco (ver Anexo C para más información al respecto).

Puede verse en la gráfica la probabilidad de que se produzca el cambio entre presentación y nudo en un rango de diez minutos alrededor del minuto veinte.

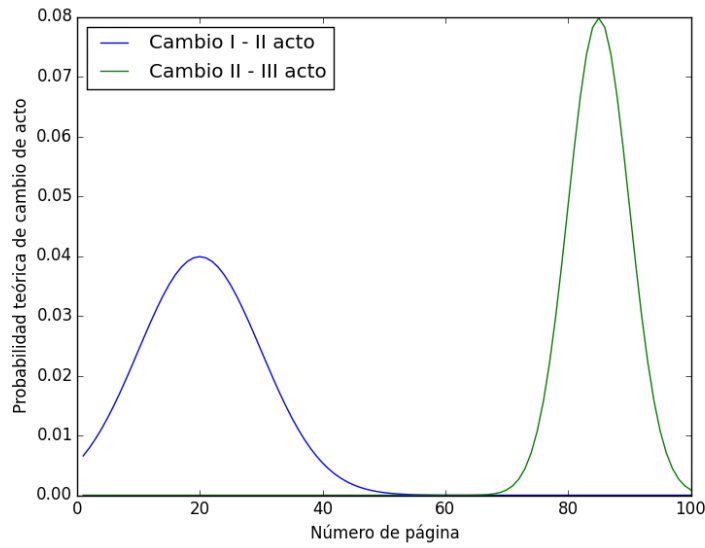


Figura 3.15: Probabilidad teórica del cambio entre actos.

Para el caso del paso del nudo al desenlace en el rango de cinco minutos alrededor del minuto ochenta y cinco. Para aplicar estas probabilidades a otros guiones primero será necesario realizar un cambio de escala. Después se llevará a cabo la mencionada corrección teórica aplicándola sobre las secuencias indicadas por las métricas como puntos de cambio entre actos.

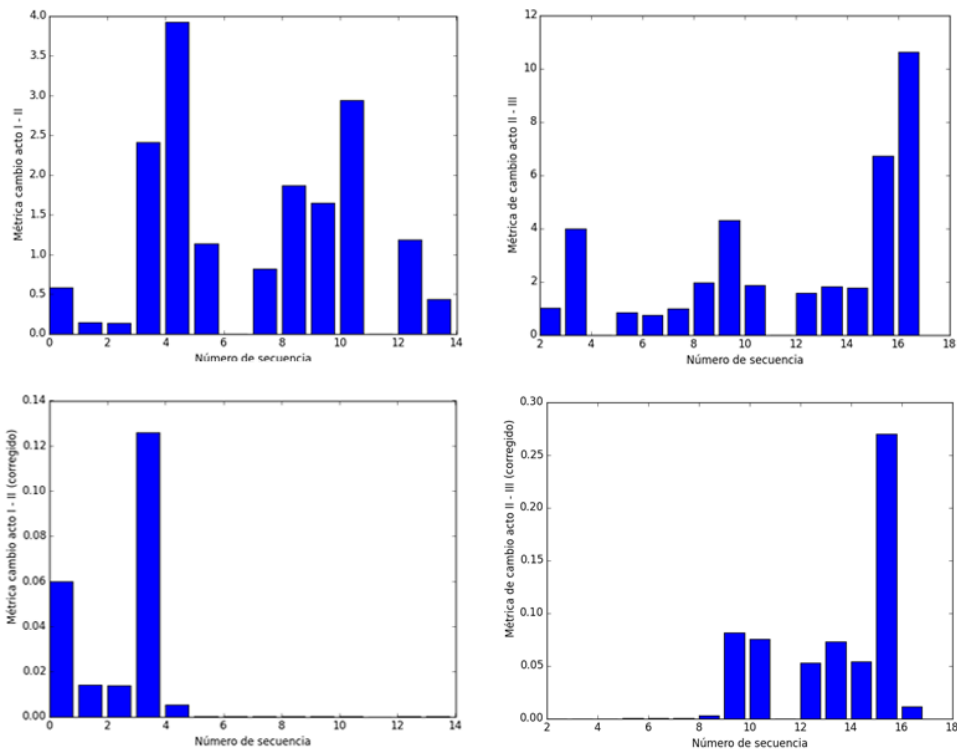


Figura 3.16: Métricas de cambio antes y después de la corrección teórica.

Sumando las probabilidades –corregidas- que cada secuencia ha obtenido se llega a un valor final para cada una de ellas. El mayor valor de todos indicará en qué secuencia se establece el cambio de acto. Como puede verse a continuación, la corrección teórica no tira por tierra todos los cálculos llevados a cabo previamente sino que centra el interés en unos puntos determinados. Comparar gráficas inferiores (corregidas) con superiores (no corregidas) en la Figura 3.16

4

Pruebas experimentales

Este capítulo es una recapitulación de las conclusiones generales obtenidas tras la aplicación de las medidas a los guiones de prueba, que están compuestos por los títulos listados a continuación.

En inglés se ha contado con los siguientes títulos:

- *Blade Runner* (1982, Hampton Fancher y David Webb Peoples, adaptación de la novela “*Do androids dream of electric sheep?*” de Philip K. Dick).
- *How to train your dragon 2* (2014, Dean DeBois y Cressida Cowell).
- *Reservoir Dogs* (1992, Quentin Tarantino).
- *The Matrix* (1999, Andy Wachowski y Lana Wachowski).
- *Se7en* (1995, Andrew Kevin Walker).
- *Star Wars: Episodie V – The Empire Strikes Back* (1980, Leigh Brackett y Lawrence Kasdan, sobre una historia de George Lucas).
- *Star Wars: Episodie VI –Return of the Jedi* (1983, Lawrence Kasdan y George Lucas, sobre una historia de George Lucas).

En castellano se ha dispuesto de dos cortometrajes que han servido sobre todo para testear el código y también como pequeñas muestras de los resultados de los índices. De hecho, varias de las gráficas presentadas hasta el momento corresponden al guión “Para Elías”, debido a su mayor facilidad de lectura por su menor longitud y número de elementos. En cuanto a los largometrajes, se ha contado con los siguientes títulos:

- *El laberinto del fauno* (2006, Guillermo del Toro).
- *Tesis* (1996, Alejandro Amenábar, sobre una historia de Alejandro Amenábar y Mateo Gil).
- *La isla mínima* (2014, Rafael Cobos y Alberto Rodríguez).

Finalmente, presentamos un contraejemplo –también en castellano– para demostrar que el sistema detecta efectivamente una estructura narrativa en un texto narrativo pero no en un texto jurídico, como es el caso aquí ofrecido.

- *Congreso de los diputados. Serie histórica. Legislatura 1810-1813. Cortes de Cádiz. Diarios de sesiones de las Cortes Generales y Extraordinarias, Sesiones de los días 25, 26, 27 y 28 de Marzo de 1811.*
 - *Versión 4 secuencias.*
 - *Versión 100 secuencias.*
 - *Versión 100 secuencias según estructura de “La isla mínima”.*

El contraejemplo consta de las actas de cuatro sesiones del Congreso de los Diputados correspondientes a los días 25, 26, 27 y 28 de Marzo de 1811. Dichas actas han sido formateadas según el estándar que venimos utilizando para los guiones cinematográficos. Para la división en secuencias se han seguido tres estrategias diferentes, que han acabado por conformar los tres guiones finales utilizados como contraejemplo. La primera versión respeta la división de las sesiones en días y tiene por tanto 4 secuencias. Para paliar todos los problemas que surgen de un guión con tan poco número de secuencias, se genera otro documento con 100 secuencias, un número más acorde a la extensión de un largometraje. Éstas se distribuyen uniformemente en un documento cada 50 líneas y en otro en las mismas líneas que uno de los guiones reales analizados anteriormente, el de “La Isla Mínima”. El objetivo es comprobar si al imponer una estructura narrativa a un contenido no narrativo se observa alguna diferencia con respecto a una estructura uniforme.

Como ya se ha indicado previamente, a los títulos en inglés tan sólo se les han aplicado los índices sobre género y personajes. El despliegue de todas las métricas sólo se realiza sobre la selección en castellano, tanto para contraejemplos como en el caso de guiones largos y cortos.

4.1 Índices sobre el género

- Estructuras “singulares” frente a estructuras “canónicas”.

En los índices sobre el género lo que pretendíamos era caracterizar mediante una serie de medidas directas aquellos valores propios de una historia de acción, un drama o un thriller –ver apartado 3.2–.

Lo que se ha obtenido como resultados refleja lo que podríamos llamar un “guión tipo” frente a otro más singular o “de autor”. Por ejemplo en cuanto a extensión –en líneas y en secuencias– y prevalencia de secuencias interiores sobre exteriores y de diálogos sobre descripciones. En líneas generales, dos guiones destacan sobre el resto: “Reservoir Dogs” y “Tesis” –ver columnas correspondientes en la Tabla D.1–. No parece aventurado decir que sus estructuras difieren de la estructura general. Con respecto a esta estructura normal, canónica o incluso podríamos decir “industrial”, encontramos algunas que son más singulares. Si hay dos autores conocidos de entre todos los guiones seleccionados, son sin duda Q. Tarantino y A. Amenábar. Lo que estamos viendo

son valores de lo que por un lado sería una estructura básica y por otro lado, una singular. Aunque obviamente, cada guión es particular y destaca en algún aspecto. El caso del contraejemplo apoya además esta teoría, por lo menos en cuanto a la relación entre diálogos y referencias.

De estas particularidades mencionadas quizás el caso más llamativo sea el de “Cómo entrenar a tu dragón 2”, que presenta importantes diferencias en cuanto al ratio de líneas por secuencia o en el hecho de ser la única película con más secuencia exteriores que interiores. Quizás parte de esta peculiaridad tenga que ver con ser del género de animación, lo cual sería una hipótesis muy interesante en la que seguir explorando. Para ver los resultados ordenados por películas e índices ir a la Tabla D.1.

A la vista de los resultados logrados quizás sea un poco aventurada la denominación inicial de “índices sobre el género”. Pero, en cualquier caso, los índices extraídos sí que parecen permitir una clasificación de los guiones en función sus características, aunque ésta sea entre “guiones tipo” y “guiones de autor.” Además, como se explica en el apartado 5.3., seguir por esta vía aplicándola a una cantidad mayor de guiones podría conducir a una clasificación de los mismos más cercana a los términos planteados inicialmente.

4.2 Índices sobre los personajes

- Caracterización satisfactoria del protagonismo de los personajes.
- Detección de la estructura narrativa en términos de los personajes.

Con estos índices el objetivo era capturar ideas asociadas típicamente a los personajes, por ejemplo, el protagonismo –ver apartado 3.3-.

Los resultados a los que hemos llegado son muy interesantes y un buen indicador de si aquello de lo que son reflejo es efectivamente una estructura narrativa o no. En general, al ver las gráficas de protagonismo podemos comprobar que la selección de personajes por grupos es bastante coherente con lo que diría una persona que conociera la historia. Un buen ejemplo puede verse para el caso de “Star Wars: Episodio V – El Imperio Contraataca” en la Figura J.1.1. En la mayoría de películas, “cantidad de diálogos” y “número de referencias” pueden ser leídas a la vez como indicadores de “protagonismo”, generando unos resultados en los que los personajes parecen dispersarse cada vez más a partir del origen de coordenadas siguiendo la diagonal. Un caso en el que no ocurre así es en el de “Reservoir Dogs” – ver Figura G.1.1, donde su bajo número de secuencias, la cantidad de personajes relevantes y lo mucho que todos hablan, define una estructura en las relaciones entre personajes lejos de la canónica-. En este caso, la asignación de personajes a grupos de protagonismo parece ser menos evidente, más estratificada. Es decir, si leemos la gráfica horizontal o verticalmente, obtendremos una agrupación u otra. En función de si nos fijamos en el número de diálogos o de referencias podemos encontrar distintas soluciones. No parece que ambos factores estén recogiendo a la vez un tercer factor “protagonismo”. Esto es lo mismo que ocurre, quizás incluso más acusadamente, en el caso del contraejemplo –ver Figuras Ñ.1.1, O.1.1 y P.1.1-.

A señalar también que, a pesar de todo, la selección de los personajes mudos no es del todo satisfactoria, siendo por ejemplo R2D2 o Chewbacca seleccionados como personajes terciarios en vez de secundarios en los dos guiones de “Star Wars” –ver Figuras J.1.1 y K.1.1- o Desdentao como secundario en vez de principal en “Cómo entrenar a tu dragón 2” – ver Figura F.1.1-.

Y también la lectura que podemos hacer de los distintos tipos de protagonistas posibles para una historia en función de si están más hacia el lado de los diálogos o de las referencias, pudiendo ver

así casos de personajes con un protagonismo más verbal o de acción –ver el caso de Neo y Morfeo en “The Matrix”, Figura H.1.1.-.

Otro caso a resaltar es el de “La isla mínima”. A la vista de los resultados –Figura N.1.1- parecería más conveniente una agrupación en dos grupos en vez de en tres. La distinción entre personajes secundarios y terciarios no parece ser clara, por lo menos en cuanto a las variables aquí tenidas en cuenta. En los siguientes apartados seguirán viéndose diferencias entre éste y los otros guiones.

Las gráficas de presentación de personajes también difieren claramente de guiones a no guiones. En el caso de los guiones la evolución que siguen los personajes principales, los secundarios y los terciarios es distinta. Y en el caso de las actas se puede apreciar que, sobre todo al principio, las tres distribuciones siguen una evolución muy parecida –ver Figuras O1.2 y P.1.2.-.

Si bien es cierto que se puede seguir profundizando en la caracterización de los personajes y refinando aspectos como la detección automática de todos los personajes o las referencias a los mismos a lo largo del guión, los resultados obtenidos son bastante satisfactorios. Y el no haber necesitado de una gran infraestructura para conseguirlos también es una virtud.

4.3 Índices sobre los temas

- Dinámica cíclica de presentación y explotación de temas.

Si bien el objetivo ideal era encontrar algo semejante a los puntos de giro –ver apartado 3.4-, la lectura de los resultados constata que todavía queda camino hasta poder llegar a ellos. No obstante, encontramos algunos resultados remarcables.

Lo primero que hay que apuntar es la dificultad que plantea dicha lectura a la luz de los ejemplos trabajados, pues en este caso contamos ya sólo con los guiones en castellano. Tanto “El laberinto del fauno” como “Tesis” parecen presentar una fuerte estructuración –corroborada además en el resto de índices- pero en cambio en “La isla mínima” parecen estarlo menos. En los dos primeros casos vemos ciclos de subida y bajada de entropía que además incrementan su frecuencia conforme se acerca el final de la historia –ver Figuras de los apartados L.3.1 y M.3.1-. En el tercer caso la distribución parece ser más ruidosa –Figuras del apartado N.3.1-. Con respecto a los contraejemplos, la diferencia es clara en el tramo final, que en los guiones suele ser en alto, y aquí está por los suelos –ver Figuras de los apartados O.3.2 y P.3.2-, aunque bien es cierto que en el caso de las actas organizadas según la estructura de secuencias de la “Isla Mínima” se percibe una mayor estructura con respecto a la versión de las secuencias distribuidas uniformemente. Más sobre la posibilidad de ganancia de estructura a raíz de la organización de las secuencias en el siguiente apartado sobre las tramas.

En definitiva, lo que reflejan los ciclos de subida y bajada de entropía son momentos en los que se ponen nuevos y variados elementos sobre la mesa seguidos de otros en los que la historia se centra en alguno de ellos y los desarrolla, haciendo así avanzar la historia en ciclos de expansión y focalización de los temas de frecuencia creciente. Teniendo esto en cuenta, y quizás en contra de lo que podría parecer a primera vista, podríamos ver en los mínimos de las gráficas los momentos de mayor interés en la historia. Esto estaría indicando que el número de temas expuesto sería menor y, por tanto, más focalizado y, cabría suponer, relacionado con algún tema importante para el desarrollo de la historia-.

Estas medidas de entropía ofrecen un primer acercamiento a la dinámica temporal del guión, frente al resto de métricas que ofrecen más bien fotos fijas de aspectos concretos. Sin embargo, la

aplicación de este tipo de métricas presenta una complejidad mayor que otros índices, ya que requiere tomar ciertas decisiones para analizar estructuras con una gran riqueza -como puede ser un guión cinematográfico- discretizándolo en series de variables simbólicas que pueden limitar el alcance de los resultados obtenidos. Así, la cantidad de variables a tener en cuenta hace que el análisis estadístico con este tipo de métricas sea una labor de gran profundidad, de la que aquí presentamos una primera aproximación. Cabría profundizar en aspectos como en el cálculo de las entropías conjuntas entre pares, triadas, etc. de temas y personajes - en esta primera aproximación únicamente hemos analizado las entropías individuales - o se podría elegir una unidad de muestreo más precisa que la secuencia - que en nuestro caso hemos elegido en consonancia con el resto de métricas aplicadas - y analizar la entropía de series temporales por ejemplo a nivel de diálogo. Esta profundización queda como una línea futura de desarrollo que podría permitir una caracterización más detallada de la estructura de los diferentes guiones cinematográficos.

No obstante, las medidas puestas en marcha han servido para hacerse una idea de las posibilidades que abre el concepto de entropía y aportar una lectura sobre el guión distinta a la del resto de índices. Además de lo muy interesante que resulta la hipótesis del ciclo de despliegue y focalización de temas con una frecuencia cada vez mayor.

4.4 Índices sobre las tramas

- Red de relaciones entre personajes y temas.

El intento de representar la trama de relaciones establecida entre los elementos de la historia se ha cristalizado en la representación en forma de grafo de aquellas distribuciones -de personajes y temas- más correlacionadas -ver apartado 3.5-.

Estos resultados son quizás los más llamativos, pues ponen de manifiesto la estructura de los guiones de una manera bastante interesante. El primer comentario es para los diferentes tipos de subgrafos: por un lado los que conectan personajes que en la historia comparten tramas, y por otro los que sólo conectan temas. En ambos casos, los resultados obtenidos para los guiones que vienen mostrando una mayor estructuración, “El laberinto del fauno” y “Tesis”, son del todo satisfactorios -ver Figuras de los apartados L.4 y M.4-. En el caso de “La Isla Mínima” encontramos en cambio unos personajes y temas bastante poco correlacionados, dando lugar a una multitud de pequeños subgrafos sin ninguna relación temática entre ellos y que además selecciona a personajes muy poco relevantes -ver Figuras del apartado N.4-.

Especialmente destacables son los casos de los contraejemplos. Para el guión que respeta la estructura original dividida por días, el grafo generado está centralizado en los temas, a los cuales se encuentran conectados los personajes como si estuvieran conectados en estrella. La centralidad de los temas en detrimento de los personajes parece captar perfectamente la idea de lo que el Congreso, frente a una película, debería ser -ver Figura Ñ.4.1-.

Para el guión de las 100 secuencias distribuidas uniformemente, el resultado es totalmente pobre, con muy pocas correlaciones y un grafo muy pequeño -ver Figura O.4.1-. En cambio, al imponer una estructura de guión comprobamos que el número de subgrafos aumenta -Figura P.4.1-, aunque el resultado se parece más al grafo desconexo de “La isla mínima” que a cualquiera de los otros dos.

Efectivamente, cabe la duda de si este resultado viene motivado por haber seguido justamente la estructura de dicho guión o simplemente por haber ganado algo de estructura con respecto a la situación previa al organizarse narrativamente. Aún a pesar de requerir pruebas adicionales para

asegurarlos, parece que, se usara la estructura del guión que se usara, la ganancia de estructura no sería tal como para poder llegar a decir que hay una narración. No es sólo la organización de los temas y personajes en general –en secuencias- lo que cuenta sino la de cada tema y personaje concreto –en cada una de sus apariciones- y en relación a la unidad total de la historia.

Parece que encontramos por tanto en estos grafos una buena manera de detectar guiones bien estructurados y de resumir su contenido.

4.5 Índices narrativos

- Habría que replantear la estrategia llevada a cabo para mejorar la división en escenas y actos.

Encontrar las estructuras narrativas de mayor nivel que la secuencia, las escenas y los actos, sería el objetivo que nos planteábamos con estos índices –ver apartado 3.6-.

Los resultados no parecen en cambio tan satisfactorios como cabría esperar. Un examen detallado de las secuencias que han sido agrupadas dentro de la misma escena parece ir en demasiadas ocasiones en contra de la división que uno podría hacer por su cuenta –ver apartados L.5, y M.5-. La métrica que indica lo cohesionado de las escenas resultantes tampoco es del todo informativa. Hay escenas que han sido indudablemente señaladas por las métricas de personajes y temas como una unidad y que no presentan dicha unidad. Aunque el caso contrario sí que se cumple, escenas con un valor de cohesión bajo sí son una importante mezcla de secuencias no relacionadas entre sí. Pero más que con haber acertado en la determinación de la heterogeneidad de las secuencias tiene que ver con que son escenas muy largas, y por tanto la probabilidad de tener más secuencias no relacionadas entre sí aumenta.

Una posible solución sería tener en cuenta las localizaciones –indicadas en la cabecera de las secuencias- a la hora de agrupar en escenas. Cabe esperar que los resultados mejoraran con ello sustancialmente.

En cuanto a los actos, pese a encontrar alguna mejora gracias al factor de corrección teórica, la respuesta no mejora mucho. El problema está entonces en ver si la mejora se debe a que la corrección se impone sobre los cálculos anteriores o es que efectivamente la conjunción de ambas es una solución aceptable.

A la vista de las gráficas de resultados sin y con corrección teórica –ver apartados L.6, M.6 y Q.6-, vemos que la que sería la solución teórica no se impone sobre aquella alcanzada en base a las métricas previas. Puede ser una solución polémica –porque supone un cambio importante en la solución tras muchos cálculos realizados-, pero su misión tan sólo es mover el foco de atención hacia la zona en la que sabemos que debería de haber un cambio de acto. De hecho, es una solución bastante conservativa.

De todas formas, los resultados alcanzados con estas métricas no son del todo satisfactorios. El problema es que estamos tratando con conceptos abstractos y que, al contrario de lo que ocurre con las secuencias, no se encuentran marcados en el propio guión y, por tanto, debemos inferir a partir de una multitud de elementos y relaciones un tanto complejas. Además, esto nos sitúa en cierta medida en contra del espíritu inicial de encontrar una formulación que pueda relacionarse de un modo estrecho con el modo en el que entendemos y tratamos las películas cotidianamente. Una revisión de la estrategia seguida en este punto sería recomendable.

5

Conclusiones

5.1 Conclusiones técnicas

A la vista de los resultados obtenidos podemos ver los problemas y limitaciones con los que nos hemos encontrado. Por ejemplo, las características seleccionadas para la detección de escenas y actos se han demostrado insuficientes. Podemos ver aquí la necesidad y conveniencia de seguir buscando y explotando más elementos del guión a parte de los ya tratados. Por otra parte, algunas de las métricas de las que hacemos uso podrían ser explotadas en mayor profundidad. Este sería el caso de las medidas de entropía, que podrían aplicarse a otra serie de distribuciones más refinadas. Por ejemplo, no sólo los temas o los personajes sino incluso extraer los temas de cada personaje particular. En resumen, profundizar en mayor medida en algunos aspectos.

No en vano, muchos de los problemas enfrentados podrían constituir el objeto de un trabajo de investigación por sí solos –corrección y manipulación automática del formato de los guiones, extracción automática de todos los personajes, selección de las palabras relevantes, problema de desambiguación de sentido....-, pero en ningún momento se ha pretendido mejorar o llevar más allá alguno de éstos -si acaso en el caso en el tema del diccionario, donde costó dar con una solución que se adecuara a lo que teníamos en mente-. En vez de buscar soluciones complejas, la estrategia ha sido más bien explorar las soluciones ya disponibles para ponerlas a trabajar juntas al servicio de la cuestión que nos hemos planteado desde un principio: *¿qué es una historia?*

La pregunta no tiene de momento una respuesta clara y ni siquiera sabemos de qué disciplina cabe esperar que venga. Pero sí que podemos poner las herramientas y conocimientos informáticos a trabajar sobre ello y aportar una visión desde un punto de vista diferente.

Se podrá criticar el hecho mismo de intentar cuantificar lo que probablemente no sea cuantificable, pero hay que insistir en que hay espacio para un estudio de este tipo, que no agota o cierra la definición de lo que es una historia sino que la complementa de la mano de otras muchas disciplinas. Estos intereses entran de lleno en una disciplina conocida como Humanidades digitales que ponen en relación la potencia de las nuevas tecnologías de la información y la comunicación para enfrentar problemas de las clásicas disciplinas humanísticas. Se han intentado operativizar las ideas e intuiciones presentadas a lo largo de la memoria, donde la propuesta para el panel de índices ha sido en todos los casos analizada y diseñada de manera crítica y fundada. Los resultados así lo demuestran en la mayoría de casos, al darnos una imagen coherente con la idea propia que cualquiera nos podríamos formar de la historias aquí trabajadas.

5.2 Conclusiones personales

Visto en perspectiva, el año dedicado al proyecto me ha aportado muchas cosas. En el aspecto técnico, aprendiendo Python he saldado una cuenta personal que tenía con los lenguajes funcionales. También he aprendido a enfrentar problemas muy variados: explorar nuevas áreas (NLP), analizar problemas alejados del dominio más frecuente de la informática, diseñar soluciones sin muchos referentes, integrar y trabajar con código de terceros, profundizar en algunos algoritmos típicos de Inteligencia Artificial, aplicar conceptos de teoría de lenguajes y de compiladores...

En un terreno más personal, tanto las lecturas como la dinámica del trabajo llevado a cabo me han brindado una primera impresión del mundo de la investigación. Y aplicar la informática y las capacidades analíticas que me ha brindado para los problemas que me preocupan ha sido completamente estimulante.

Poder leer y profundizar sobre temas que me interesan –la inteligencia artificial, las historias, el lenguaje...-, pero que en un principio no parecen formar un objeto de estudio común, me ha permitido ampliar los horizontes y comprobar sí hay gente interesada en ello.

5.3 Líneas abiertas

Muchas e interesantes son las líneas que se abren a partir de este punto. Por un lado tenemos algunas conclusiones del propio proyecto que apuntan a nuevas hipótesis que habría que comprobar; y por el otro, ideas que se han ido barajando a lo largo del proceso pero que por unas y otras cuestiones se han ido dejando de lado.

En el primer grupo lo más inmediato sería comprobar –en la misma línea que se ha seguido con el último contraejemplo, ver Anexo P- cómo afecta el imponer la estructura de un guión a otro. Otras vías de expansión pasarían por explotar las localizaciones, llevar a cabo medidas sobre distribuciones más elaboradas –temas de personajes, referencias de personajes a personajes, etc.- o realizar medidas más elaboradas, como por ejemplo las de entropías conjuntas sugeridas en la discusión de los resultados del apartado 4.3.

En el segundo grupo de ideas tenemos propuestas con un gran potencial. Por ejemplo, desarrollar un clasificador de guiones en función de todas las métricas con las que hemos caracterizado los guiones. Catalogarlos podría valer tanto como aporte para teorías sobre géneros cinematográficos como para ayudar en una productora a automatizar en cierta medida en el proceso de cribado de la gran cantidad de guiones que puedan recibir. O incluso determinar si un guion comparte

características o no con otros cuyos resultados en taquilla ya han sido comprobados, e intentar predecir en cierta manera el éxito de una nueva historia (Hall, 2016). Algunas de las técnicas exploradas pero finalmente no implementadas, por ejemplo el análisis de sentimientos, podrían servir de base para montar sistemas como una red de personajes (Valiente, 2011) que contemple las relaciones positivas o negativas que se establecen entre ellos. Una red compleja de este tipo en la que se reflejara el estado del mundo que cada personaje, es decir, las relaciones que cada personaje cree que se establecen entre todos los personajes de la historia, podría ser también una manera de encontrar las fuentes de tensión del relato y los puntos de giro.

A parte de para los guiones cinematográficos, podría pensarse en alguna manera de trasladar todo esto a otros formatos, por ejemplo, el teatro o las novelas.

Y por último, otra posibilidad sería aplicar en la generación automática de historias los conocimientos y las métricas sobre la estructura que hemos obtenido. De manera que, sabiendo qué forma debería tener el tipo de historia que queremos contar, ésta pueda construirse con mayor eficacia.

6

Carga y desarrollo de tiempos

6.1 Gestión de tiempos

El proyecto se ha desarrollado a lo largo de un año, entre mediados de Marzo de 2015 y mediados de Febrero de 2016. El procedimiento seguido en muchas ocasiones ha sido cíclico, con el planteamiento, análisis, diseño e implementación de múltiples problemas que iban surgiendo. No obstante, podemos resumir la distribución de tareas de la siguiente manera:

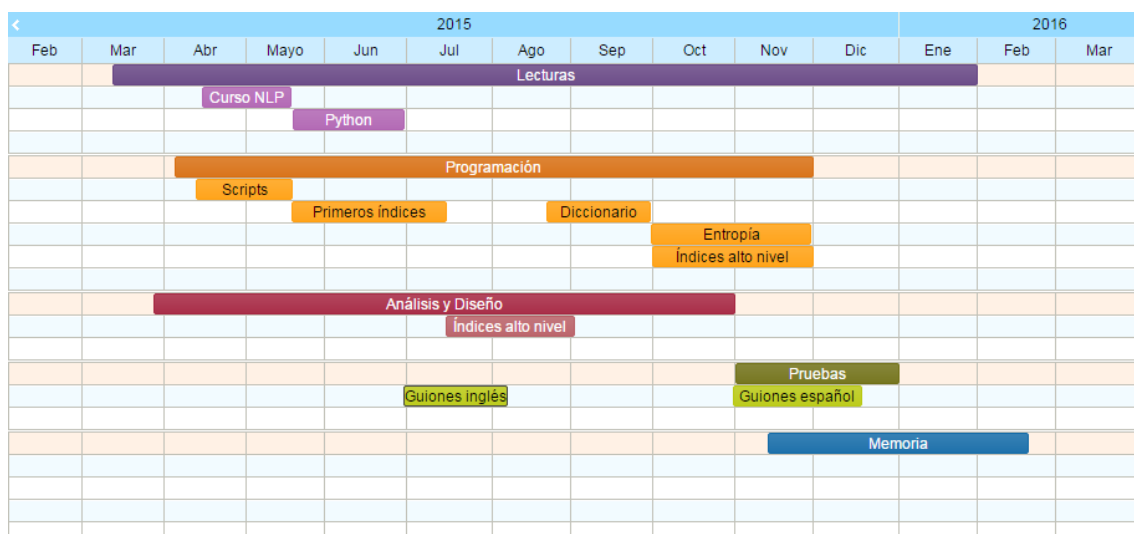


Figura 6.1: Diagrama de Gantt

Se incluye a continuación una descripción de las tareas señaladas en Figura.

- Lecturas. Se engloba en este apartado por un lado la lectura y trabajo de la bibliografía y por otro los dos subapartados tratados en los siguientes puntos. La razón de que esta tarea se extienda a lo largo de la práctica totalidad de la duración del proyecto está en que, si bien el trabajo se ha focalizado al inicio, en todo momento se ha estado abierto a nuevas lecturas y trabajos que no por estar en un momento avanzado del desarrollo han sido obviadas.
- Curso NLP. Uno de los subapartados mencionados anteriormente corresponde a un curso online sobre Procesamiento de Lenguaje Natural al que le fue dedicado un mes de trabajo.
- Python. En este otro subapartado se incluyen varios libros dedicados específicamente a este lenguaje. Aunque el aprendizaje de un lenguaje no termina nunca, la primera toma de contacto en este caso llevó en torno a un mes, mes y medio.
- Programación: Se incluyen aquí no sólo los subapartados destacados a continuación sino también otras labores como búsqueda, instalación y actualización de todo el software utilizado. Es la tarea que mayor tiempo consume de todo.
- Scripts: Pequeñas pruebas realizadas en BASH para testear algunas primeras ideas. Se llevan a cabo a la par que el estudio del curso sobre NLP.
- Primeros índices: Aquí se trata ya de código en Python. Son dos meses dedicados principalmente al preprocesamiento y los índices hasta la parte del Diccionario.
- Diccionario: Comporta las tareas de programación relativas al diccionario como son la modificación y añadidos al código en Java en el que nos basamos y conexión con el programa principal. En conjunto todo esto llevó algo más de un mes.
- Entropía e índices de alto nivel: Ambas tareas son llevadas a cabo a la vez durante dos meses de manera conjunta.
- Análisis y Diseño: Esta tarea también se ha extendido a lo largo de todo el proyecto, llevándose a cabo el análisis de los resultados de una fase previa y realizando el diseño de la siguiente. Por ejemplo después de las primeras lecturas y antes de empezar con los scripts y después de éstos y antes de la fase de implementación en Python. A parte de estos momentos puntuales hay un periodo de preparación de la última fase de programación que correspondería con la tarea señalada como “Índices de alto nivel”.
- Pruebas: La mayor parte del tiempo de pruebas se ha dedicado a conseguir –digitalizar- y preparar los guiones. También quedan englobados en este punto los tiempos de ejecución del programa y el posterior análisis de sus resultados.
- Memoria: El proceso de elaboración de la presente memoria es la última tarea que se lleva a cabo. La dedicación a la misma se intensifica durante el último mes y medio.

6.2 Tiempo invertido por tareas

Tarea	Número de horas
Lecturas	192
Curso NLP	65
Python	30
Libros y Papers	97
Programación	432
Scripts	78
Primeros índices	97
Diccionario	132
Entropía e índices de alto nivel	125
Análisis y Diseño	161
Índices alto nivel	53
Otros	108
Pruebas	86
Obtención guiones (inglés)	23
Obtención guiones (castellano)	45
Ejecución pruebas	18
Memoria	90
TOTAL	961

Tabla 6.2: Horas dedicadas por tarea

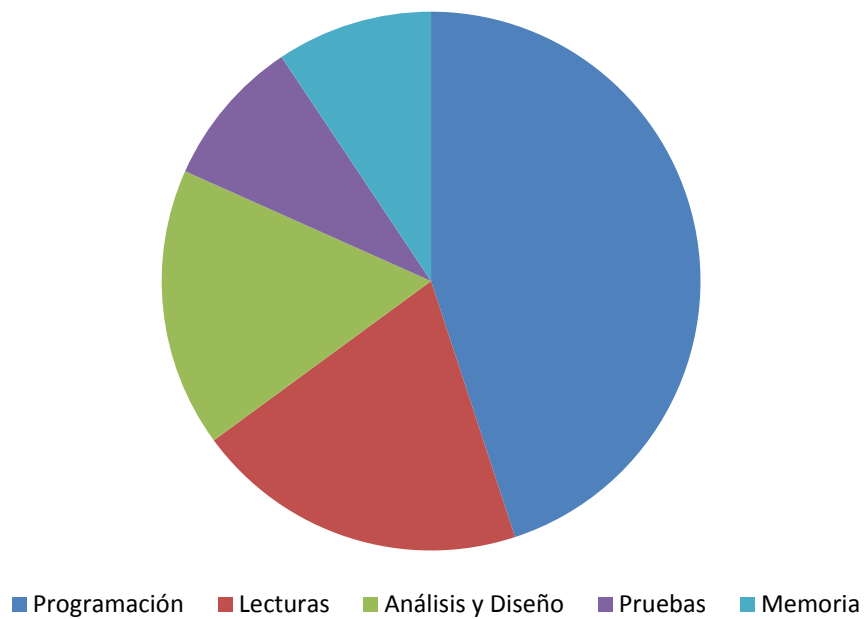


Figura 6.3: Horas dedicadas por tarea

Referencias

- Adams, B., Dorai, C., & Venkatesh, S. (2001, January). Automated film rhythm extraction for scene analysis. In ICME 2001: Proceedings of the IEEE International Conference on Multimedia and Expo (pp. 1056-1059). IEEE.
- Aristoteles (2013). Poética. CreateSpace Independent Publishing Platform.
- Barthes, R. & Duisit, L. (1975). An introduction to the structural analysis of narrative. *New Literary History*, 6(2):237–272.
- Blau, J. J., Petrusz, S. C., & Carello, C. (2013). Fractal Structure of Event Segmentation: Lessons From Reel and Real Events. *Ecological Psychology*, 25(1), 81-101.
- Boychuk, E., Paramonov, I., Kozhemyakin, N., & Kasatkina, N. (2014, April). Automated approach for rhythm analysis of french literary texts. In Open Innovations Association FRUCT, Proceedings of 15th Conference of (pp. 15-23). IEEE.
- Bringsjord, S. & Ferrucci, D. (1999). Artificial Intelligence and Literary Creativity: Inside the mind of Brutus, a StoryTelling Machine. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Bruner, J. (1990). Actos de significado más allá de la revolución cognitiva.
- Bruner, J. (2003). La fábrica de historias: derecho, literatura, vida. Fondo de cultura económica.
- Bruner, J. (1986). Realidad mental y mundos posibles. Barcelona. Gedisa.
- Burke, K. (1969). A grammar of motives. Univ of California Press.
- Chalmers, D. W., Francis, C. C., Pepper, N. M., & Bedau, M. A. (2010). High-content Words in Patent Records Reflect Key Innovations in the Evolution of Technology. In ALIFE (pp. 838-845).
- Cunningham, D. Maynard, K. Bontcheva, and V. Tablan (2002). GATE: an Architecture for Development of Robust HLT Applications. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 7–12 July 2002, ACL '02, pages 168–175, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.
- Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013, May). No country for old members: User lifecycle and linguistic change in online communities. In Proceedings of the 22nd international conference on World Wide Web (pp. 307-318). International World Wide Web Conferences Steering Committee.
- Dewey, J (2007). Cómo pensamos: La relación entre pensamiento reflexivo y proceso educativo. Paidós

- Guzdial, M., Harrison, B., Li, B., & Riedl, M. O. (2014). Crowdsourcing Open Interactive Narrative. In the 10th International Conference on the Foundations of Digital Games. Pacific Grove, CA.
- Hall, J (2016). Scriptbook will predict if a screenplay will make money; will also possibly, maybe trigger nuclear armageddon. <https://www.slashfilm.com/scriptbook-software/>
- Hargood, C., Millard, D., & Weal, M. (2011). Measuring narrative cohesion: A five variables approach.
- Herman, D. (2000). Narratology as a cognitive science. *Image and Narrative*.
- <https://www.narrativescience.com/narrative-analytics-white-paper>
- <https://www.nlp.lsi.upc.edu/freeling/doc/userman/html/>
- Hutto, D. D. (2012). Folk psychological narratives: The sociocultural basis of understanding reasons. MIT Press.
- Zhang, B., & Jhala, A. (2014, June). Mid-Scale Shot Classification for Detecting Narrative Transitions in Movie Clips. In Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence.
- Jarmasz, M. (2003). Roget's Thesaurus as a Lexical Resource for Natural Language Processing
- Jordan, A. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, 14, 841.
- Kassarnig, V. (2016). Political Speech Generation. arXiv:1601.03313
- Klein, D., & Manning, C. D. (2003, July). Accurate unlexicalized parsing. In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1 (pp. 423-430). Association for Computational Linguistics.
- Klein, S., Aeschliman, J. F., Balsiger, D., Converse, S. L., Court, C., Foster, M., Lao, R., Oakley, J. D., and Smith, J. (1973). Automatic novel writing: A status report. Technical Report 186, Computer Science Department, The University of Wisconsin, Madison, Wisconsin.
- Latasa, L. (2011). Sistemas Analizador y Recomendador de Blogs. Ingeniería Informática. Universidad de Zaragoza.
- Lebowitz, M. (1985). Storytelling as Planning and Learning. *Poetics*, 14:483–502.
- Mateas, M. and Sengers, P. (1999). Narrative Intelligence: An Introduction to the NI Symposium. In *Working Notes of the Narrative Intelligence Symposium*, pages 1–10.
- McKee, R. *El guión*. Alba Editorial.
- Meehan, J. (1976). *The Metanovel: Writing Stories by Computer*. PhD thesis, Yale University.
- Miller, G. A. 1995. WordNet: a lexical database for English. *Commun. ACM*, 38:39–41.
- Ouyang J, McKeown K. (2014). Towards Automatic Detection of Narrative Structure. *LREC 2014*: 4624-4631
- Pavel, A., Goldman, D. B., Hartmann, B., & Agrawala, M. (2015, November). SceneSkim: Searching and Browsing Movies Using Synchronized Captions, Scripts and Plot Summaries. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 181-190). ACM.
- Peinado, F., Navarro, A., and Gervás, P. (2008). A Testbed Environment for Interactive Storytellers. In *International Conference on Intelligent Technologies for Interactive Entertainment*. ACM Digital Library.
- Propp, V. (1998). *Morfología del cuento* (Vol. 139). Ediciones Akal.

-
- Reyes, A., & Rosso, P. (2012). Making objective decisions from subjective data: Detecting irony in customer reviews. *Decision Support Systems*, 53(4), 754-760.
 - Riedl, M. (2004). *Narrative Planning: Balancing Plot and Character*. PhD thesis, Department of Computer Science, North Carolina State University.
 - Riedl, M. and Young, R. (2006). Story Planning as Exploratory Creativity: Techniques for Expanding the Narrative Search Space. *New Generation Computing, Computational Paradigms and Computational Intelligence. Special Issue: Computational Creativity*, 24(3):303–323
 - Rumelhart, D. E. (1975). Notes on a schema for stories. *Representation and Understanding: Studies in Cognitive Science*, pages 211–236.
 - Schank, R. C. (1982). *Dynamic Memory : A Theory of Reminding and Learning in Computers and People*. Cambridge University Press.
 - Soto, RM and Garcia, RA (2009). Comparación de tres modelos de texto para la generación automática de resúmenes. *Procesamiento de Lenguaje Natural*, 43 (2009), pp- 303-311
 - Theune, M., Faas, E., Nijholt, A., and Heylen, D. (2003). The virtual storyteller: Story creation by intelligent agents. In *Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment*, pages 204–215.
 - Todorov, T. (1977). *The Grammar of Narrative*, pages 108–119. *The Poetics of Prose*. Ithaca: Cornell UP
 - Turner, S. (1992). *MINSTREL: A Computer Model of Creativity and Storytelling*. PhD thesis, University of California at Los Angeles, Los Angeles, CA, USA.
 - Valiente, C. (2012). *Caracterización de procesos digitales en Historia Digital*. Ingeniería Informática. Universidad de Zaragoza.
 - Van Trijp, R., Steels, L., Beuls, K., & Wellens, P. (2012, April). Fluid construction grammar: The new kid on the block. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 63-68). Association for Computational Linguistics.
 - Wasserman, M., Zeng, X. H. T., & Amaral, L. A. N. (2015). Cross-evaluation of metrics to estimate the significance of creative works. *Proceedings of the National Academy of Sciences*, 112(5), 1281-1286.
 - Zhang, B., & Jhala, A. (2014, June). Mid-Scale Shot Classification for Detecting Narrative Transitions in Movie Clips. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

ANEXOS

A

Preprocesado

Antes de pasar a calcular los índices es necesario procesar el guión para comprobar que siguen las convenciones esperadas así como realizar algunas tareas básicas como por ejemplo eliminar líneas superfluas. El proceso incorpora algunas de las fases iniciales que encontraríamos en un compilador y añade algunas otras pensadas a propósito del contexto del proyecto. En general, la pipeline consta de las siguientes fases:

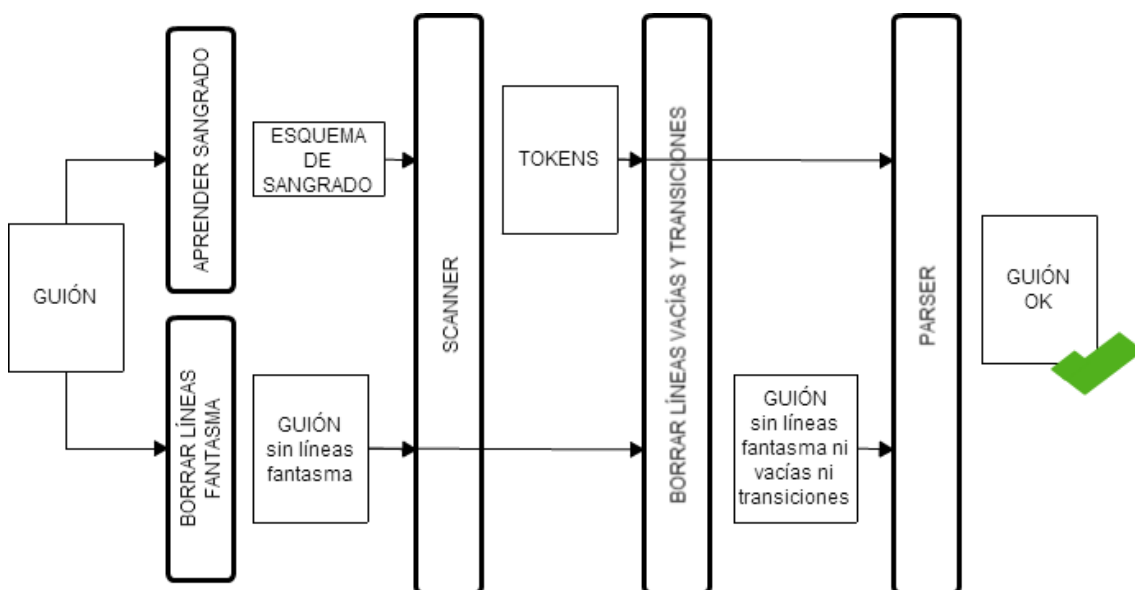


Figura A.1: Esquema de la pipeline de preprocesamiento

- **Aprender sangrado.** Recibe como entrada el guión en bruto y ofrece como salida el esquema de sangrado del guión.
- **Borrar líneas fantasma.** Recibe como entrada el guión en bruto y el esquema de sangrado, y como salida devuelve el guión habiendo detectado y borrado aquellas líneas que presentan un sangrado válido pero que no tiene ningún contenido, líneas que parecen vacías pero no lo son.

```

Para Elías.txt
11» INT. SALÓN - DÍA
2
3 Montaje de un vídeo casero. Imagen granulada
4 luz. MARÍA (35), la madre, y ELÍAS (5), el hi
5 cada uno en un taburete. Ella intenta enseñar
6 chico anda despistado mirando a la cámara. Su
7 la escena, habla desde detrás de la cámara.
8
9 » » » » » PADRE (V.O.)
10 » » » ¡Elías, Elías! Haz caso a mamá.
11 » » »
12 » » »
13
14 MARÍA llama cuidadosamente la atención de su
15 tocar las primeras notas de LA MELODÍA. ELÍAS
16 cortando abruptamente la preciosa MELODÍA.
17
18 » » » » » MARÍA
19 » » » (Suavemente)
20 » » » Hay que tratar bien al piano, car

Para Elías.txt
11» INT. SALÓN - DÍA
2
3 Montaje de un vídeo casero. Imagen granulada
4 luz. MARÍA (35), la madre, y ELÍAS (5), el hi
5 cada uno en un taburete. Ella intenta enseñar
6 chico anda despistado mirando a la cámara. Su
7 la escena, habla desde detrás de la cámara.
8
9 » » » » » PADRE (V.O.)
10 » » » ¡Elías, Elías! Haz caso a mamá.
11
12 MARÍA llama cuidadosamente la atención de su
13 tocar las primeras notas de LA MELODÍA. ELÍAS
14 cortando abruptamente la preciosa MELODÍA.
15
16 » » » » » MARÍA
17 » » » (Suavemente)
18 » » » Hay que tratar bien al piano, car
19
20 Coge las manos del niño y las coloca sobre el

```

Figura A.2: Eliminación de líneas fantasma

- **Scanner (anализador léxico).** Recibe como entrada el guión sin líneas fantasma y el esquema de sangrado y devuelve un stream de tokens. Los tokens o etiquetas disponibles para asignar a cada línea son: secuencia, descripción, diálogo, nombre de personaje, acotación, transición y línea vacía.

```

Para Elías.txt
11 INT. SALÓN - DÍA
2
3 Montaje de un vídeo casero. Imagen granulada
4 luz. MARÍA (35), la madre, y ELÍAS (5), el hi
5 cada uno en un taburete. Ella intenta enseñar
6 chico anda despistado mirando a la cámara. Su
7 la escena, habla desde detrás de la cámara.
8
9 PADRE (V.O.)
10 ¡Elías, Elías! Haz caso a mamá.
11
12 MARÍA llama cuidadosamente la atención de su
13 tocar las primeras notas de LA MELODÍA. ELÍAS
14 cortando abruptamente la preciosa MELODÍA.
15
16 MARÍA
17 (Suavemente)
18 Hay que tratar bien al piano, car
19
20 Coge las manos del niño y las coloca sobre el

Para Elías.tok
2 "ISSEQUENCE",
3 "LN",
4 "DESCRIPTIO",
5 "DESCRIPTIO",
6 "DESCRIPTIO",
7 "DESCRIPTIO",
8 "DESCRIPTIO",
9 "LN",
10 "CHARACTERN",
11 "DIALOGTEXT",
12 "LN",
13 "DESCRIPTIO",
14 "DESCRIPTIO",
15 "DESCRIPTIO",
16 "LN",
17 "CHARACTERN",
18 "ANNOTATION",
19 "DIALOGTEXT",
20 "LN",
21 "DESCRIPTIO",

```

Figura A.3: Etiquetado de las líneas

- **Borrar líneas vacías y transiciones.** Recibe como entrada el guión sin líneas fantasma y el stream de tokens. Y la salida es un guión sin líneas fantasma, líneas vacías seguidas ni transiciones. Este es un paso importante para homogeneizar la longitud de los guiones.

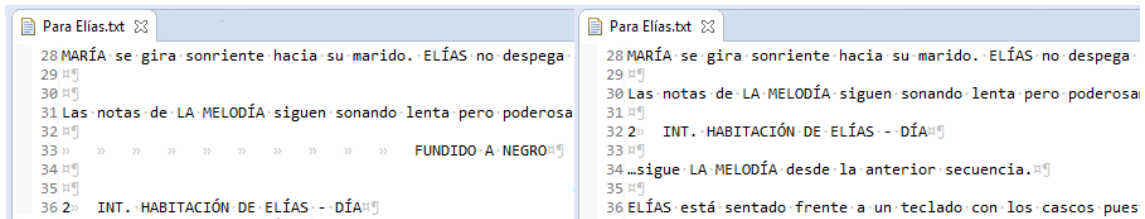


Figura A.4: Eliminación de líneas vacías y transiciones

- **Parser (analizador sintáctico).** Recibe como entrada el guión sin líneas fantasma, líneas vacías seguidas ni transiciones y el stream de tokens y como resultado informa del cumplimiento o no del mismo con respecto a la estructura convencional de un guión cinematográfico.

De entre estos procesos hay dos destacables: el aprendizaje del sangrado y el parser.

A.1 Aprendizaje del sangrado

La primera estrategia tomada para descomponer el guión en sus partes no proporcionaba resultados satisfactorios. A grandes rasgos, el peso de la solución recaía en expresiones regulares y variables booleanas acerca de lo que se había hecho o dejado de hacer en anteriores iteraciones, resultando en un código nada intuitivo.

En una segunda iteración, se replantea la estrategia intentando simplificarla. Al fin y al cabo, cualquiera que se enfrente a la hoja de un guión no realiza operaciones tan exactas como ponerse a contar a ver si hay entre 20 ó 40 espacios ó 5 u 8 tabuladores. Más bien se sabe que hay unos ciertos elementos que tienen que aparecer (descripciones, diálogos, acotaciones, nombres de personaje...) y que además lo harán siguiendo un mismo patrón a lo largo del guión. (Pavel, 2015).

Esta idea es la que se quiere implementar en la parte de aprendizaje del sangrado. Sabiendo los elementos que pueden aparecer y el orden de sangrado que deben seguir (encabezados de secuencia y descripciones tienen el mismo nivel de sangrado, después los diálogos, etc...), se extraen los espaciados iniciales de todas las líneas y se clasifican en función de su longitud. Luego, si no hay más tipos de espaciados de los esperados –en cuyo caso se informaría de las líneas en las que esto ocurre para su revisión- se procede a asignar a cada uno de ellos una etiqueta que los identifique. Con respecto a la solución anterior, se ha introducido la detección y etiquetado de transiciones. Para el caso de las secuencias nos basamos también en expresiones regulares, aunque relajando en este caso la expresión con respecto a la primera versión. Y para el etiquetado de los nombres de personajes lo que se hace es comprobar que están escritos en mayúsculas; y para las acotaciones, que empiecen por paréntesis.

A.2 Parser

La estructura de un guión cinematográfico que venimos tratando es muy sencilla. Podemos representarla a través del siguiente autómata:

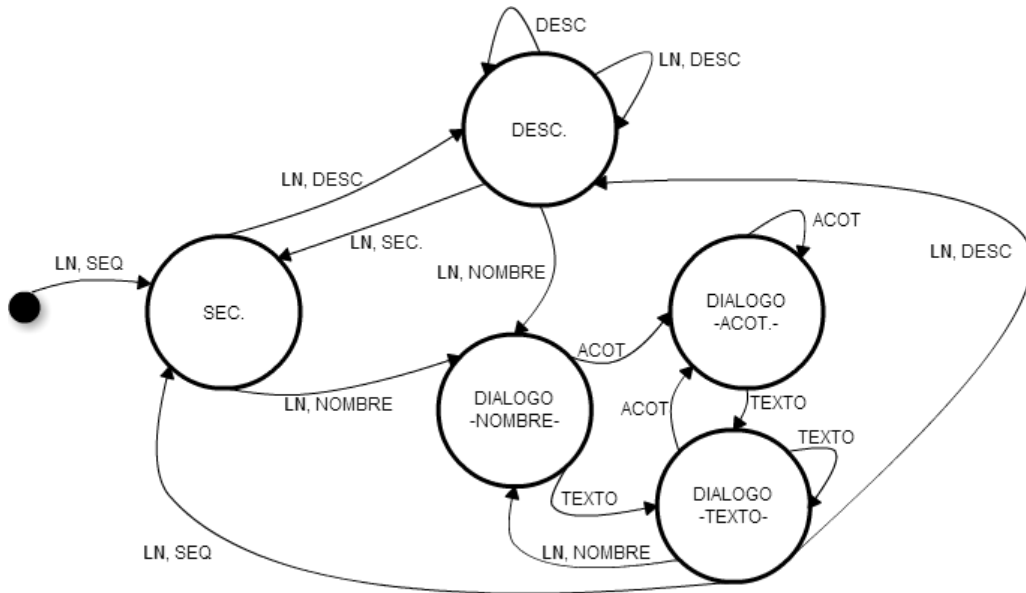


Figura A.5: Autómata reconocedor de guiones cinematográficos

Se puede ver, por ejemplo, que todo comienzo de secuencias, descripción o diálogo debe estar precedido por una línea vacía (LN). O que las descripciones son bloques de texto continuo o bloques separados por una línea vacía, es decir, párrafos.

A estas conexiones se les van a añadir otros casos que, aun no siendo válidos, se dan con frecuencia y pueden ser solucionados “al vuelo” sin necesidad de parar la ejecución y relegar la responsabilidad del arreglo en el usuario. Son casos en los que el espaciado entre los bloques no es el requerido y es fácil añadir o quitar líneas vacías.

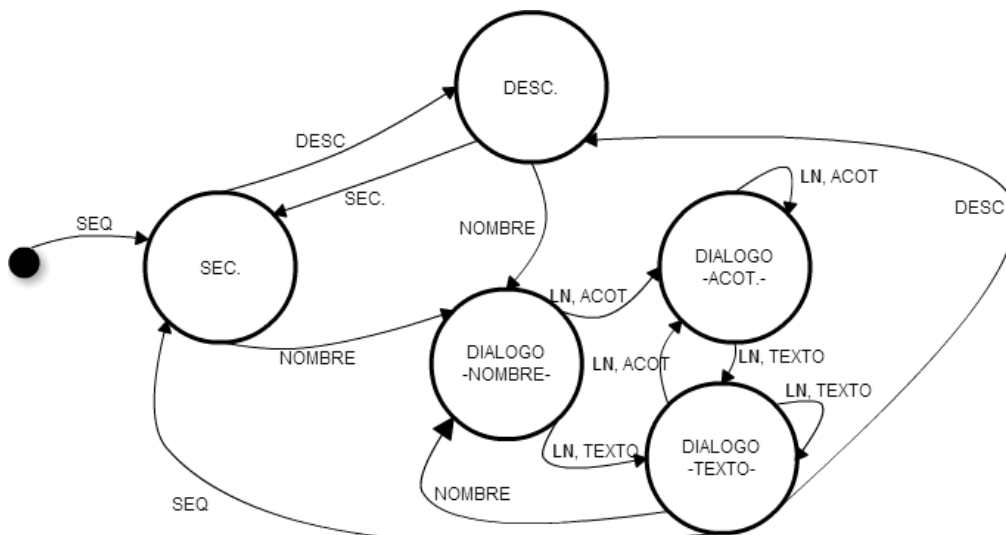


Figura A.6: Transiciones no válidas pero recuperables

Estos son ejemplos de casos erróneos que el parser es capaz de solucionar.

```

Para Elías.txt
23
24 » » » » » » » PADRE (V.O.)
25
26 » » » » » (En tono cómplice)
27
28 » » » Va a ser tan bueno o más que su madre.
29
30 MARÍA se gira sonriente hacia su marido. ELÍAS no despega.

Para Elías.txt
23
24 » » » » » » » PADRE (V.O.)
25 » » » » » (En tono cómplice)
26 » » » » » Va a ser tan bueno o más que su madre.
27
28 MARÍA se gira sonriente hacia su marido. ELÍAS no despega.
29
30 Las notas de LA MELODÍA siguen sonando lenta pero poderosa
  
```

Figura A.7: Eliminación automática de líneas vacías

```

Para Elías.txt
348 » » » » » allá.
349 » » » » » » » ELÍAS
350 » » » » » No, no, no.
351 7 INT. BAR - DÍA
352 Es un bar no muy grande con un pequeño escenario a
353 tocando. La gente lo está pasando bien.
354
355 ELÍAS se encuentra entre el público. Se fija en lo
356 RAÚL y sus amigos. Una sonrisa se dibuja en su car

Para Elías.txt
348 » » » » » allá.
349
350 » » » » » » » ELÍAS
351 » » » » » No, no, no.
352
353 7 INT. BAR - DÍA
354
355 Es un bar no muy grande con un pequeño escenario a
356 tocando. La gente lo está pasando bien.
  
```

Figura A.8: Añadido automático de líneas vacías.

B

Comparativa de algoritmos para la obtención de temas

El objetivo de este Anexo es el de profundizar en la explicación acerca del funcionamiento de los algoritmos para la obtención de los temas. Una primera descripción se ha dado ya en el apartado 3.4.2.1. Lo que aquí se pretende es dar un ejemplo detallado de cómo se llega a los resultados, mostrar la diferencia entre la ejecución local –por secuencias- y global y el porqué de diferentes soluciones de una a otra ejecución.

Supongamos que tenemos un guión con dos secuencias. Seleccionamos para la primera las tres palabras más importantes, y para la segunda las cuatro palabras más importantes. De cada una de estas palabras obtenemos del diccionario tres sentidos distintos, y para cada uno de estos sentidos varias ideas asociadas. Para simplificar el proceso nos quedamos con la primera de todas las ideas.

Podemos representar el escenario en una tabla como la que sigue:

Secuencia 1	Palabra 1	Sentido 1	Idea 1: A
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: C
	Palabra 2	Sentido 1	Idea 1: B
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: X
	Palabra 3	Sentido 1	Idea 1: C
		Sentido 2	Idea 1: C
		Sentido 3	Idea 1: H
Secuencia 2	Palabra 4	Sentido 1	Idea 1: H
		Sentido 2	Idea 1: H
		Sentido 3	Idea 1: C
	Palabra 5	Sentido 1	Idea 1: P
		Sentido 2	Idea 1: P
		Sentido 3	Idea 1: P
	Palabra 6	Sentido 1	Idea 1: A
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: C
	Palabra 7	Sentido 1	Idea 1: H
		Sentido 2	Idea 1: P
		Sentido 3	Idea 1: H

Tabla B.1: Distribución de palabras, sentidos e ideas para un ejemplo con dos secuencias

B.1 Algoritmo de Fuerza Bruta

B.1.1 Ejecución por secuencias

Para la Secuencia 1, en una primera iteración obtenemos los siguientes datos:

IDEA	PALABRAS Y SENTIDOS EN DONDE APARECE		
A	Palabra 1, Sentido 1		
B	Palabra 1, Sentido 2	Palabra 2, Sentido 1	Palabra 2, Sentido 2
C	Palabra 1, Sentido 3	Palabra 3, Sentido 1	Palabra 3, Sentido 2
X	Palabra 2, Sentido 3		
H	Palabra 3, Sentido 3		

Tabla B.1.1: Palabras y sentidos compartidos por las ideas para la Secuencia 1, 1ª iteración

- Se ordena de mayor a menor número de relaciones: B, C, X, H, A
- Se seleccionan las ideas con más relaciones: B, C
- Se ordena de mayor a menor número de palabras relacionadas: B, C
- Se selecciona el primer elemento: B o C

Dentro de las ideas con un mismo número de relaciones el orden es indiferente. Y como se señala en el último punto, tanto B como C presentan el mismo número de relaciones, por lo que cualquiera podrá ser elegida. De hecho, en función de cuál sea seleccionada obtendremos una respuesta u otra.

En una segunda iteración, si B ha resultado elegida, las palabras que restan por ser asignada a ideas han disminuido:

Secuencia 1	Palabra 1	Sentido 1	Idea 1: A
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: C
	Palabra 2	Sentido 1	Idea 1: B
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: X
	Palabra 3	Sentido 1	Idea 1: C
		Sentido 2	Idea 1: C
		Sentido 3	Idea 1: H

Tabla B.1.2: Distribución de palabras, sentidos e ideas para la Secuencia 1 seleccionado B

Los datos obtenidos en este caso son:

IDEA	PALABRAS Y SENTIDOS EN DONDE APARECE	
C	Palabra 3, Sentido 1	Palabra 3, Sentido 2
H	Palabra 3, Sentido 3	

Tabla B.1.3: Palabras y sentidos compartidos por las ideas en la Secuencia 1 seleccionado B, 2ª iteración

- Se ordenan de mayor a menor número de relaciones: C, H
- Se seleccionan las ideas con más relaciones: C
- Se ordenan de mayor a menor número de palabras relacionadas: C
- Se selecciona el primer elemento: C

No quedarían más palabras por asignar a ideas, por lo tanto, se terminaría la ejecución. El resultado sería que los temas de la Secuencia 1 son: B, con una importancia de $\frac{2}{3}$, y C, con una importancia de $\frac{1}{3}$.

En una segunda iteración, si C ha resultado elegida, las palabras que restan por ser asignada a ideas han disminuido y son diferentes al caso en el que se seleccionó B:

Secuencia 1	Palabra 1	Sentido 1	Idea 1: A
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: C
	Palabra 2	Sentido 1	Idea 1: B
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: X
	Palabra 3	Sentido 1	Idea 1: C
		Sentido 2	Idea 1: C
		Sentido 3	Idea 1: H

Tabla B.1.4: Distribución de palabras, sentidos e ideas para la Secuencia 1, seleccionando C

Los datos obtenidos en este caso son:

IDEA	PALABRAS Y SENTIDOS EN DONDE APARECE	
B	Palabra 2, Sentido 1	Palabra 2, Sentido 2
X	Palabra 2, Sentido 3	

Tabla B.1.5: Palabras y sentidos compartidos por las ideas en la Secuencia 1, seleccionado C, 2ª iteración

- Se ordenan de mayor a menor número de relaciones: B, X
- Se seleccionan las ideas con más relaciones: B
- Se ordena de mayor a menor número de palabras relacionadas: B
- Se selecciona el primer elemento: B

No quedarían más palabras por asignar a ideas, por lo tanto, se terminaría la ejecución. El resultado sería que los temas de la Secuencia 1 son: C, con una importancia de $\frac{2}{3}$, y B, con una importancia de $\frac{1}{3}$.

Cualquiera de las dos distribuciones valdría como respuesta: o B, con una importancia de $\frac{2}{3}$, y C, con una importancia de $\frac{1}{3}$ o C, con una importancia de $\frac{2}{3}$, y B, con una importancia de $\frac{1}{3}$.

Para la Secuencia 2, en una primera iteración obtenemos los siguientes datos:

IDEA	PALABRAS Y SENTIDOS EN DONDE APARECE			
A	Palabra 6, Sentido 1			
B	Palabra 6, Sentido 2			
C	Palabra 4, Sentido 3	Palabra 6, Sentido 3		
P	Palabra 5, Sentido 1	Palabra 5, Sentido 2	Palabra 5, Sentido 3	Palabra 7, Sentido 2
H	Palabra 4, Sentido 1	Palabra 4, Sentido 2	Palabra 7, Sentido 1	Palabra 7, Sentido 3

Tabla B.1.6: Palabras y sentidos compartidos por las ideas en la Secuencia 2, 2ª iteración

- Se ordena de mayor a menor número de relaciones: P, H, C, A, B
- Se seleccionan las ideas con más relaciones: P, H
- Se ordena de mayor a menor número de palabras relacionadas: P, H
- Se selecciona el primer elemento: P o H

De nuevo, en función de si se elige a P o a H obtendremos una respuesta u otra. En el primer caso además todavía se abrirán más adelante varias posibilidades. Las posibles distribuciones a las que llegamos al final son:

- H con importancia $\frac{1}{2}$, P con importancia $\frac{1}{4}$, A con importancia $\frac{1}{4}$.
- H con importancia $\frac{1}{2}$, P con importancia $\frac{1}{4}$, B con importancia $\frac{1}{4}$.
- H con importancia $\frac{1}{2}$, P con importancia $\frac{1}{4}$, C con importancia $\frac{1}{4}$.
- P con importancia $\frac{1}{2}$, C con importancia $\frac{1}{2}$.

B.1.2 Ejecución a nivel global

A la hora de ejecutar el algoritmo globalmente se tienen en cuenta las apariciones de una idea en todas las palabras del guión, no se contabilizan por secuencias como en el caso anterior. Así pues, siguiendo con el ejemplo planteado, los datos que obtendríamos serían los siguientes:

IDEA	PALABRAS Y SENTIDOS EN DONDE APARECE				
A	Palabra 1, Sentido 1	Palabra 6, Sentido 1			
B	Palabra 1, Sentido 2	Palabra 2, Sentido 1	Palabra 2, Sentido 2	Palabra 6, Sentido 2	
C	Palabra 1, Sentido 3	Palabra 3, Sentido 1	Palabra 3, Sentido 2	Palabra 4, Sentido 3	Palabra 6, Sentido 3
P	Palabra 5, Sentido 1	Palabra 5, Sentido 2	Palabra 5, Sentido 3	Palabra 7, Sentido 2	
H	Palabra 3, Sentido 3	Palabra 4, Sentido 1	Palabra 4, Sentido 2	Palabra 7, Sentido 1	Palabra 7, Sentido 3
X	Palabra 2, Sentido 3				

Tabla B.1.7: Palabras y sentidos compartidos por las ideas a nivel global. 1ª iteración

- Se ordena de mayor a menor número de relaciones: H, C, P, B, A, X
- Se seleccionan las ideas con más relaciones: H, C
- Se ordena de mayor a menor número de palabras relacionadas: C, H
- Se selecciona el primer elemento: C

Destacar de entre los anteriores el tercer punto. La idea C se relaciona con cuatro palabras: la 1, la 3, la 4 y la 6, mientras que H se relaciona con 3 palabras: la 3, la 4 y la 7.

Al eliminar las palabras asociadas a la idea C, las posibilidades para la siguiente iteración se reducen:

Secuencia 1	Palabra 1	Sentido 1	Idea 1: A
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: C
	Palabra 2	Sentido 1	Idea 1: B
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: X
	Palabra 3	Sentido 1	Idea 1: C
		Sentido 2	Idea 1: C
		Sentido 3	Idea 1: H
Secuencia 2	Palabra 4	Sentido 1	Idea 1: H
		Sentido 2	Idea 1: H
		Sentido 3	Idea 1: C
	Palabra 5	Sentido 1	Idea 1: P
		Sentido 2	Idea 1: P
		Sentido 3	Idea 1: P
	Palabra 6	Sentido 1	Idea 1: A
		Sentido 2	Idea 1: B
		Sentido 3	Idea 1: C
	Palabra 7	Sentido 1	Idea 1: H
		Sentido 2	Idea 1: P
		Sentido 3	Idea 1: H

Tabla B.1.8: Distribución de palabras, sentidos e ideas a nivel global, 2ª iteración

Siguiendo con la ejecución de los pasos como venimos haciendo llegamos a la solución, que en este caso es única: C con importancia $\frac{4}{7}$, P con importancia $\frac{2}{7}$ y B con importancia $\frac{1}{7}$.

Lo que este ejemplo pone de manifiesto es la facilidad con la que se pueden obtener distintas distribuciones temáticas. Y también cómo un tema muy importantes en una secuencia –en este ejemplo H- pueden no salir reflejados en la configuración global. Del mismo modo, el tema más importante globalmente podría aparecer en las distribuciones locales nada más que residualmente.

B.2 Algoritmo Genético

Como se explica en el apartado referido al principio del Anexo, este algoritmo se basa fundamentalmente en la aleatoriedad: para establecer los individuos de origen, para mezclar las distintas partes de los mejores individuos de anteriores iteraciones al generar los nuevos individuos y para las mutaciones. Por eso se presentan directamente los resultados que el algoritmo devuelve al introducir el mismo ejemplo tratado en el apartado anterior.

- Secuencia 1: C con importancia $\frac{2}{3}$, B con importancia $\frac{1}{3}$.
Secuencia 2: C con importancia $\frac{1}{4}$, P con importancia $\frac{1}{2}$ y B con importancia $\frac{1}{4}$.
- Secuencia 1: C con importancia $\frac{1}{3}$, B con importancia $\frac{2}{3}$.
Secuencia 2: C con importancia $\frac{1}{4}$, P con importancia $\frac{1}{2}$ y H con importancia $\frac{1}{4}$.
- Secuencia 1: H con importancia $\frac{1}{3}$, B con importancia $\frac{2}{3}$.
Secuencia 2: B con importancia $\frac{1}{4}$, P con importancia $\frac{1}{2}$ y H con importancia $\frac{1}{4}$.
- Secuencia 1: C con importancia $\frac{2}{3}$, X con importancia $\frac{1}{3}$.
Secuencia 2: C con importancia $\frac{1}{2}$ y P con importancia $\frac{1}{2}$.

En el caso de la Secuencia 1 tan sólo las dos primeras opciones eran posibles en el caso del algoritmo por fuerza bruta. En el caso de la Secuencia 2 esto ocurre sólo con la última respuesta. En todas las demás H es seleccionada con importancia $\frac{1}{4}$, cuando en el otro algoritmo lo es con valor $\frac{1}{2}$. En cualquier caso, estas respuestas son perfectamente válidas, como puede comprobarse a la vista de la Tabla C.1.

C

El guión literario cinematográfico

En este Anexo se pretenden recopilar juntas todas las ideas sobre el guión cinematográfico manejadas a lo largo de la Memoria. Algunas son meras cuestiones de terminología y otras convenciones de uso habitual.

Un ejemplo de un guión real puede encontrarse en el apartado Q.7.

C.1 Secuencia, escena y acto

La primera distinción la realizamos entre secuencia y escena. Con la primera nos referimos a la estructura narrativa mínima de la historia, que se encuentra además señalada en el guión. Toda secuencia debería suponer idealmente el progreso o desarrollo de algún elemento de la historia o, en otras palabras, debería aportar algo al conjunto y, por tanto, eliminarla del guión resultaría en una pérdida de coherencia interna.

La escena por su parte sería un conjunto de secuencias relacionadas por un mismo tema. Puede por tanto extenderse varias secuencias manteniéndose de una localización a otra. Por ejemplo, la escena de una persecución. Estamos ante una estructura narrativa de mayor nivel que la secuencia.

Dependiendo de las referencias que manejemos o la tradición en que nos movamos, el sentido de estas palabras –secuencia y escena- puede intercambiarse. Así ocurre sin ir más lejos en la anglosajona.

Por último hablamos del acto. Normalmente encontramos tres en las historias, que se corresponden con las conocidas partes de presentación, nudo y desenlace. El final de cada uno de estos actos vendría marcado por un punto de giro o de no retorno, que supone uno de los momentos más relevantes de la historia y sobre el que se articula el resto de la trama. Esta supone la estructura narrativa de mayor nivel.

C.2 Cambios de acto

En una película de 100 minutos y, por tanto, de unas 100 páginas, entendemos que el cambio del I al II acto debe producirse en torno a la página 20, con una variación de ± 10 páginas. Es decir, más o menos a los 20 minutos del principio.

En el caso del II al III acto este momento debería producirse teóricamente alrededor de la página 85 ± 5 páginas. Es decir, en el último cuarto de hora. En este caso el intervalo es más reducido, pues conforme se acerca el final de la película las acciones se supone que se aceleran.

C.3 Una página, un minuto

Una regla de uso completamente extendido es el que establece la equivalencia entre una hoja de guión y un minuto de metraje en la pantalla. Bien es cierto que esta no es una regla matemática y que depende de muchos factores, como por ejemplo la dirección, el montaje o el contenido de la secuencia (no sería lo mismo un pasaje de diálogo o uno de acción). Pero, en cualquier caso, si se respeta el formato –tipografía Courier New, tamaño 12 de letra...- el resultado se ajustará bastante a lo esperado. La importancia de seguir esta convención radica en el hecho de que el guión es una herramienta fundamental para todo el proceso de producción posterior, por lo que cuanto mejor refleje éste el que debería ser el resultado final de la película –duración incluida-, mejor.

D

Resultados de bajo nivel

Se adjunta a continuación la Tabla D.1 y unos breves párrafos comentando los resultados sobre el género o de bajo nivel –ver apartado 3.2-.

En cuanto a longitud, “Tesis”, “Seven” y “The Matrix” son las más largas. No parece aventurado relacionar duración con complejidad de la trama. No obstante, “La isla mínima”, un thriller también, es el guión más corto.

En cuanto a número de secuencias tenemos, en número descendente: “Star Wars Episodio V”, “The Matrix”, “Tesis”, “El laberinto del fauno”, “Seven”... Que se desarrolle la acción en muchos lugares puede ser signo del dinamismo en la película –por las localizaciones, no por la trama en sí-. Llamamos la atención “Reservoir Dogs” y, sobre todo, “Cómo entrenar a tu dragón 2”, que no da la sensación de tener tan poca variedad de secuencias. Sería interesante saber si es una característica común de las películas de animación.

En cuanto al ratio líneas por secuencia se podría establecer un valor medio en torno a las 40 líneas por secuencia. Estas son más o menos las líneas que tiene una hoja, por lo que podríamos estar hablando de una media de una secuencia por minuto. No es descabellado, aunque obviamente sabemos que la duración de las escenas es muy variable. Pero quizás se pueda tener una idea del ritmo en función de si el ratio queda por encima o por debajo de esa cantidad. Destacarían en este caso “The Matrix” y “Star Wars Episodio V”, que son películas de acción. “Cómo entrenar a tu dragón 2” y “Reservoir Dogs” no las tendríamos en cuenta por ser tan discordantes con el resto.

A la luz de los números de secuencias interiores y exteriores vemos que dominan claramente las escenas interiores. Excepto en el caso de “Star Wars Episodio VI” y “Cómo entrenar a tu dragón 2”. Quizás se deba de nuevo a ser una película de animación. En “Seven” o “Tesis” se pone de manifiesto la gran proporción de interiores. Sin embargo, “La isla mínima” –otro thriller- se desmarca de nuevo al no presentar una diferencia tan marcada. El entorno y el paisaje son parte importante de la película. Llamamos la atención “The Matrix”, donde el número de interiores parece a priori elevado para una película de acción. Hay que recordar no obstante que fue una película de presupuesto bastante ajustado. “Star Wars Episodio V” en cambio está bastante equilibrada.

Las líneas de diálogo dominan también claramente sobre las de descripción. Una razón de peso al respecto es el formato: las líneas de diálogos son más cortas y, por tanto, el mismo contenido ocupa más espacio. Lo que sí podemos apreciar es que en general los números están bastante igualados. Con la excepción de “Tesis”, “Reservoir Dogs” y las Actas que, como bien se señala en la tabla, presentan diferencias significativas de +2000 líneas. Después de haber visto cómo estas historias sobresalían en algún caso anterior, lo vemos también aquí. Quizás habría que destacar también a “Cómo entrenar a tu dragón”, aunque el hecho de ser una película de animación puede que tenga que ver.

	INGLES										CASTELLANO									
	BR	Dragon	SW V	Matrix	Dogs	Se7en	SW VI	Elías	Testigos	Fauno	Tesis	Isla	Actas4	pruebas	Actas100	ActasIsla				
Nº líneas	5755	5100	6435	6508	5021	6954	5092	673	589	4961	7085	4990	4771	5129	5175					
Nº de secuencias	128	39	277	217	45	155	138	17	11	161	195	110	4	106	111					
Ratio líneas/secuencia	44.95	130.74	23.23	29.98	111.55	44.85	36.89	39.52	54.27	30.81	36.33	45.35	1192.75	48.38	46.62					
Nº secuencias interiores,	94	11	159	180	33	115	66	17	11	105	180	69	4	106	111					
Nº secuencias exteriores	34	28	118	37	12	40	72	0	0	56	15	41	0	0	0					
Nº líneas de descripción	2558	2228	2358	2874	1019	2664	2178	237	237	2295	1931	1997	1445	1469	1465					
Nº líneas de diálogo	2940	2793	3523	3199	3911	3979	2638	401	338	2344	4764	2772	3318	3448	3488					
	+400	+500	+1200	+300	+3000	+1500	+500	+170	+100	+100	+2700	+700	+2000	+2000	+2000					

Tabla D.1: Resultados métricas de bajo nivel

E

Blade Runner

E.1 Personajes

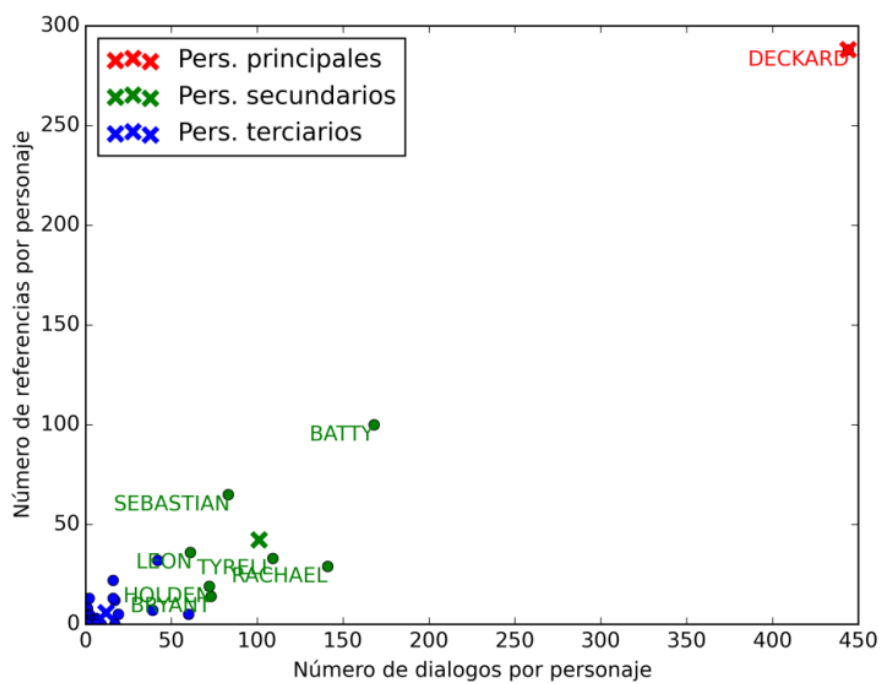


Figura E.1.1: Protagonismo de los personajes en "Blade Runner"

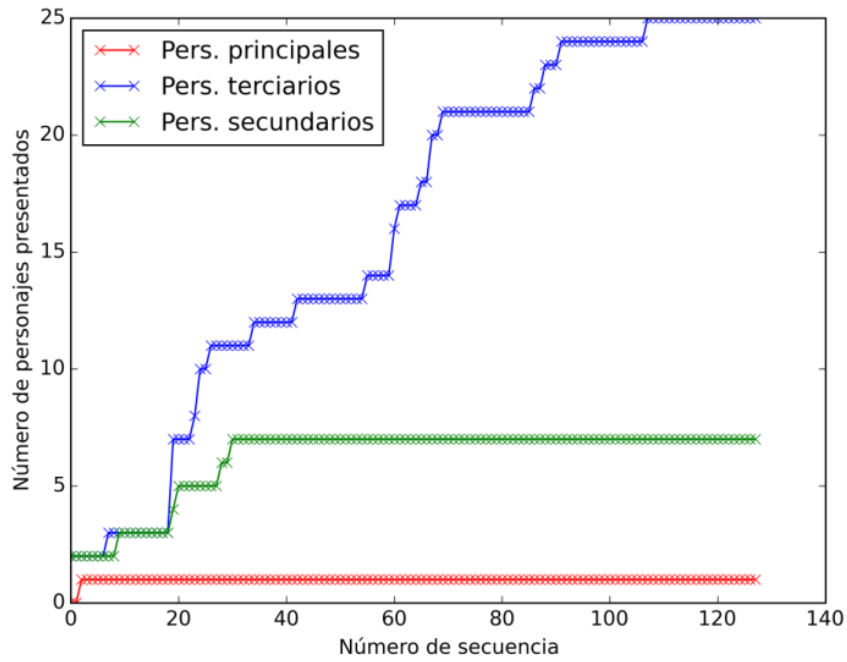


Figura E.1.2: Presentación de los personajes en “Blade Runner”

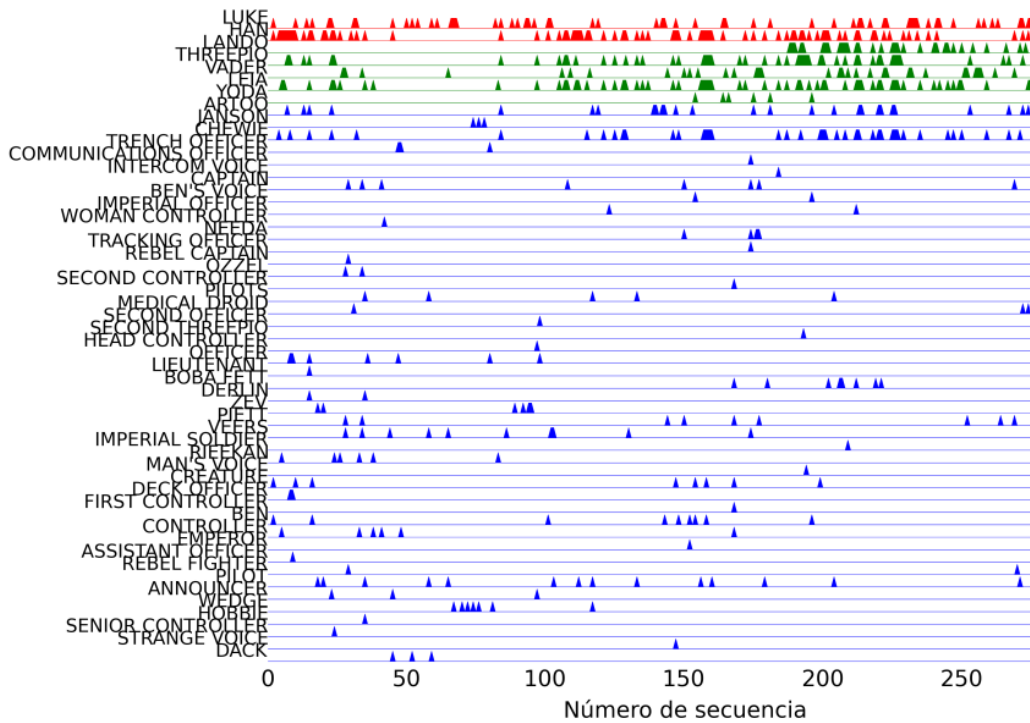


Figura E.1.3: Aparición de los personajes en “Blade Runner”

F

Cómo entrenar a tu dragón 2

F.1 Personajes

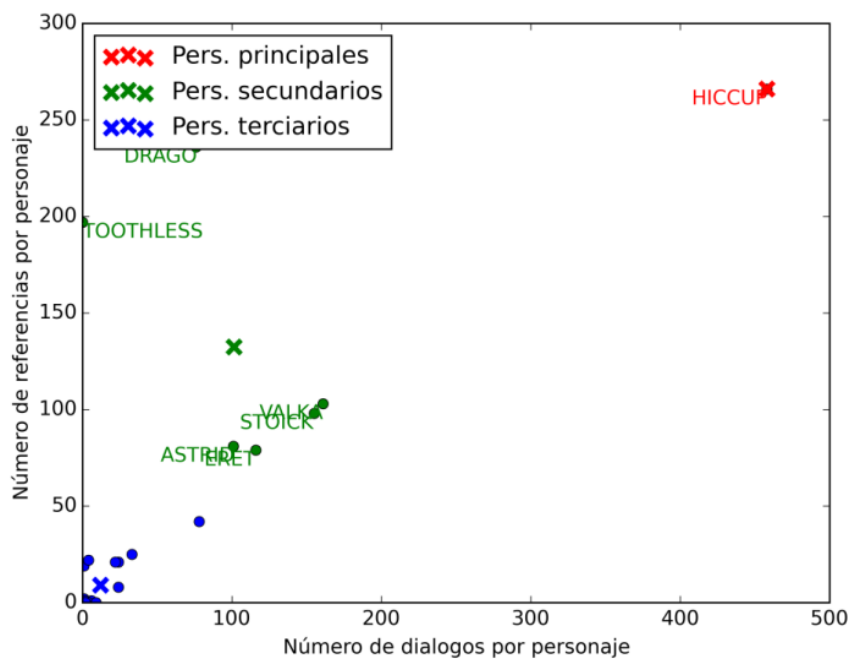


Figura F.1.1: Protagonismo de los personajes en "Cómo entrenar a tu dragón 2"

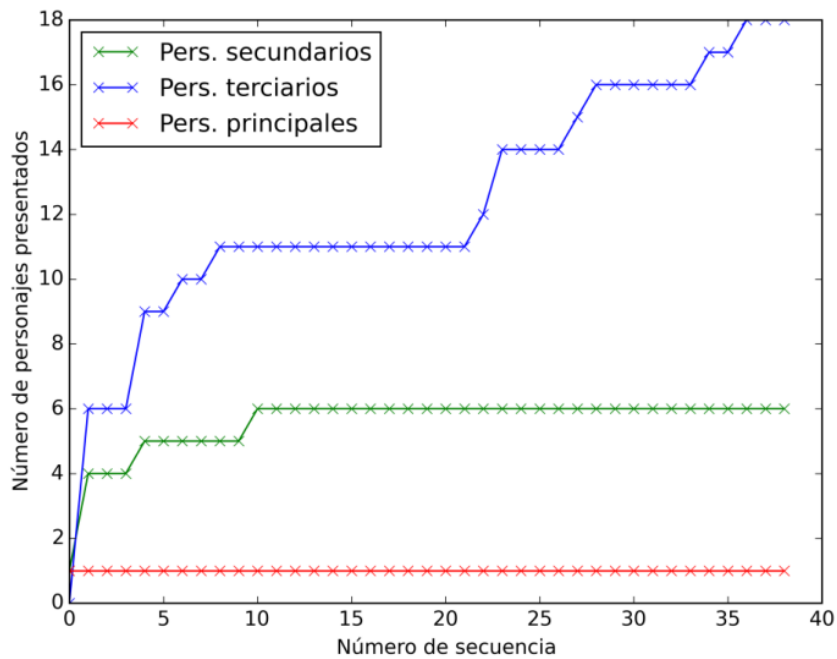


Figura F.1.2: Presentación de los personajes en “Cómo entrenar a tu dragón 2”

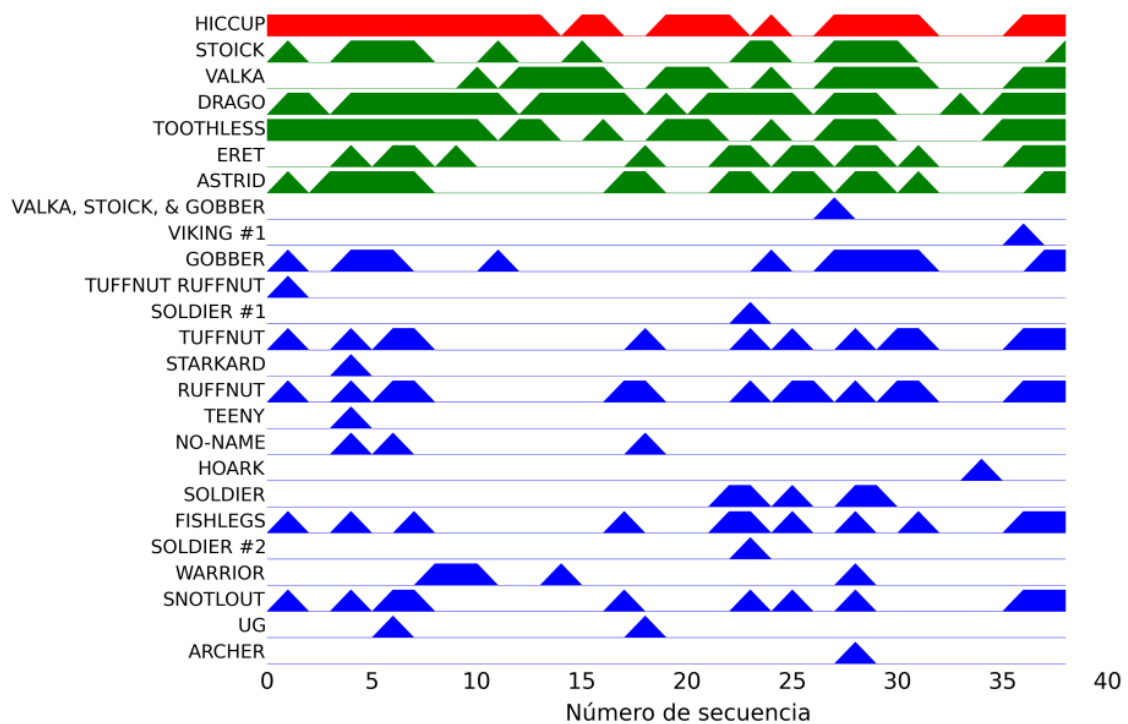


Figura F.1.3: Aparición de los personajes en “Cómo entrenar a tu dragón 2”

G

Reservoir Dogs

G.1 Personajes

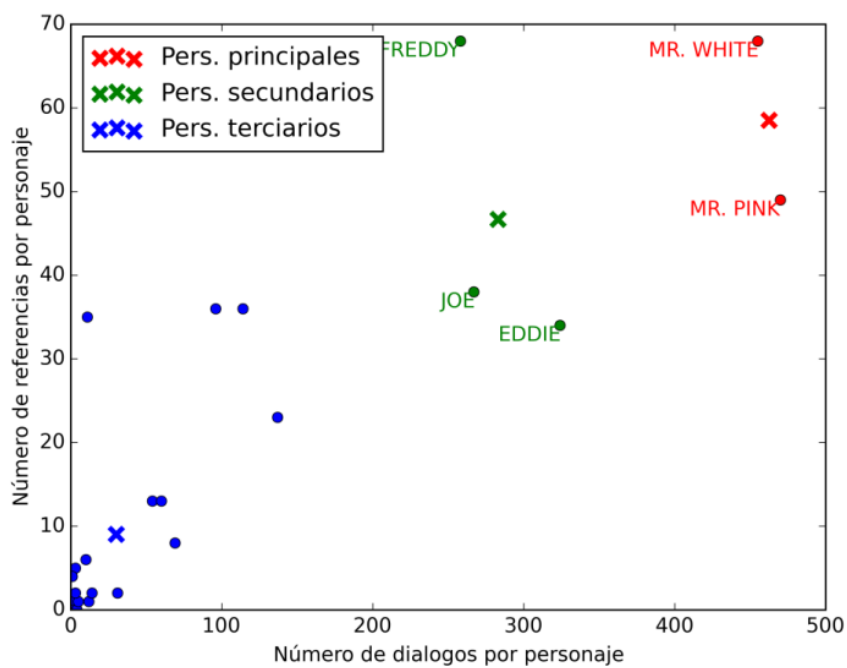


Figura G.1.1: Protagonismo de los personajes en "Reservoir Dogs"

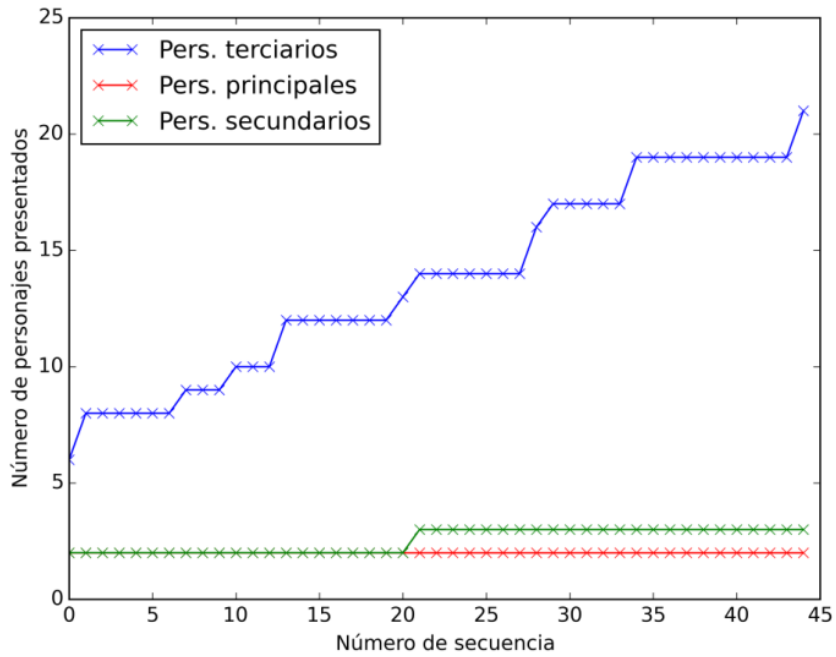


Figura G.1.2: Presentación de los personajes en "Reservoir Dogs"

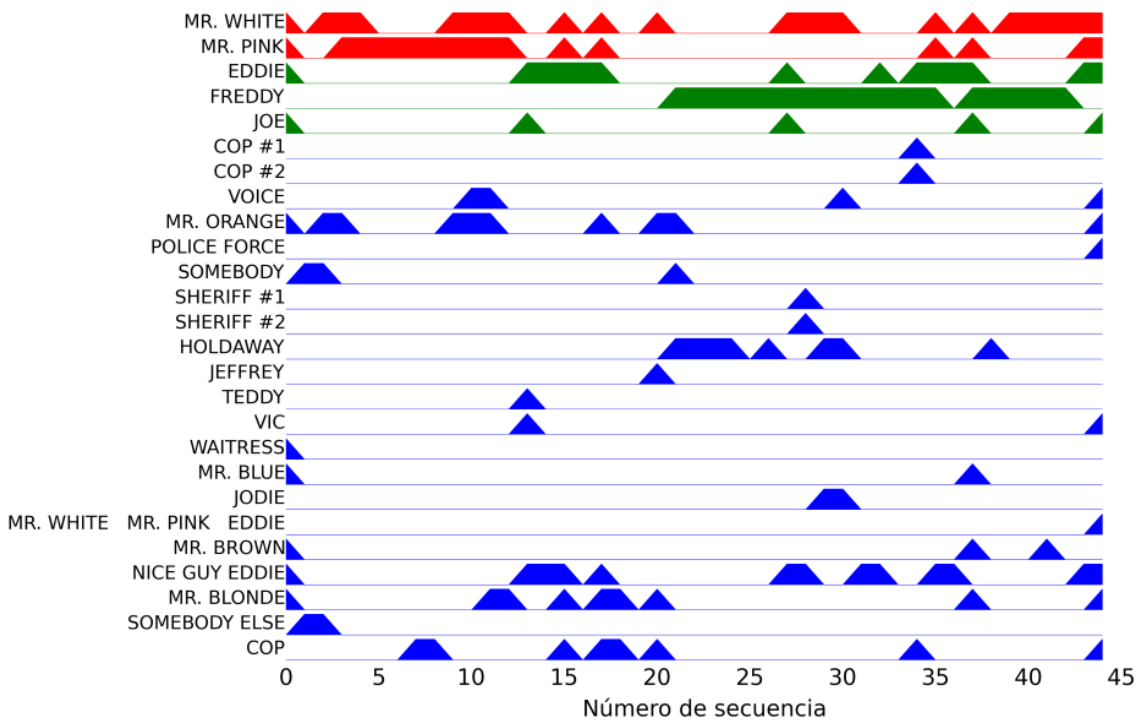


Figura G.1.3: Aparición de los personajes en "Reservoir Dogs"

H

The Matrix

H.1 Personajes

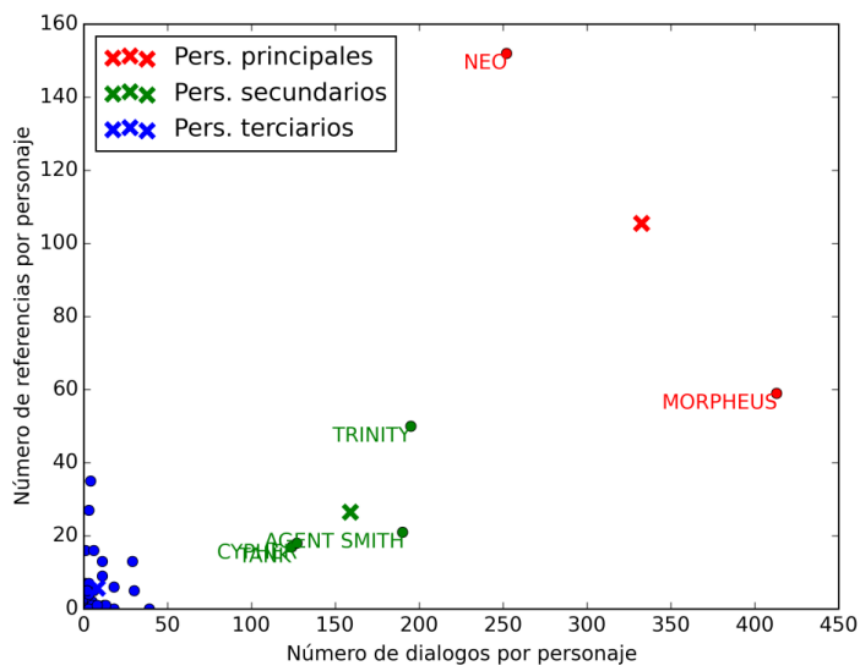


Figura H.1.1: Protagonismo de los personajes en "The Matrix"

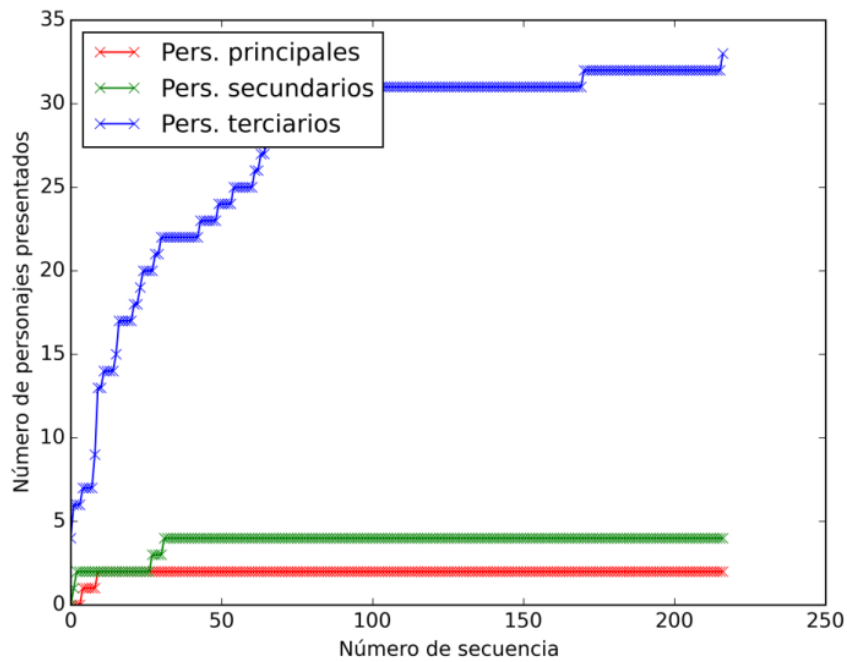


Figura H.1.2: Presentación de los personajes en “The Matrix”

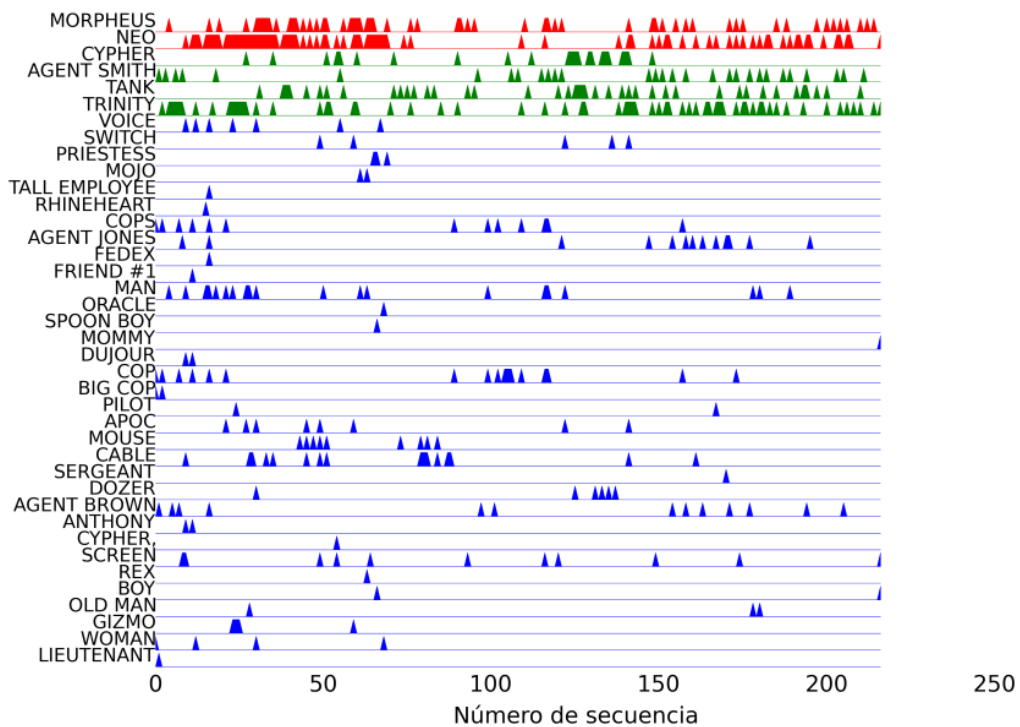


Figura H.1.3: Aparición de los personajes en “The Matrix”

I

Se7en

I.1 Personajes

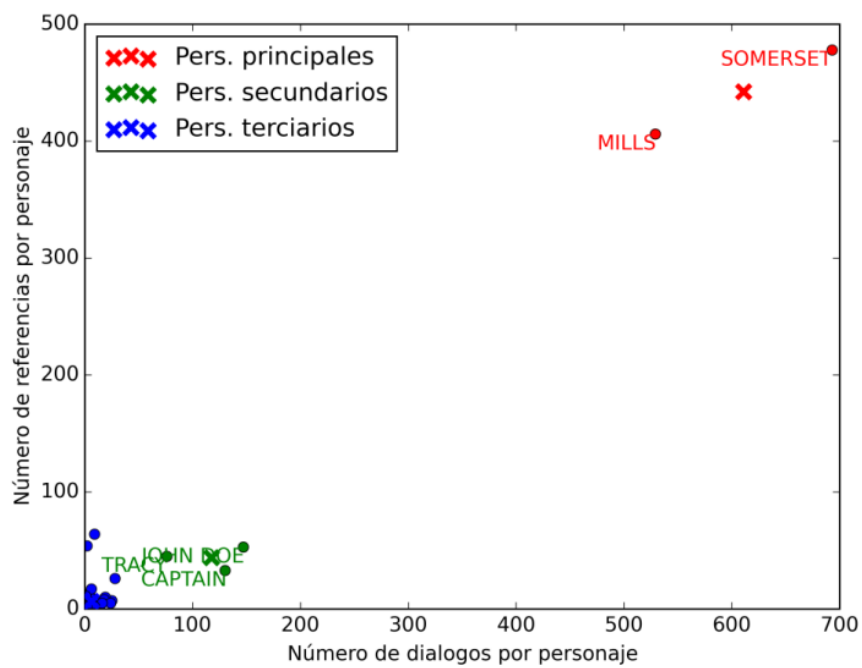


Figura I.1.1: Protagonismo de los personajes en "Se7en"

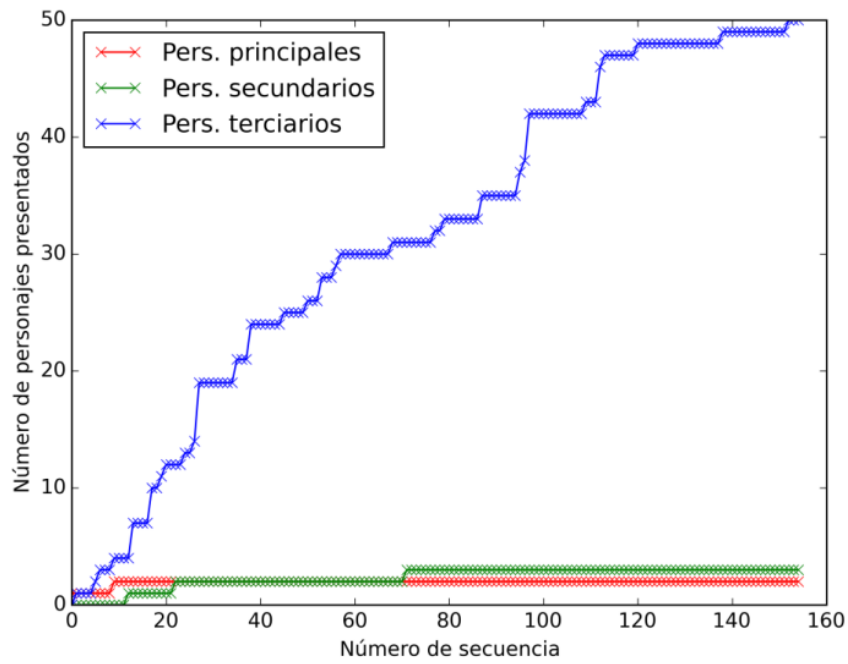


Figura I.1.2: Presentación de los personajes en “Se7en”

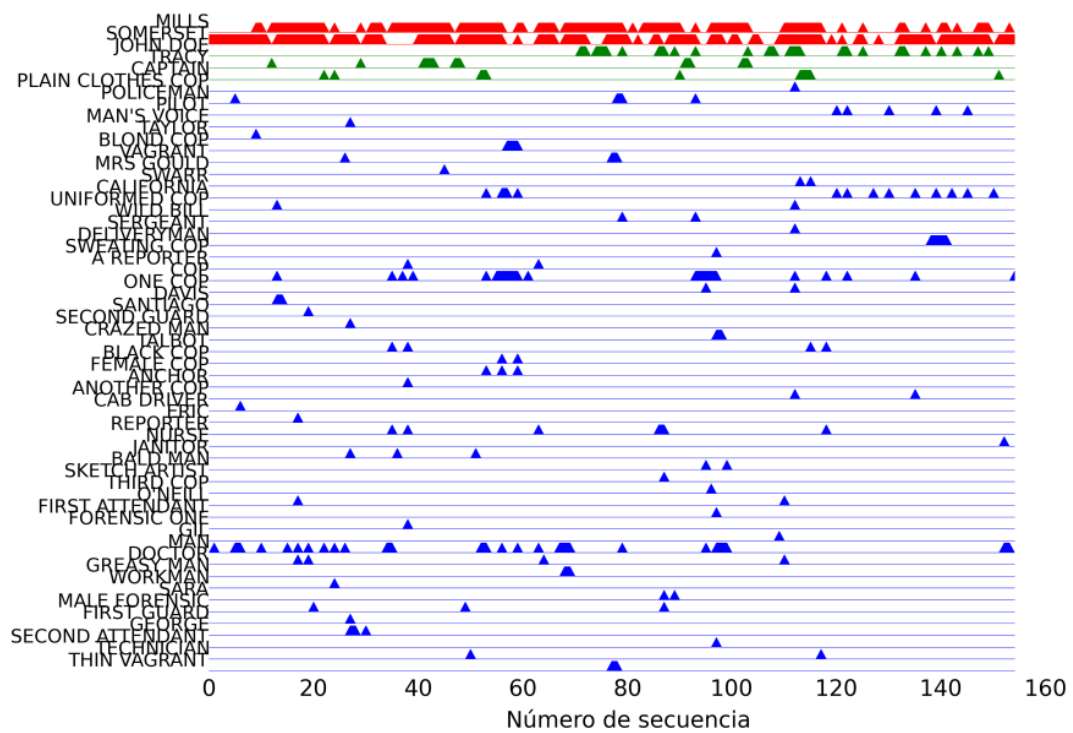


Figura I.1.3: Aparición de los personajes en “Se7en”

J

Star Wars: Episodio V – El Imperio Contraataca

J.1 Personajes

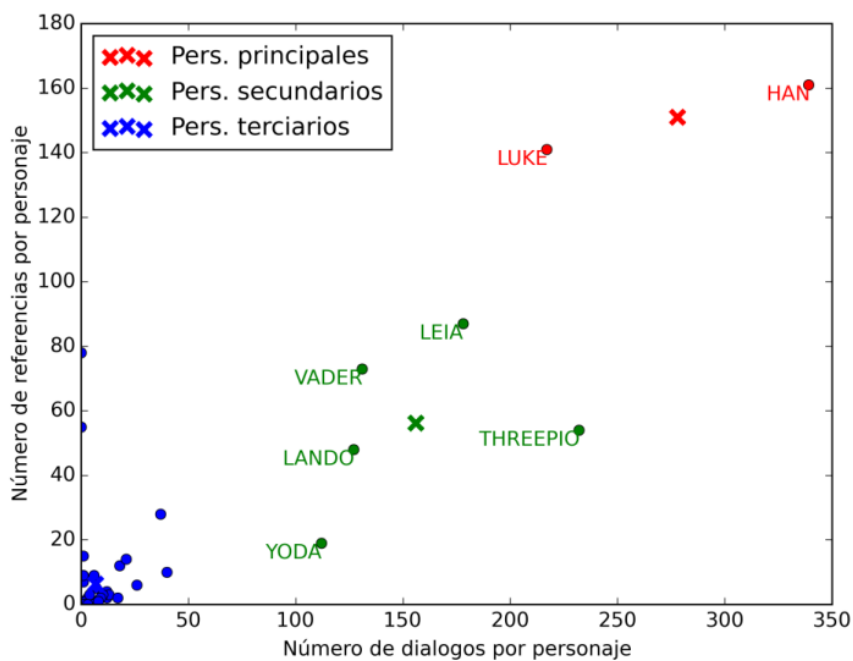


Figura J.1.1: Protagonismo de los personajes en “Star Wars: Episodio V – El Imperio Contraataca”

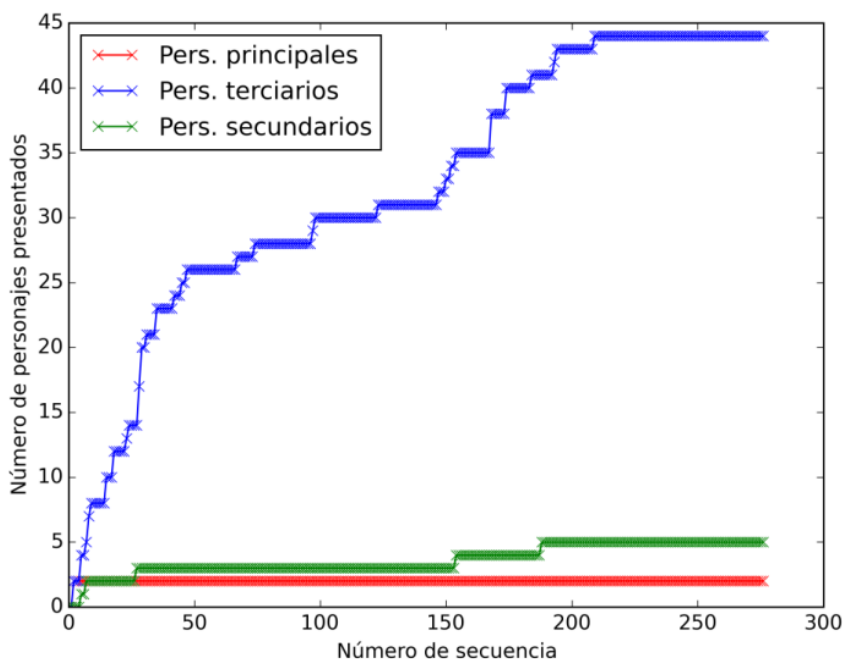


Figura J.1.2: Presentación de los personajes en “Star Wars: Episodio V – El Imperio Contraataca”

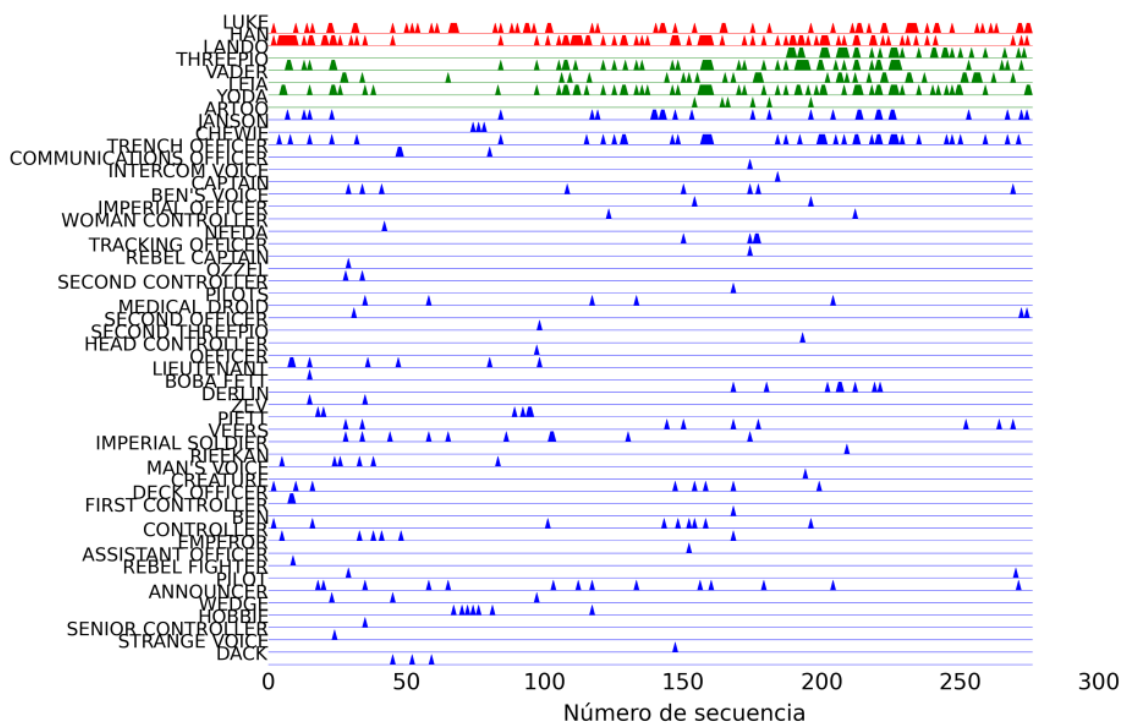


Figura J.1.3: Aparición de los personajes en “Star Wars: Episodio V – El Imperio Contraataca”

K

Star Wars: Episodio VI – El Retorno del Jedi

K.1 Personajes

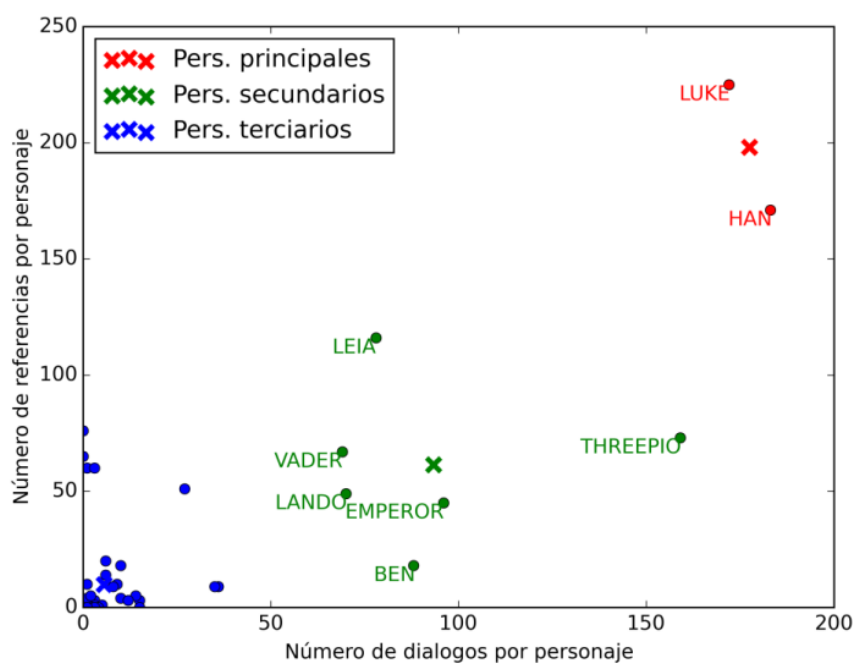


Figura K.1.1: Protagonismo de los personajes en “Star Wars: Episodio VI – El Retorno del Jedi”

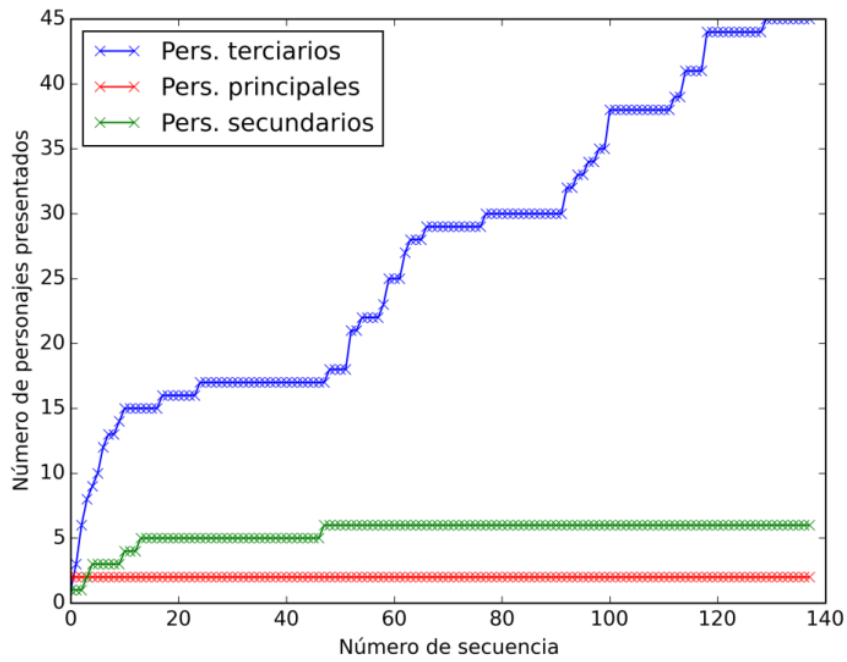


Figura K.1.2: Presentación de los personajes en “Star Wars: Episodio VI – El Retorno del Jedi ”

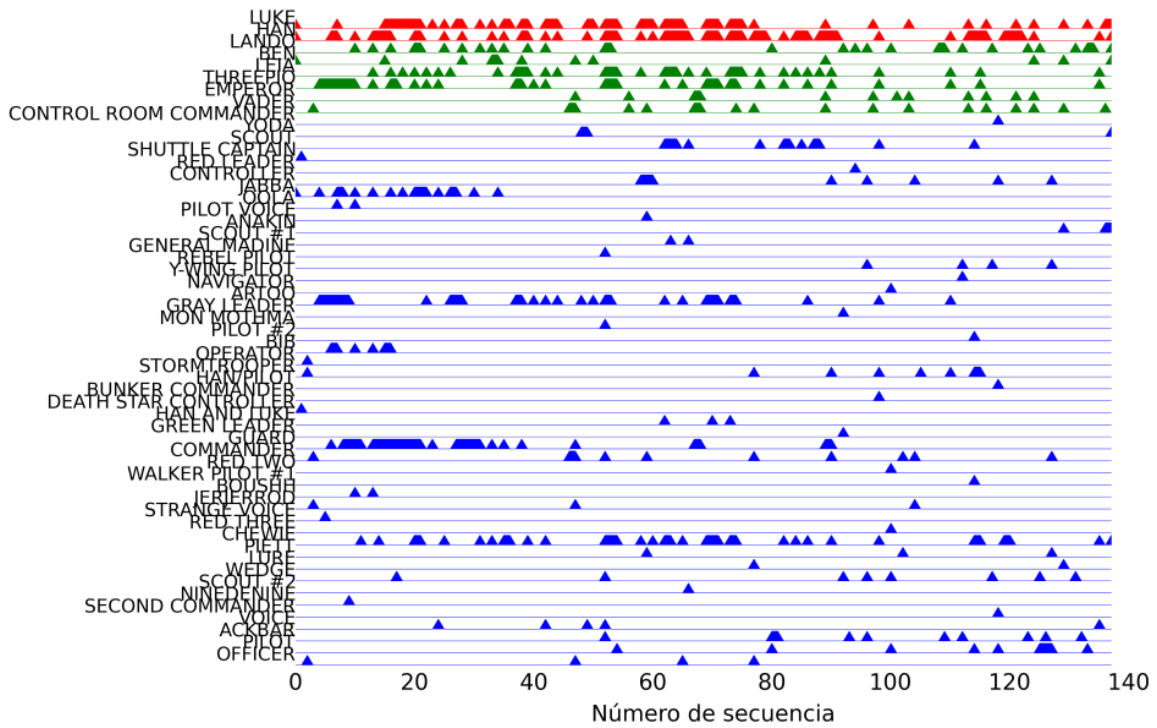


Figura K.1.3: Aparición de los personajes en “Star Wars: Episodio VI – El Retorno del Jedi ”

L

El laberinto del fauno

L.1 Personajes

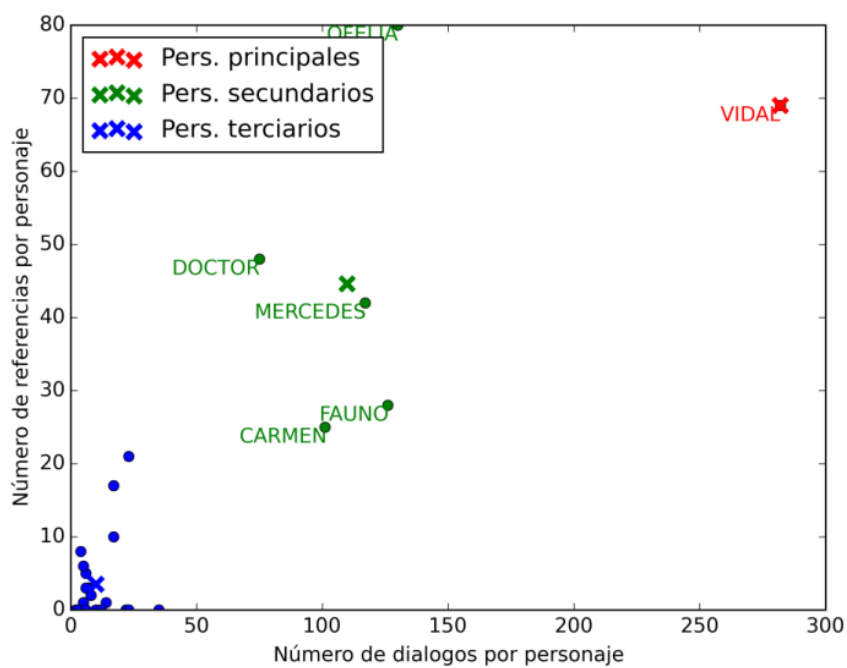


Figura L.1.1: Protagonismo de los personajes en "El laberinto del fauno"

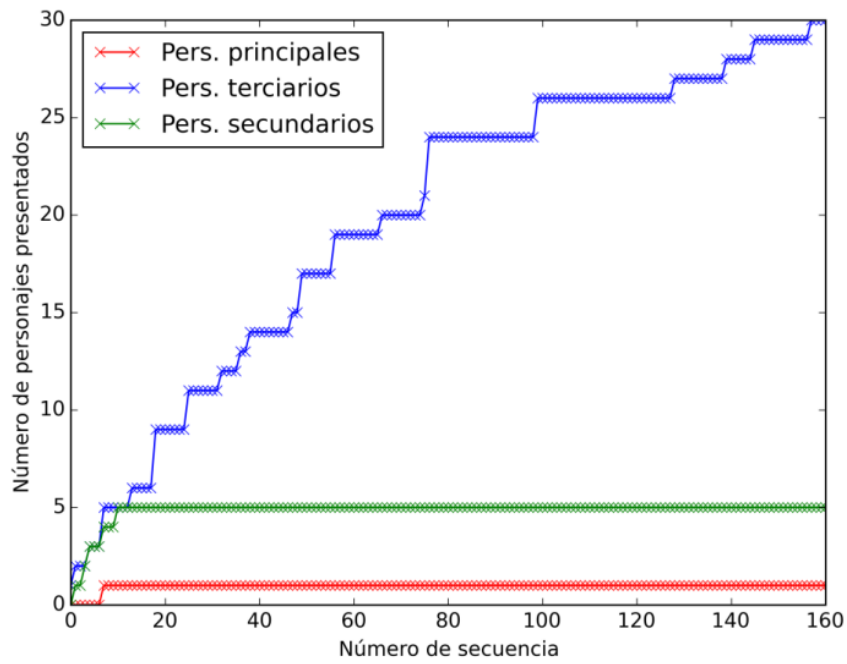


Figura L.1.2: Presentación de los personajes en “ El laberinto del fauno ”

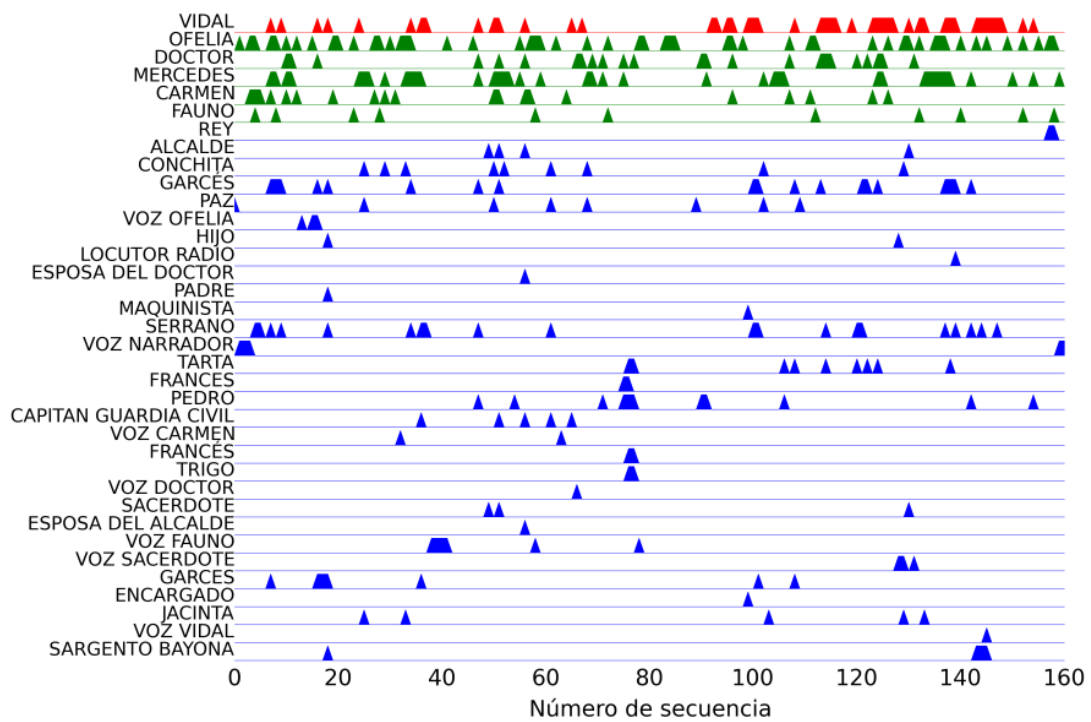


Figura L.1.3: Aparición de los personajes en “ El laberinto del fauno ”

L.2 Temas

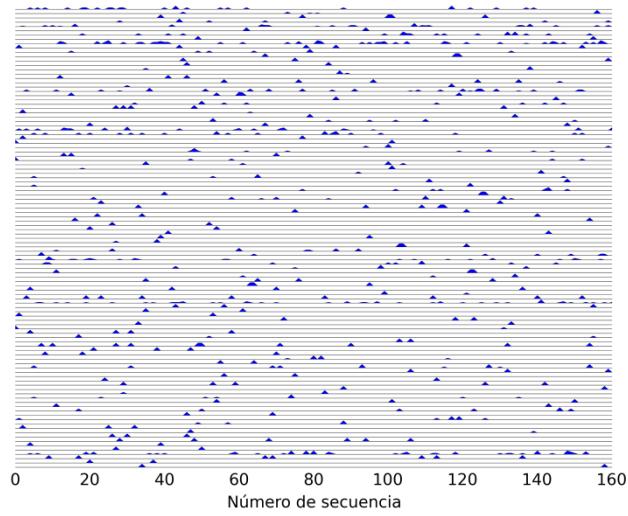


Figura LM.2: Distribución de los temas en “El laberinto del fauno”

L.3 Entropía

L.3.1 Entropía temas

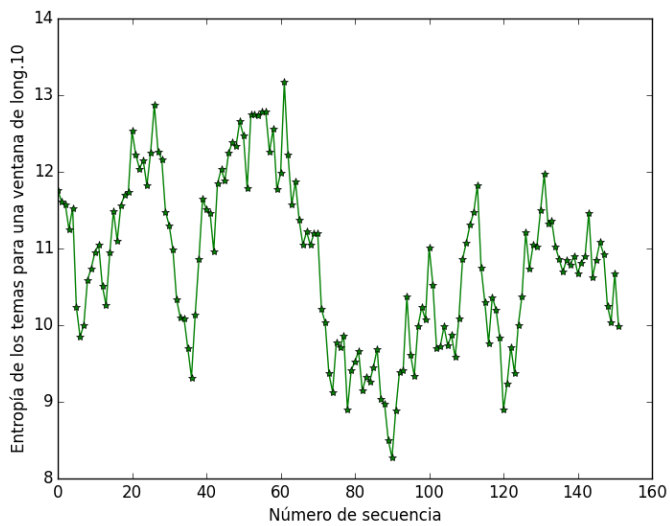


Figura L.3.1.1: Entropía de los temas en “El laberinto del fauno”

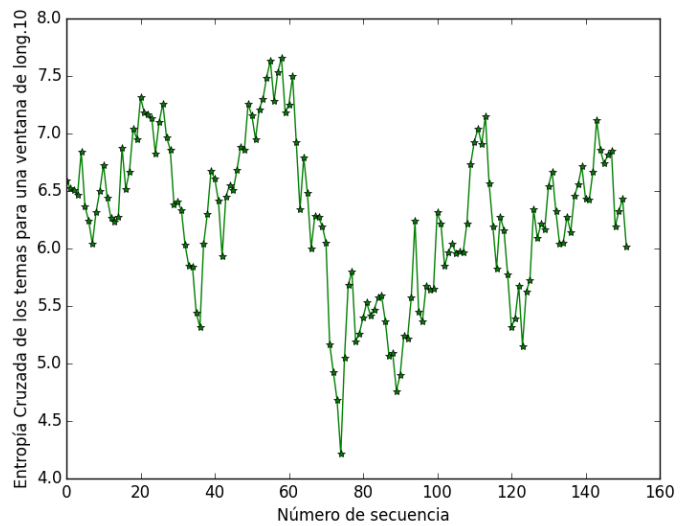


Figura L.3.1.2: Entropía cruzada de los temas en “El laberinto del fauno”

M.3.2 Entropía personajes

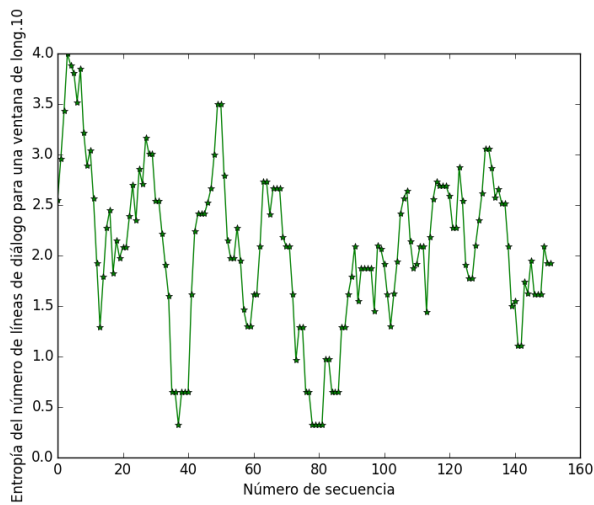


Figura L.3.2.1: Entropía del número de líneas de diálogo en “El laberinto del fauno”

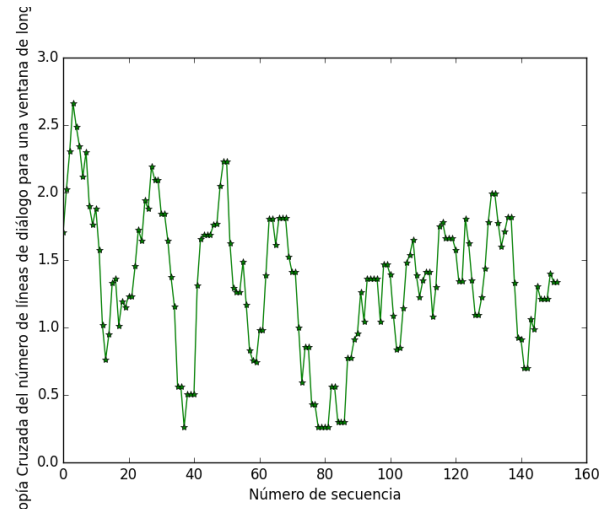


Figura L.3.2.2: Entropía cruzada del número de líneas de diálogo en “El laberinto del fauno”

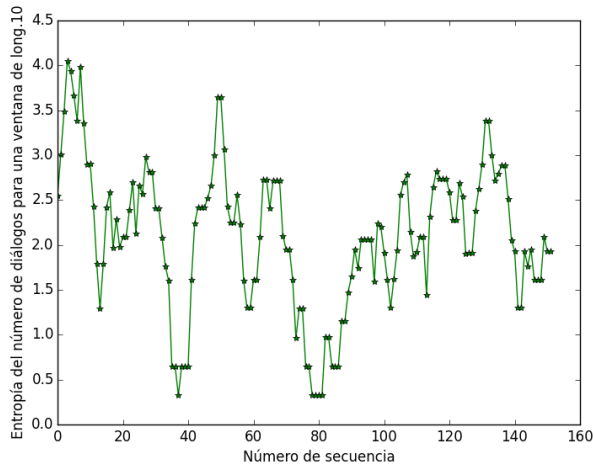


Figura L.3.2.3: Entropía del número de diálogos en “El laberinto del fauno”

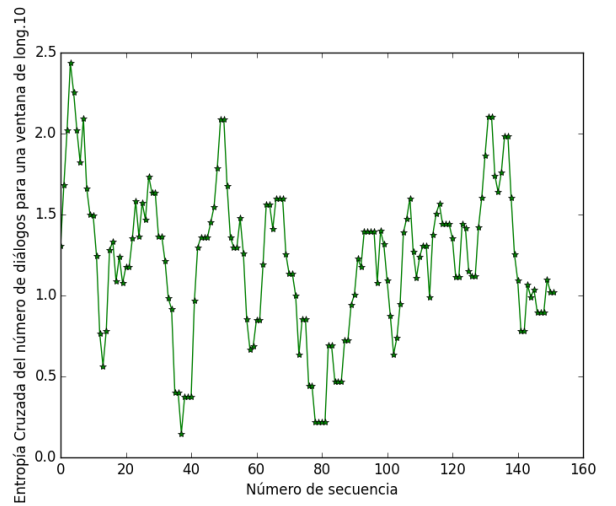


Figura L.3.2.4: Entropía cruzada del número de diálogos en “El laberinto del fauno”

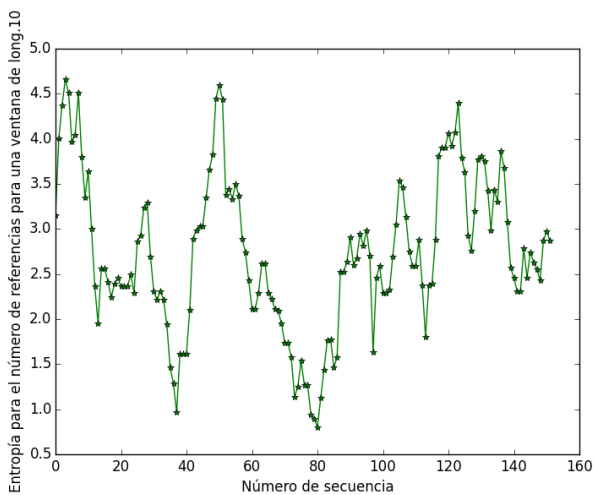


Figura L.3.2.5: Entropía del número de referencias en “El laberinto del fauno”

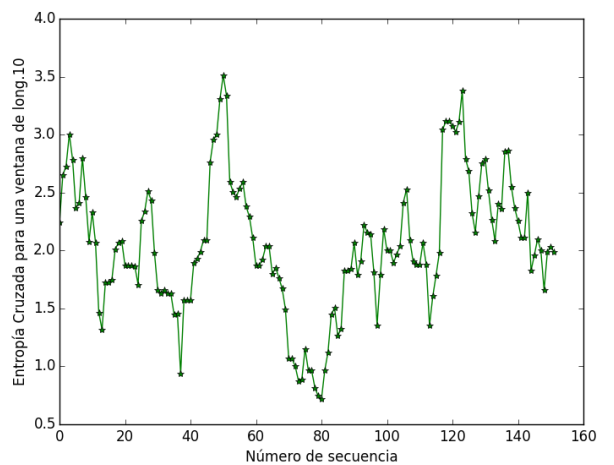


Figura L.3.2.6: Entropía del número de referencias en “El laberinto del fauno”

L.4 Grafos

Se presentan a continuación tres grafos resultados de tres ejecuciones diferentes del programa. Algunos constan de menos o más elementos que otros, pero son en cualquier caso representativos de los resultados, no son casos extremos.

Los nodos con letras corresponden a personajes, mientras que los marcados con letras corresponden a temas. Los nodos unidos mediante aristas de trazo continuo han obtenido un alto valor de correlación de Información Mutua –ver Apartado 3.6.3-. Los nodos pintados del mismo color son aquellos que han resultado agrupados según esta métrica. Cuanto mayor correlación más grueso el trazo de la arista. Por otra parte, los nodos unidos mediante trazos discontinuos comparten alguna de las ideas que conforman los temas.

Lo primero a destacar son los diferentes subgrafos que se repiten de una ejecución a otra. Entre ellos, el más importante es sin duda el centrado en el seleccionado como protagonista: Vidal. Se relaciona con “Juegos de azar”, y este tema a su vez con “Cuerpo de ser vivo”. Esta conexión parece hacer referencia a los momentos en los que tortura a otros dos personajes, aunque la expresión resulte sin duda un tanto frívola. Por otro lado, se relaciona fuertemente con el tema “Familia”. Sin duda un tema capital para el personaje, y que según el grafo, le pone en relación con el Doctor, quien cuida justamente de su esposa. Se le suele relacionar también con “Sentimiento de desagrado”.

Otro subgrafo que aparecer recurrentemente es el que une a Ofelia y el Fauno mediante el tema mitología. A Ofelia se la relaciona con “Vivienda de animales salvajes” y al Fauno con “Sentimiento de agrado”, al contrario que a Vidal.

Mercedes es relacionada con “Planta según su hábitat”. También se le suele relacionar con “Actividades y estados físicos – Voluntad”, acertada e interesante respuesta, y con “Fibra, hilo”.

Carmen, la madre, se relaciona con “Tecnología (física y química) aplicadas”. También con “Provincia – Provincial” y, en algunos otros casos, también con “Comportamiento – interés”, lo cual resulta también acertado e interesante como respuesta, describe bastante bien al personaje.

El Doctor, por su parte, aparece siempre relacionado con Vidal por el tema “Familia”, no en vano la relación que mantienen es a causa de la mujer enferma del capitán, Carmen.

En cuanto a los subgrafos temáticos, destacar “Ejército” y “Movilidad, inmovilidad”. Y también en algún caso puntual –por tanto no recogido en la selección que sigue a continuación- Vidal unido a “País (Estado)” por el tema “Ejército (armas)” y, también mediante este mismo tema, a “Ejército (guerra y paz)”, reseñables conexiones en una película ambientada en el contexto de la Guerra Civil.

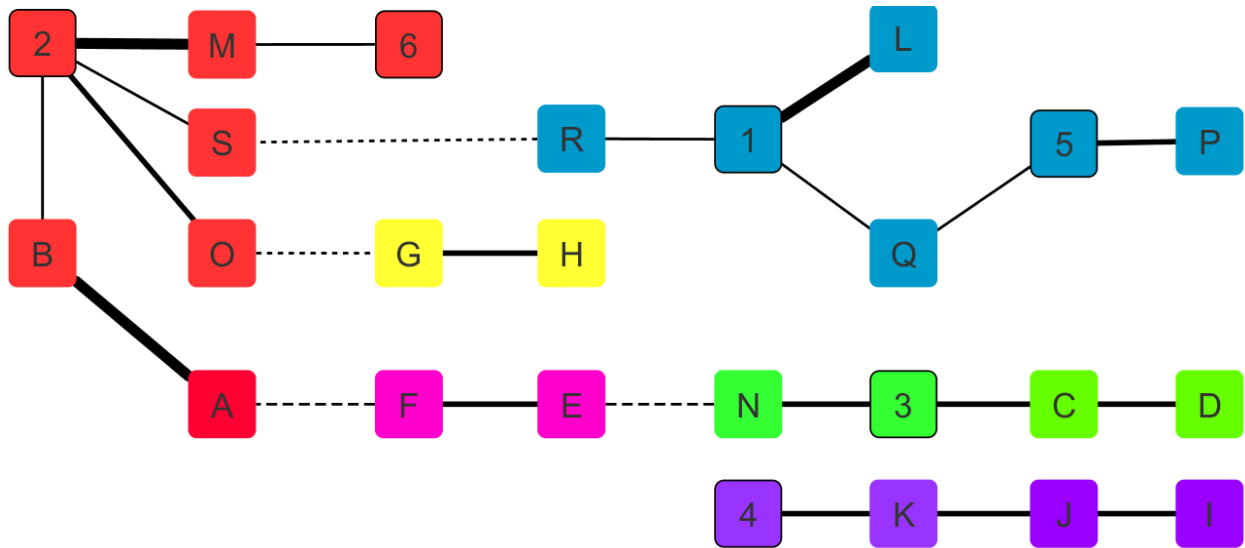


Figura L.4.1: Grafo de relaciones entre temas y personajes para “El laberinto del fauno”

- TEMAS -	
A	Cuerpo de ser vivo – Sistema nervioso
B	Juego de suerte – Jugar (juegos de azar)
C	Tecnología (Física y Química) – Mecánica
D	Provincia – Provincial
E	Comportamiento – Honrar
F	Cuerpo de ser vivo
G	Ejército (guerra y paz)
H	Movilidad, inmovilidad
I	Fibra, hilo
J	Actividades y estados físicos – Voluntad
K	Planta según hábitat – Planta perenne
L	Propiedad. Posesión – Poseer
M	Familia (en general) – Familia nuclear
N	Comportamiento – Interés
O	Ejército (armas)
P	Mitología – Mitología. Diosas
Q	Vivienda de animales salvajes – Vivienda de animales
R	Sentimientos (en general) - Sorpresa
S	Sentimientos (en general) – Desagrado
- PERSONAJES -	
1	FAUNO
2	VIDAL
3	CARMEN
4	MERCDES
5	OFELIA
6	DOCTOR
-ARISTAS-	
—————	Distribución relacionada (MI)
-----	Temática relacionada (DIC)

Tabla L.4.1: Leyenda de temas y personajes de la Figura L.4.1

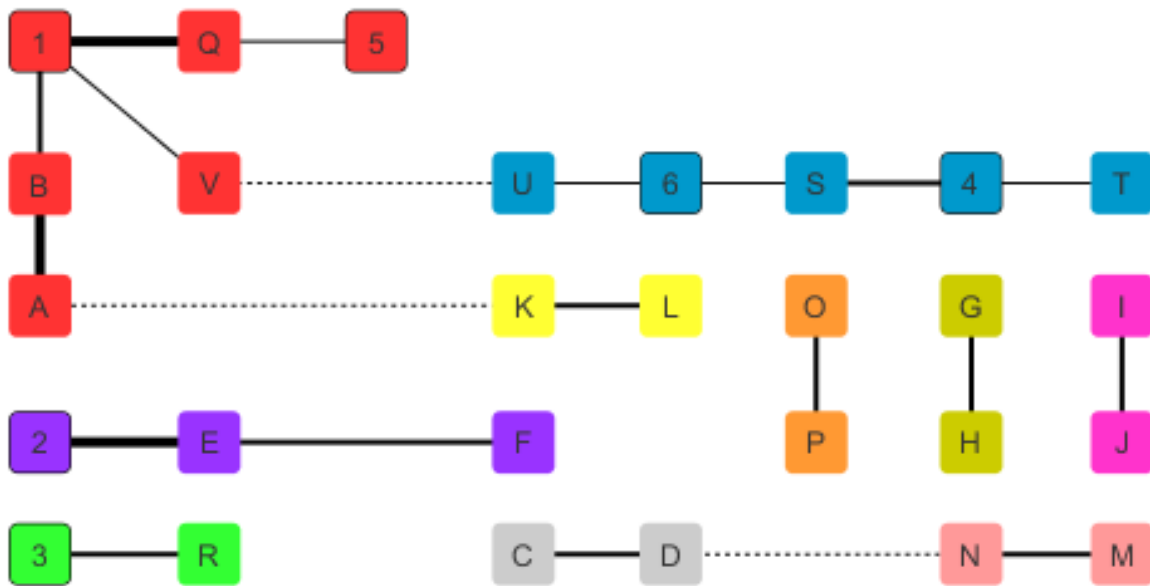


Figura L.4.2: Grafo de relaciones entre temas y personajes para “El laberinto del fauno”

- TEMAS -

- A Cuerpo de ser vivo – Sistema nervioso
- B Juego de suerte – Jugar (juegos de azar)
- C Fibra, hilo
- D Actividades y estados físicos – Voluntad
- E Tecnología (Física y Química) - Mecánica
- F Provincia - Provincial
- G Cantidad (número)
- H Tener, carecer
- I Movilidad, inmovilidad.
- J Ejército
- K Cuerpo de ser vivo
- L Comportamiento – Honrar
- M Construcción (en general)
- N Actividades y estados físicos
- O Artes plásticas
- P Desarrollo – Edad (de personas)
- Q Familia (en general) – Familia nuclear
- R Planta según su hábitat – Planta perenne
- S Mitología – Mitología. Diosas
- T Vivienda de animales salvajes – Vivienda de animales
- U Sentimientos (en general) – Sorpresa
- V Sentimientos (en general) – Desagrado

- PERSONAJES -

- 1 VIDAL
- 2 CARMEN
- 3 MERCEDES
- 4 OFELIA
- 5 DOCTOR
- 6 FAUNO

-ARISTAS-

- Distribución relacionada (MI)
- - - - - Temática relacionada (DIC)

Tabla L.4.2: Leyenda de temas y personajes de la Figura L.4.2

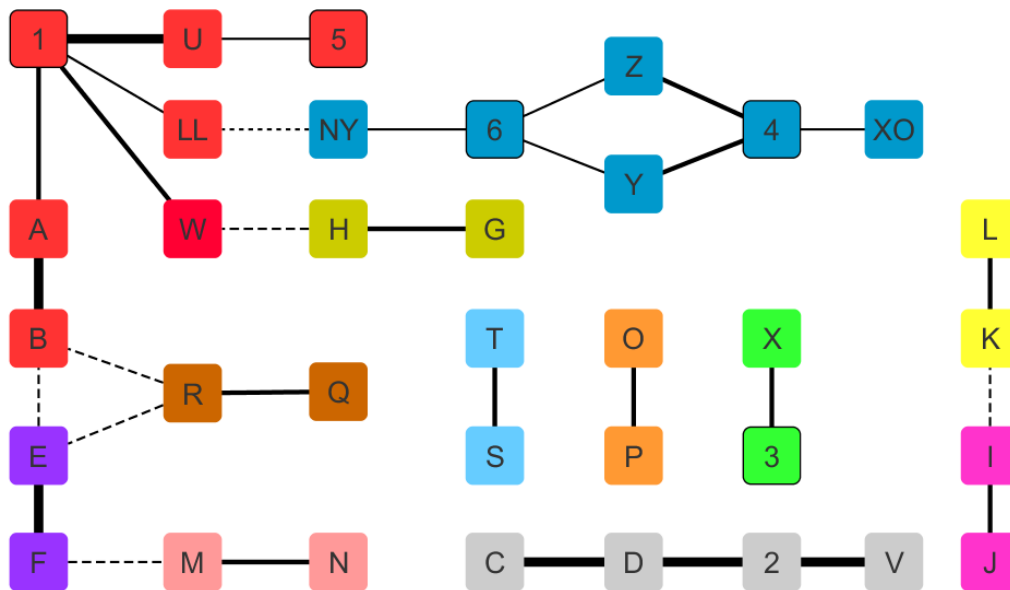


Figura L.4.3: Gráfico de relações entre temas e personagens para “El laberinto del fauno”
- TEMAS -

A	Juego de suerte – Jugar (juegos de azar)
B	Cuerpo de ser vivo – Sistema nervioso
C	Provincia – Provincial
D	Tecnología (Física y Química) – Mecánica
E	Cuerpo de ser vivo
F	Artes plásticas – Baile
G	Movilidad, inmovilidad
H	Ejército (guerra y paz)
I	Actividades y estados físicos – Voluntad
J	Fibra, hilo
K	Actividades y estados físicos
L	Construcción (en general)
LL	Sentimientos (en general) – Desagrado
M	Artes plásticas
N	Desarrollo – Edad (de personas)
NY	Sentimientos (en general) – Sorpresa
O	Enfermedades y accidentes
P	Energía (combustión)
Q	Fábrica – Preparar (disponer)
R	Cuerpo de ser vivo – Cara (rostro)
S	Ganadera - Granja
T	Política
U	Familia (en general) – Familia nuclear
V	Comportamiento – Interés
W	Ejército (armas)
X	Joyería y bisutería – Piedra preciosa
XO	Planta según hábitat – Planta herbácea
Y	Mitología – Mitología. Diosas
Z	Vivienda de animales salvajes – Vivienda de animales

- PERSONAJES -

1	VIDAL
2	CARMEN
3	MERCEDES
4	OFELIA
5	DOCTOR
6	FAUNO

Tabla L.4.3: Leyenda de temas y personajes de la Figura L.4.3

L.5 Escenas

ESCENA	PROB. DE CONFORMAR UNA ESCENA	DESCRIPCIÓN DE LA ESCENA
(0, 7)	0.5	Tras la parada en el bosque, vuelven al coche.
(8, 8)	1.0	De vuelta en el coche.
(9, 10)	1.0	El insecto observa el coche y llega al molino.
(11, 11)	1.0	Ofelia sigue al insecto al laberinto y conoce a Mercedes.
(12, 14)	0.75	Presentación del Doctor y del interior de la casa. Madre e hija en la habitación.
(15, 16)	1.0	Cuento de Ofelia.
(17, 21)	0.5	Final del cuento ya en la habitación. Médico y Vidal. Vidal mata al hijo de un campesino.
(22, 27)	0.5	Ofelia sigue al hada hasta el fauno y habla con él. Vidal en la cocina.
(28, 42)	0.225	Mercedes cocina. Baño. Vestido. Marca de la princesa. Ordeñando las vacas. Bodega. Guardias al monte. Ofelia frente al árbol.
(43, 49)	0.33	Ofelia debajo del árbol. Guardias en el claro del monte, los guerrilleros escondidos.
(50, 55)	0.75	Ofelia sale del árbol. Comida con las autoridades. Mercedes en la cocina. Sale de la casa.
(56, 57)	1.0	Guerrilleros responden a Mercedes. Ofelia llega del bosque y se encuentra con Mercedes.
(58, 58)	1.0	La madre se ausenta de la comida.
(59, 60)	1.0	Ofelia bañándose. Ofelia en el pozo con el fauno.
(61, 64)	0.5	Mercedes sale del molino y Ofelia vuelve. Vidal afeitándose. Racionamiento de comida.
(65, 70)	0.5	Ofelia lee, madre ensangrentada. Avisan a Vidal. Ofelia debe dormir sola. Mercedes canta nana.
(71, 72)	1.0	Doctor sorprende a Mercedes en la cocina. Ambos por el bosque.
(73, 80)	0.416	Se encuentran con guerrilleros. Visita del fauno mandrágora. Refugio y amputación. Puerta con tiza.
(81, 82)	1.0	Ofelia baja al pasillo. El reloj corre.
(83, 96)	0.225	Hombre pálido. Guerrilleros. Vidal afeitándose. Ofelia prepara mandrágora.
(97, 99)	0.75	Ofelia coloca la mandrágora. Vidal y el Doctor. Vidal sale al jardín y ve una columna de humo.
(100, 101)	1.0	Ofelia habla con su hermano. Vidal llega al lugar del siniestro.
(102, 117)	0.26	Al molino y explosión. Prisionero. Doctor y Carmen. Tortura. Fauno enfadado. Vidal descubre al Doctor.
(118, 123)	0.4166	Vidal ve la mandrágora. Doctor pone inyección. Guardias entran al palomar.
(124, 160)	0.3055	La mandrágora al fuego. Vidal mata al Doctor. Muere Carmen, la madre. Vidal y Ofelia. Vidal y Mercedes. Mercedes se va con Ofelia y son capturados. Torturan a Mercedes. Vuelve el fauno. Los guerrilleros salvan a Mercedes. Batalla. Ofelia llega al círculo. Vidal mata a Ofelia. Los guerrilleros matan a Vidal. Salón del trono.

Tabla L.5: Escenas de "El laberinto del fauno". Secuencias, probabilidad y descripción

L.6 Actos

L.6.1 Cambio Acto I - II

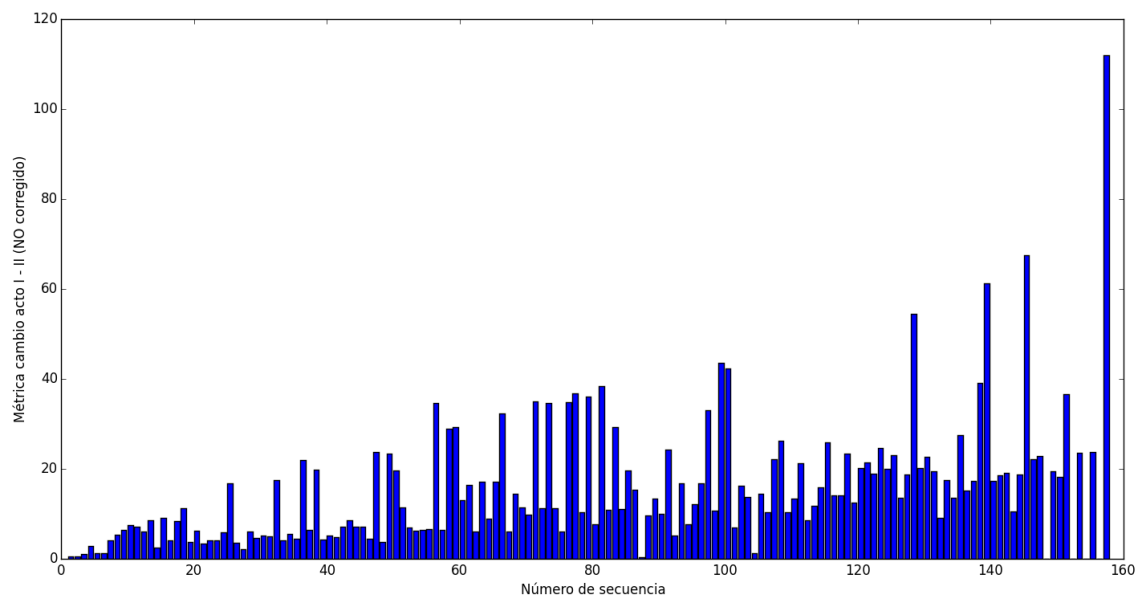


Figura L.6.1.1: Probabilidad de cambio del I al II Acto sin corrección teórica para "El laberinto del fauno"

Cambio de I a II ACTO en la secuencia 157, con valor 112.0

Líneas 4883 a 4892.

Cambio teórico en la secuencia 22 +/- 11.

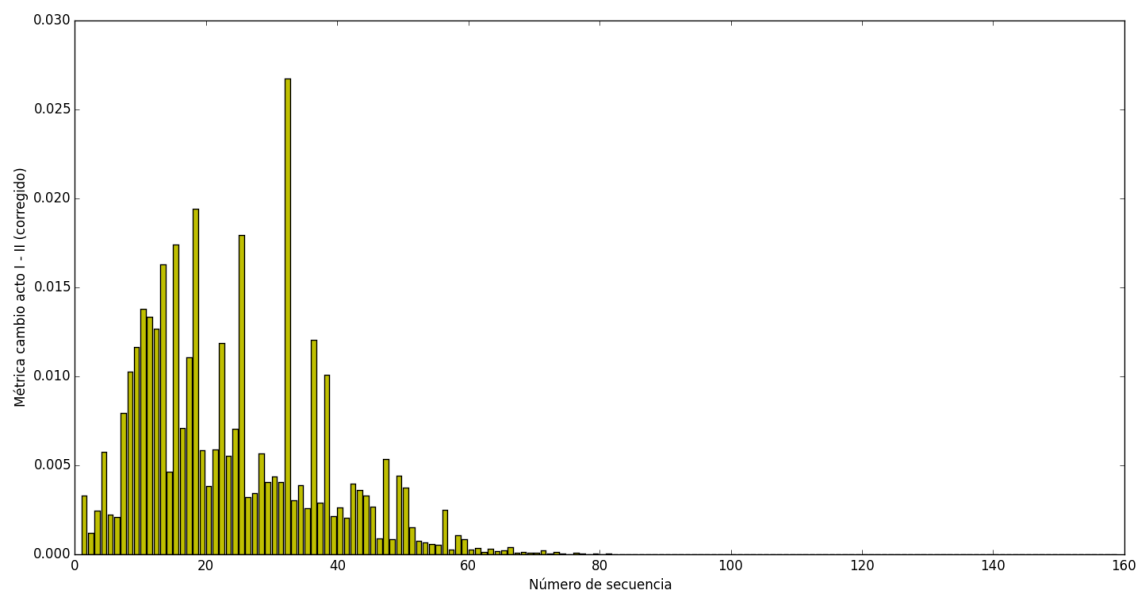


Figura L.6.1.2: Probabilidad de cambio del I al II Acto con corrección teórica para "El laberinto del fauno"

Cambio de I a II ACTO en la secuencia 32, con valor 0.02671.

Líneas 1337 a 1348.

Cambio teórico en la secuencia 22 +/- 11.

L.6.2 Cambio Acto II - III

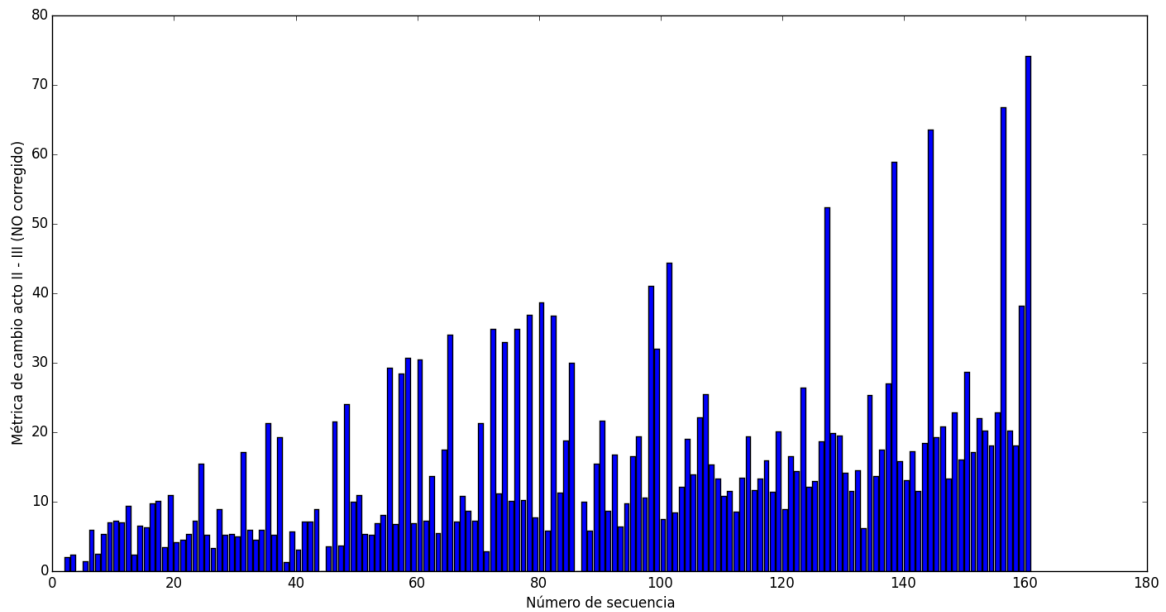


Figura L.6.2.1: Probabilidad de cambio del II al III Acto sin corrección teórica para “El laberinto del fauno”

Cambio de II a III ACTO en la secuencia 160, con valor 74.1562.

Líneas 4947 a 4961.

Cambio teórico en la secuencia 95 +/- 5.

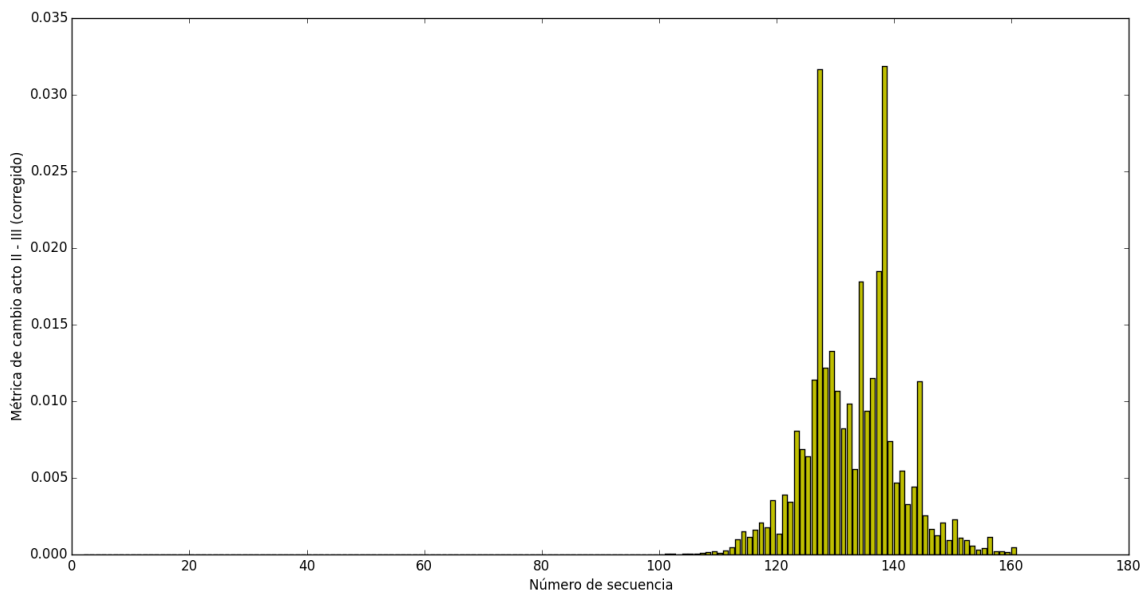


Figura L.6.2.2: Probabilidad de cambio del II al III Acto con corrección teórica para “El laberinto del fauno”

Cambio de II a III ACTO en la secuencia 138, con valor 0.03186.

Líneas 4328 a 4397.

Cambio teórico en la secuencia 95 +/- 5.

M

Tesis

M.1 Personajes

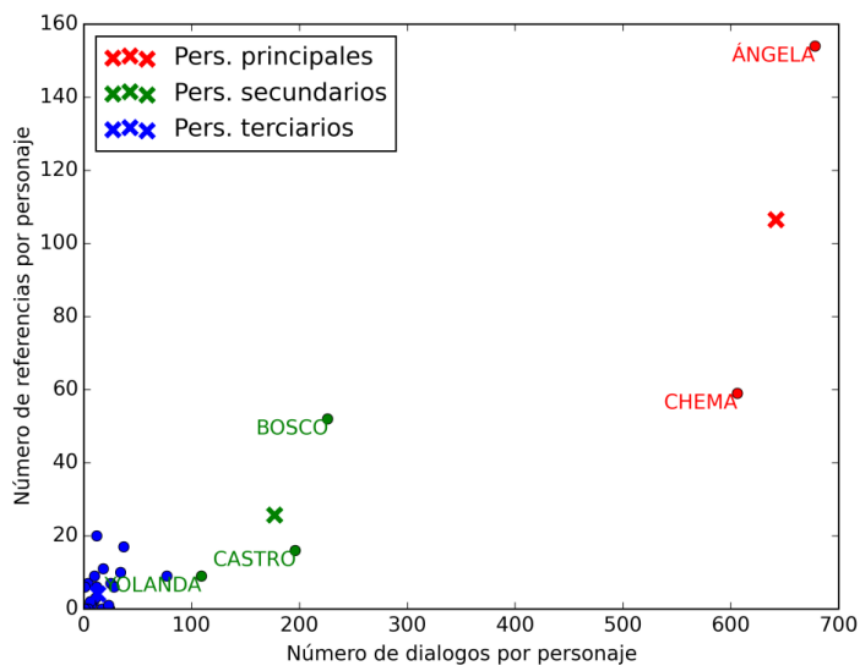


Figura M.1.1: Protagonismo de los personajes en “Tesis”

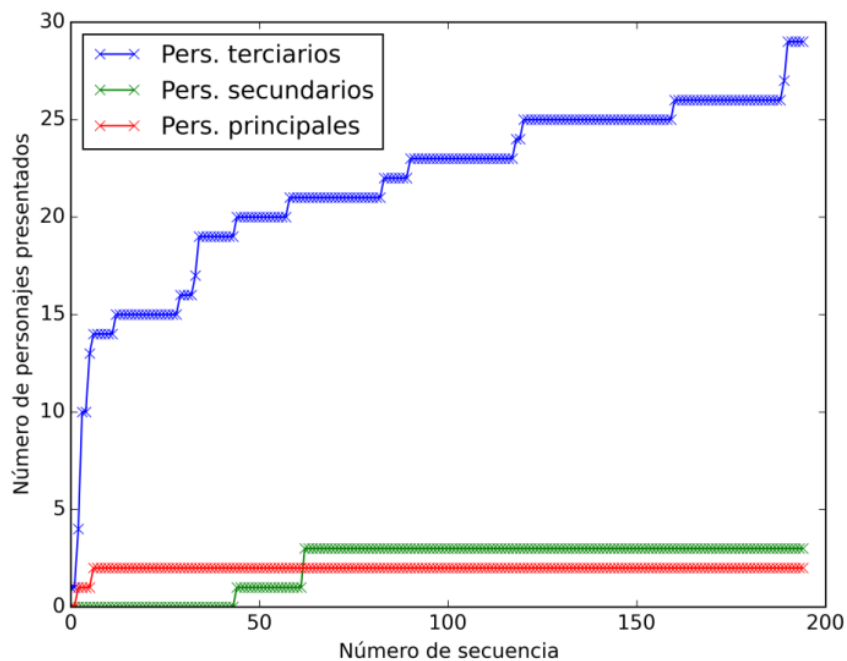


Figura M.1.2: Presentación de los personajes en “Tesis”

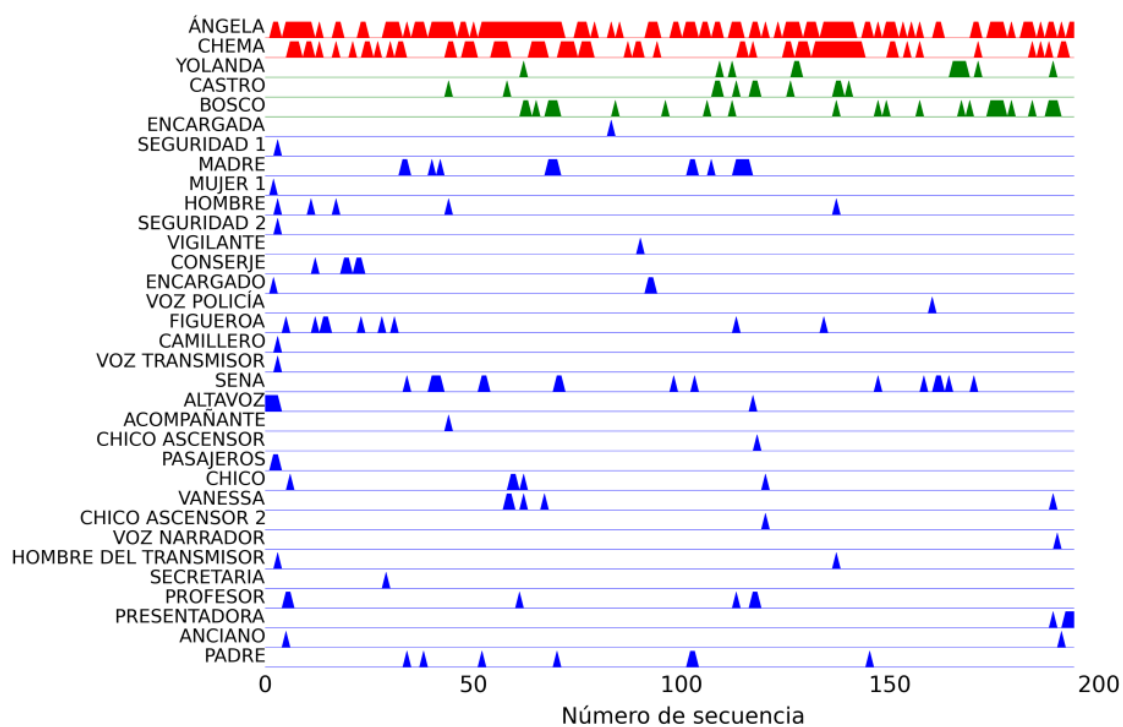


Figura M.1.3: Aparición de los personajes en “Tesis”

M.2 Temas

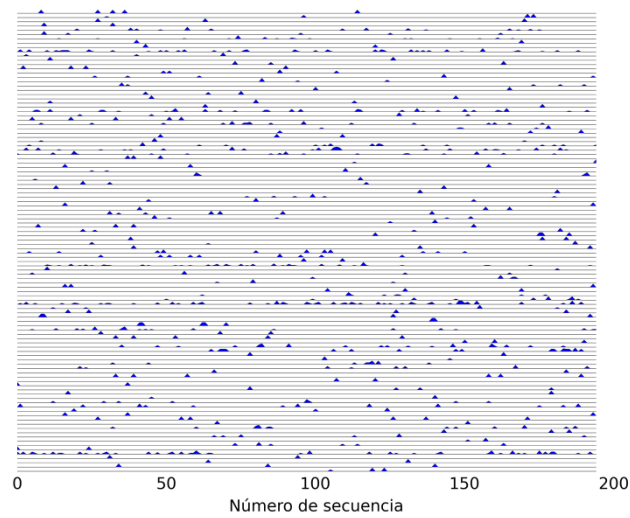


Figura M.2: Distribución de los temas en "Tesis"

M.3 Entropía

M.3.1 Entropía temas

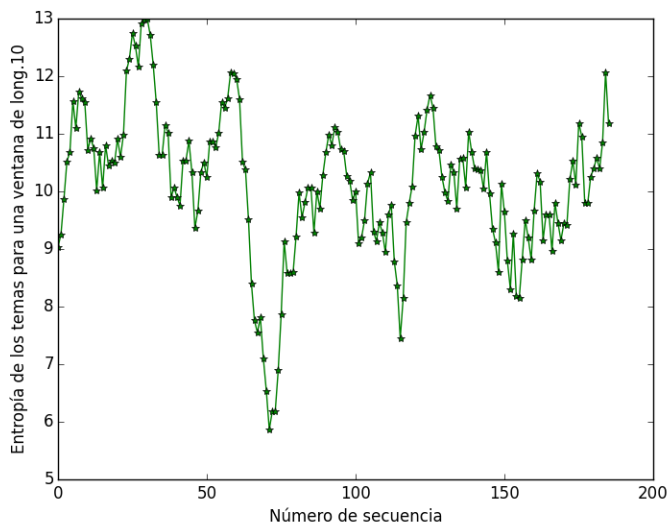


Figura M.3.1.1: Entropía de los temas en "Tesis"

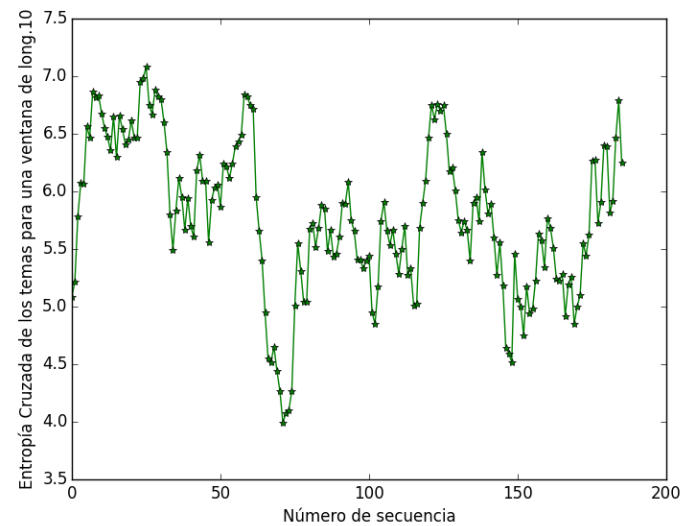


Figura M.3.1.2: Entropía cruzada de los temas en "Tesis"

M.3.2 Entropía personajes

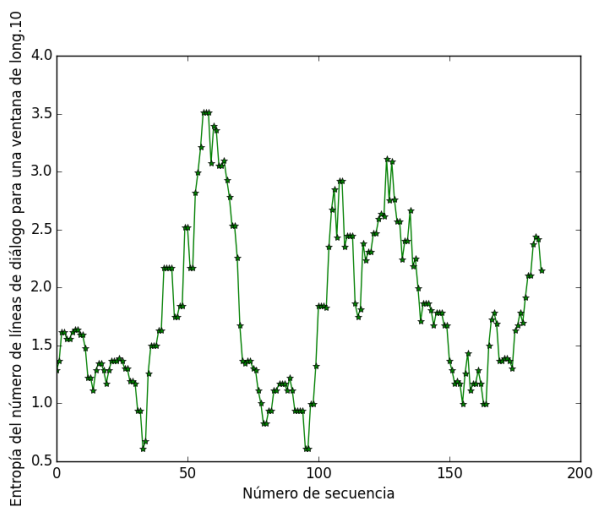


Figura M.3.2.1: Entropía del número de líneas de diálogo en “Tesis”

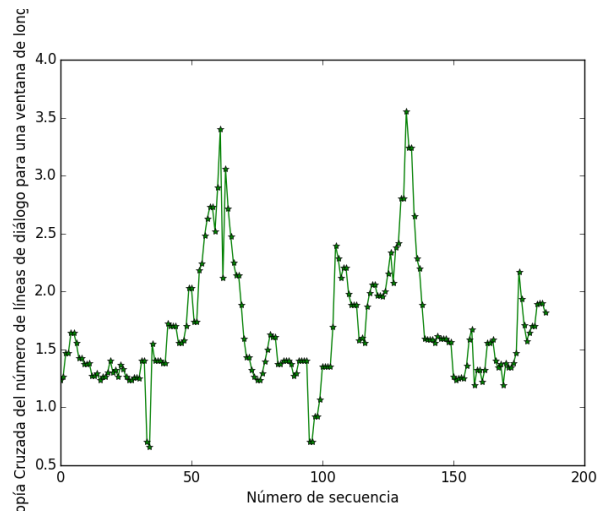


Figura M.3.2.2: Entropía cruzada del número de líneas de diálogo en “Tesis”

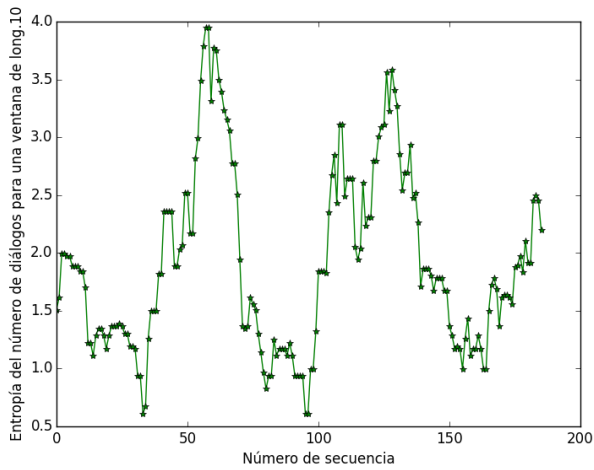


Figura M.3.2.3: Entropía del número de diálogos en “Tesis”

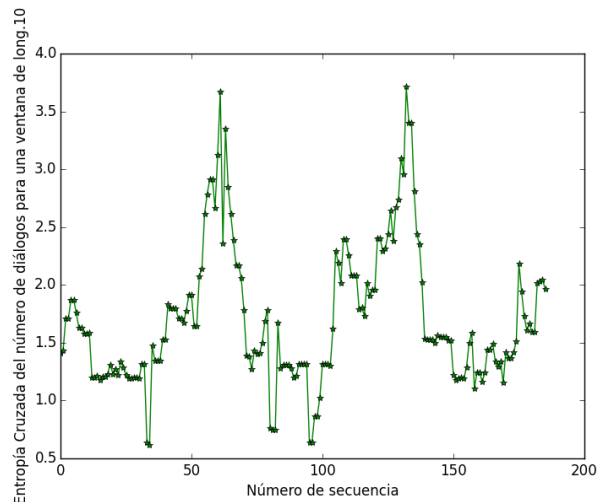


Figura M.3.2.4: Entropía cruzada del número de diálogos en “Tesis”

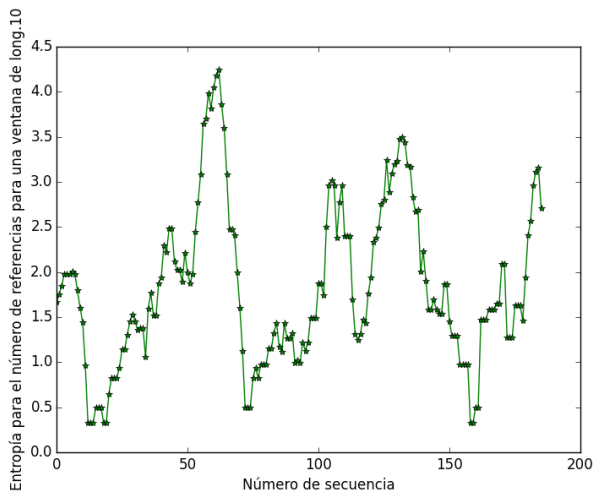


Figura M.3.2.5: Entropía del número de referencias en “Tesis”

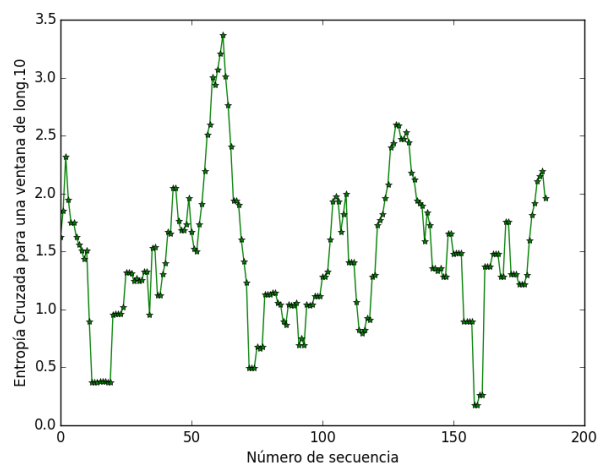


Figura M.3.2.6: Entropía del número de referencias en “Tesis”

M.4 Grafos

Se presentan a continuación tres grafos resultados de tres ejecuciones diferentes del programa. Algunos constan de menos o más elementos que otros, pero son en cualquier caso representativos de los resultados, no son casos extremos.

Los nodos con letras corresponden a personajes, mientras que los marcados con letras corresponden a temas. Los nodos unidos mediante aristas de trazo continuo han obtenido un alto valor de correlación de Información Mutua –ver Apartado 3.6.3-. Los nodos pintados del mismo color son aquellos que han resultado agrupados según esta métrica. Cuanto mayor correlación más grueso el trazo de la arista. Por otra parte, los nodos unidos mediante trazos discontinuos comparten alguna de las ideas que conforman los temas.

Encontramos unos grafos bastante compactos. Por un lado un subgrafo principal en el que encontramos a los protagonistas Ángela y Chema. Éstos se encuentran bastante relacionados por 2, 3 o incluso 4 nodos. “Energía – Combustión”, “Cantidad (número) – La matemática”. “Emplead@s en el hogar”. En un caso se relacionan también por “Ser humano – Persona” y “Complemento (accesorios) – Bolsa (saco)”, que en los grafos en los que no son compartidos se relacionan solamente o con Ángela en el primer caso o con Chema en el segundo. Ángela se relaciona también, bastante acertadamente cabe decir, mediante “Pelea / Discusión / Agresión” con el profesor Castro. Éste se relaciona también con “Sexualidad” y “Cuerpo de ser vivo”, que en algún caso también comparte con Ángela.

El tema “Energía – Combustión” aparece por una escena que hay en la que van encendiendo cerillas. “Complemento (accesorios) – Bolsa (saco)” porque están con el tema de la bolsa de cámara preocupados todo el rato. Interesante que a los protagonistas se les relacione con “Ser humano”, parece humanizar un poco a los personajes -aunque es una lectura quizás demasiado poética, habría que ver qué palabras causan este tema-. La ya señalada relación entre Castro y Ángela por “Pelea / Discusión / Agresión” resulta bastante interesante, pues el primero está a punto de matarla.

Hay otro subgrafo a parte, que es el de Bosco. Relacionado con “Árboles”, “Comportamiento – Permitir”, y “Cantidad (número) – Geometría”. En un caso, relacionado también con “Artes plásticas – Fotografía”. La relación con los árboles es por una secuencia en la que están en un parque y la de fotografía porque se le relaciona constantemente con la cámara (de vídeo, eso sí). Los nodos que menos destacan a simple vista, pueden resultar más interesantes a la luz de las conexiones que brindan con el otro subgrafo, pues se conectan justamente con aquellos nodos que conectan a Ángela con el resto de los personajes.

A parte tenemos un subgrafo temático que relaciona “Inteligencia – Verdad” con “Enfermedad y accidentes”. Podría estar refiriéndose al final: hospital y verdad.

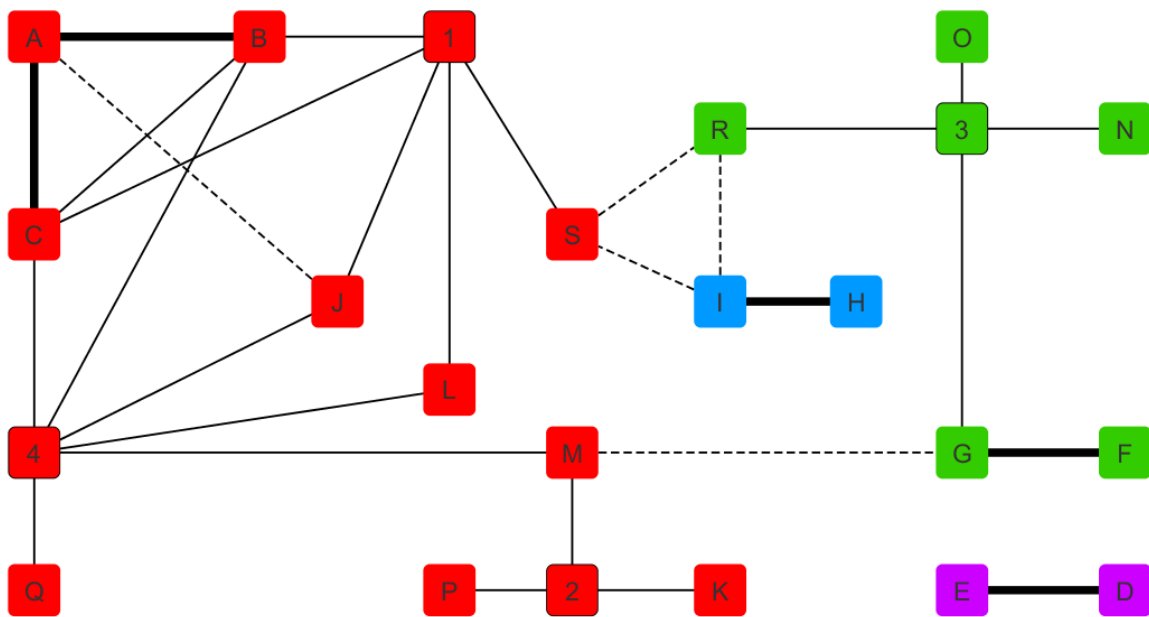


Figura M.4.1: Grafo de relaciones entre temas y personajes para “Tesis”

- TEMAS -

- A Física (ciencia)
- B Trabajadora (persona que trabaja) – Emplead@s en el hogar
- C Complemento (accesorios) – Bolsa (saco)
- D Inteligencia – Verdad
- E Enfermedades y accidentes – Enfermedad
- F Geografía – Meteorología
- G Comportamiento – Permitir
- H Teléfono, telégrafo
- I Cantidad (número) – Nada
- J Física (ciencia) – Luz / Oscuridad
- K Cuerpo de ser vivo – Nombre general de animales
- L Energía – Combustión
- M Comportamiento – Pelea / Discusión / Agresión
- N Artes plásticas – Fotografía (técnica)
- O Árboles – Conjunto de árboles
- P Reproducción (biología) – Sexualidad
- Q Ser humano – Persona (ser humano)
- R Cantidad (número) – Geometría
- S Cantidad (número) – La matemática

- PERSONAJES -

- 1 CHEMA
- 2 CASTRO
- 3 BOSCO
- 4 ÁNGELA

-ARISTAS-

- Distribución relacionada (MI)
- - - - - Temática relacionada (DIC)

Tabla M.4.1: Leyenda de temas y personajes de la Figura M.4.1

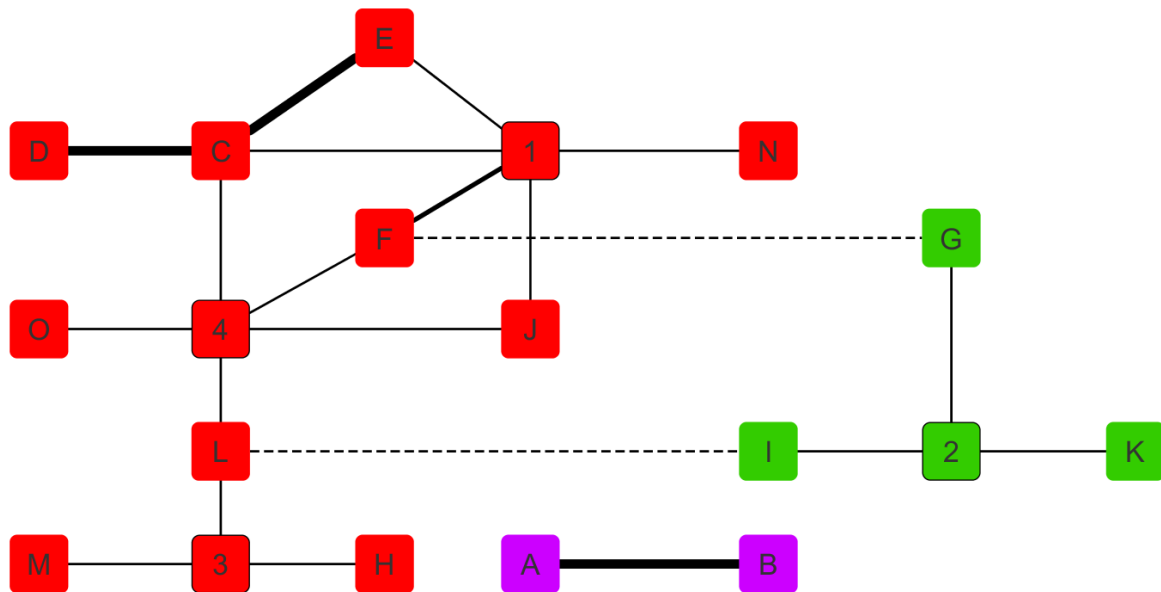


Figura M.4.2: Grafo de relaciones entre temas y personajes para “Tesis”

- TEMAS -	
A	Enfermedades y accidentes – Enfermedad
B	Inteligencia – Verdad
C	Trabajadora (persona que trabaja) – Emplead@s en el hogar
D	Física (ciencia)
E	Complemento (accesorios) – Bolsa (saco)
F	Cantidad (número) – La matemática
G	Cantidad (número) – Geometría
H	Cuerpo de ser vivo – Nombre general de animales
I	Comportamiento – Permitir
J	Energía – Combustión
K	Árboles – Conjunto de árboles
L	Comportamiento – Pelea / Discusión / Agresión
M	Reproducción (biología) – Sexualidad
N	Situación en general – Situación (sitio)
O	Ser humano – Persona (ser humano)
- PERSONAJES -	
1	CHEMA
2	BOSCO
3	CASTRO
4	ÁNGELA
-ARISTAS-	
—————	Distribución relacionada (MI)
-----	Temática relacionada (DIC)

Tabla M.4.1: Leyenda de temas y personajes de la Figura M.4.2

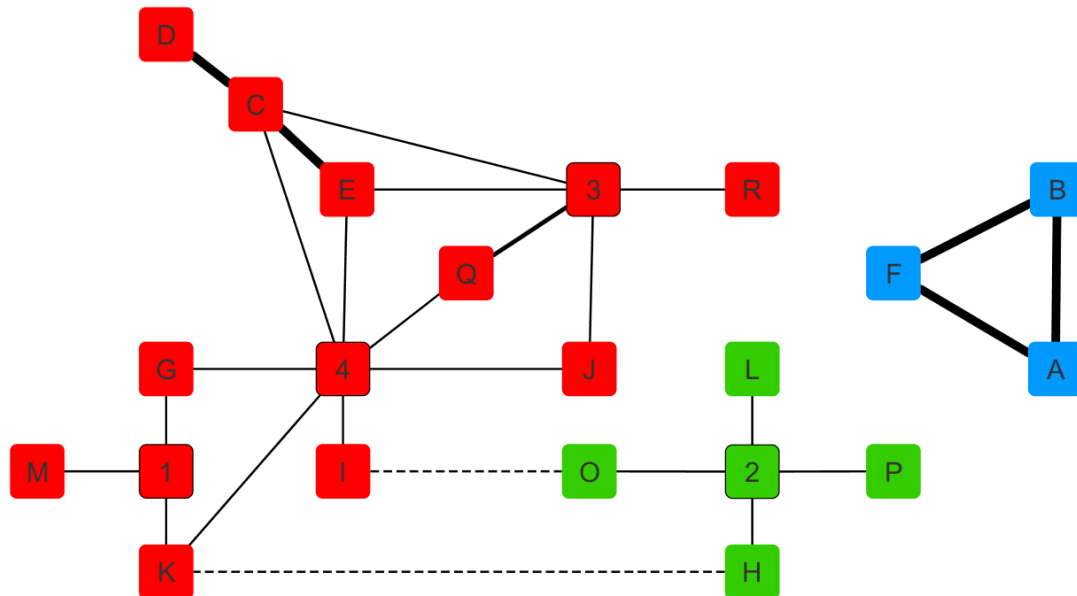


Figura M.4.3: Grafo de relaciones entre temas y personajes para “Tesis”

- TEMAS -

- A Inteligencia – Verdad
- B Enfermedades y accidentes – Enfermedad
- C Trabajadora (persona que trabaja) – Emplead@s en el hogar
- D Física (ciencia)
- E Complemento (accesorios) – Bolsa (saco)
- F Situación en general – Sacarse (extraerse)
- G Cuerpo de ser vivo – Nombre general de animales
- H Comportamiento – Permitir
- I Cantidad (número) – Parte de
- J Energía – Combustión
- K Comportamiento – Pelea / Discusión / Agresión
- L Artes plásticas – Fotografía (técnica)
- M Reproducción (biología) – Sexualidad
- N Sustancia química – Química inorgánica
- O Cantidad (número) – Geometría
- P Árboles – Conjunto de árboles
- Q Ser humano – Persona (ser humano)
- R Sentimientos (en general) - Afecto

- PERSONAJES -

- 1 CASTRO
- 2 BOSCO
- 3 CHEMA
- 4 ÁNGELA

-ARISTAS-

- Distribución relacionada (MI)
- - - - - Temática relacionada (DIC)

Tabla M.4.3: Leyenda de temas y personajes de la Figura M.4.3

M.5 Escenas

ESCENA	PROB. DE CONFORMAR UNA ESCENA	DESCRIPCIÓN DE LA ESCENA
(0, 5)	0.75	Metro.
(6, 7)	1.0	Ángela entra en la facultad y habla con el profesor Figueroa.
(8, 12)	0.416	Conoce a Chema y van a su casa a ver películas.
(13, 18)	0.416	Terminan de ver la película. Figueroa anda por el depósito y descubre una sala.
(19, 34)	0.25	Inspecciona la sala. Siguen viendo películas en casa de Chema. Ángela descubre a Figueroa muerto.
(35, 40)	0.33	Ángela llega a casa y pone la película de la sala.
(41, 42)	1.0	Descripción de lo que ella está viendo.
(43, 51)	0.66	Ángela se despierta. Discute con su hermana. Castro anuncia que Figueroa ha muerto. Ángela y Chema en casa de éste viendo películas.
(52, 53)	1.0	Ángela vomita. Hablan sobre la película. Descubren qué marca de cámara debió ser utilizada.
(54, 56)	0.75	Ángela llega a casa, discute con su hermana por si ha entrado en su cuarto.
(57, 63)	0.375	Conversación telefónica con Chema. Descubren que la muerta debió conocer a su asesino. Ángela ve la cámara en la cafetería. Persecución.
(64, 67)	0.5	Ángela conoce a Bosco y luego a Yolanda. Éste le deja la cámara. Ángela con Chema.
(68, 71)	0.5	Chema conoce a Bosco. Chema y Ángela hablando. Ángela en casa, Bosco está también ahí.
(72, 75)	0.75	En el cuarto. Cena. Viendo la televisión. Conversación con Chema sobre el tema.
(76, 80)	0.5	Ángela roba información de la tienda. Ángela llama a Chema.
(81, 90)	0.2916	Chema en la universidad con el vigilante. Ángela en el depósito. Ya en casa ve la entrevista a Bosco.
(91, 94)	0.5	Sena confiesa haber husmeado. Ángela se despierta.
(95, 96)	1.0	Baja hasta la cocina y cree ver a alguien. Cuchillo en mano y estalla de tensión.
(97, 99)	0.5	Sus padres le tranquilizan. Durmiendo en el cuarto.
(100, 103)	0.5	Pesadilla con Bosco. Llamada de Castro. Yolanda llama al timbre y Ángela no responde.
(104, 105)	1.0	Ángela ve a Yolanda marcharse por la ventana.
(106, 126)	0.1339	Conversación entre Ángela y Castro. Chema le alerta. Persecución por la Universidad. Ángela con Chema. Conversación con Yolanda. En los túneles.
(127, 128)	1.0	Ángela y Chema guiándose con las cerillas.
(129, 138)	0.2916	Ángela se despierta sola. Le tiene Castro en la sala de edición. Chema mata Castro. Van a casa de Chema. Ángela a su casa.
(139, 142)	0.75	Ángela va a la discoteca. Salen. Llegan al parque. Casa de Chema, Ángela quiere ir a la policía.
(143, 147)	1.0	Ángela sospecha de Chema.
(148, 194)	0.29545	Sena asustada. Ángela va a casa de Bosco. Hablan. Llega Chema. Bosco amordaza a Ángela. Ángela mata a Bosco. Hospital. Programa de televisión. Final.

Tabla M.5: Escenas de "Tesis". Secuencias, probabilidad y descripción

M.6 Actos

M.6.1 Cambio Acto I – II

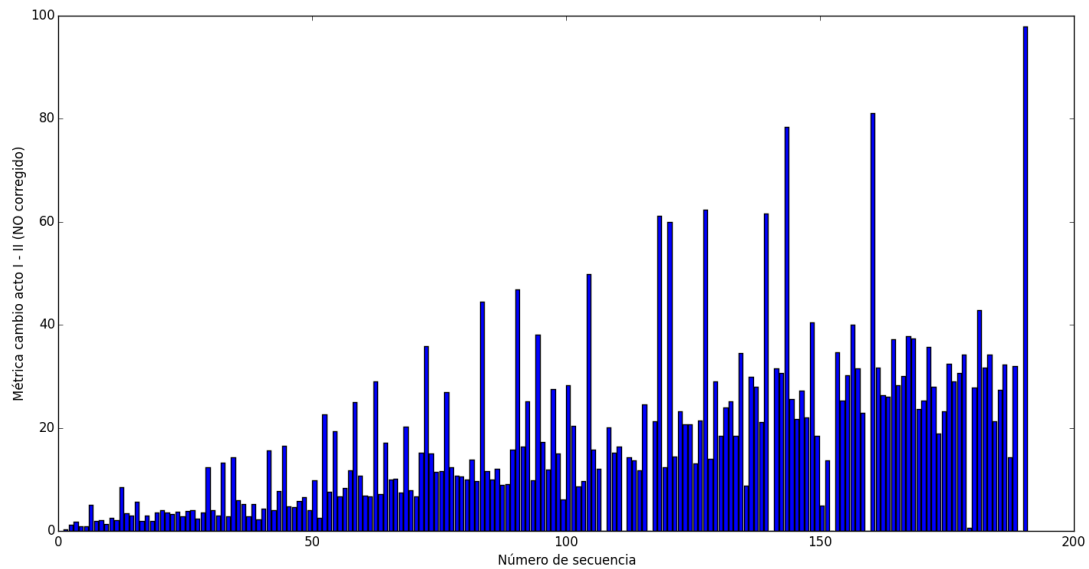


Figura M.6.1.1: Probabilidad de cambio del I al II Acto sin corrección teórica para "Tesis"

Cambio de I a II ACTO en la secuencia 190, con valor asignado 97.9176.

Líneas 6840 a 6874.

Cambio teórico en la secuencia 32 +/- 16

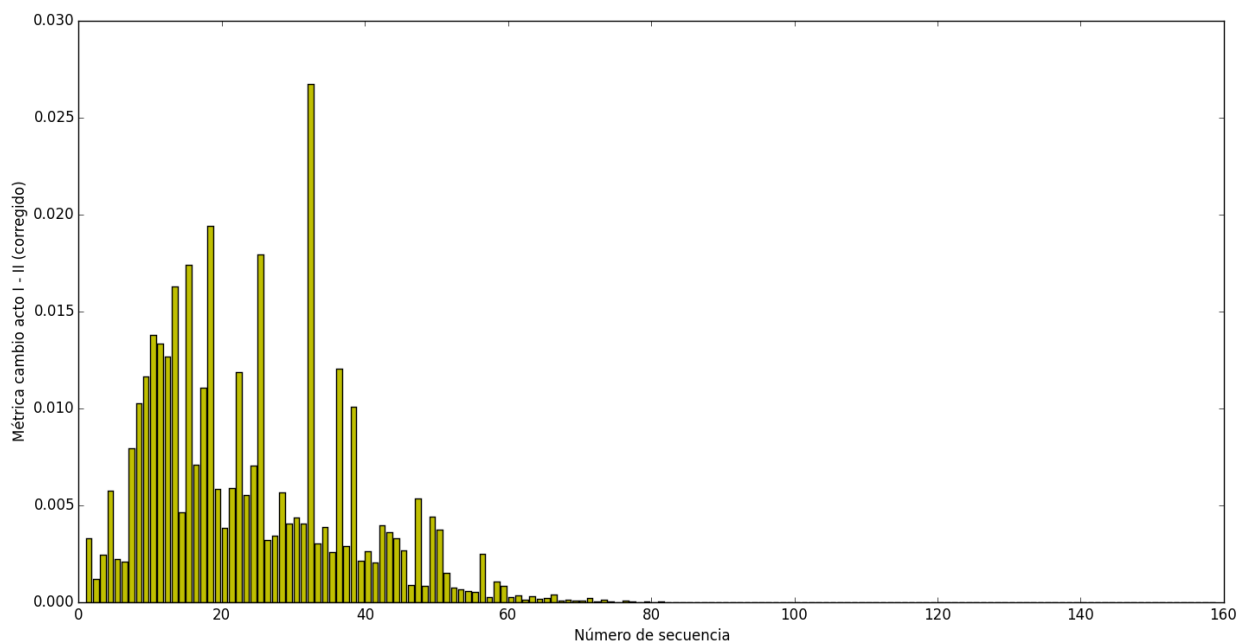


Figura M.6.1.1: Probabilidad de cambio del I al II Acto con corrección teórica para "Tesis"

Cambio de I a II ACTO en la secuencia 44, con valor asignado 0.01424.

Líneas 1163 a 1243.

Cambio teórico en la secuencia 32 +/- 16

M.6.2 Cambio Acto II – III

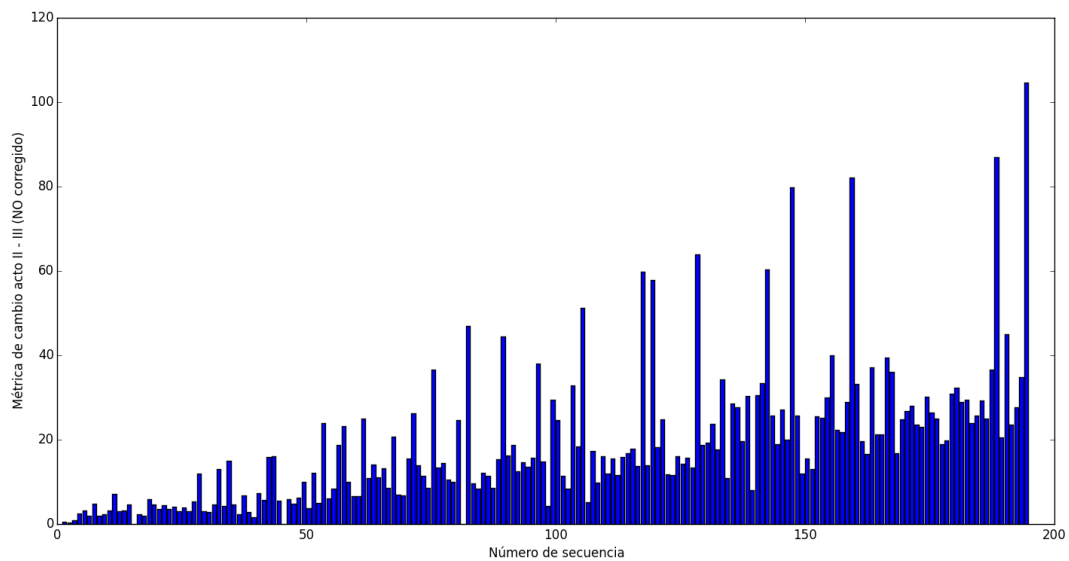


Figura M.6.2.1: Probabilidad de cambio del II al III Acto sin corrección teórica para "Tesis"

Cambio de II a III ACTO en la secuencia 194, con valor asignado 104.65.

Líneas 7065 a 7085.

Cambio teórico en la secuencia 136 +/- 8

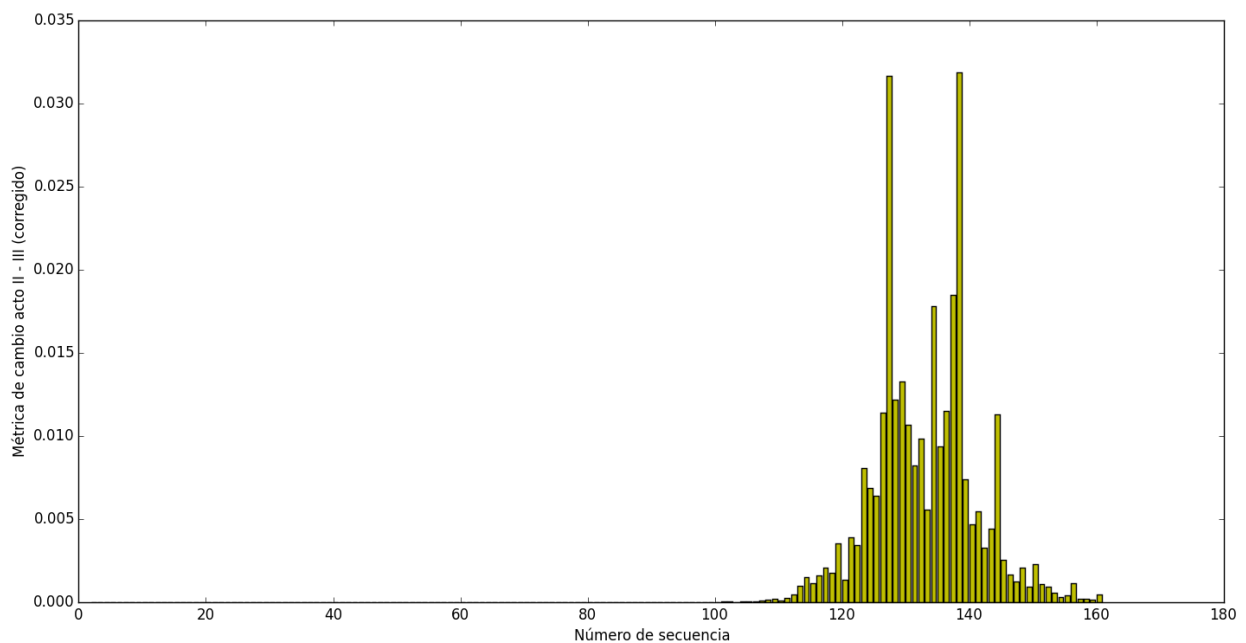


Figura M.6.2.2: Probabilidad de cambio del II al III Acto con corrección teórica para "Tesis"

Cambio de II a III ACTO en la secuencia 147, con valor asignado 0.02713.

Líneas 5841 a 5909.

Cambio teórico en la secuencia 136 +/- 8

N

La isla mínima

N.1 Personajes

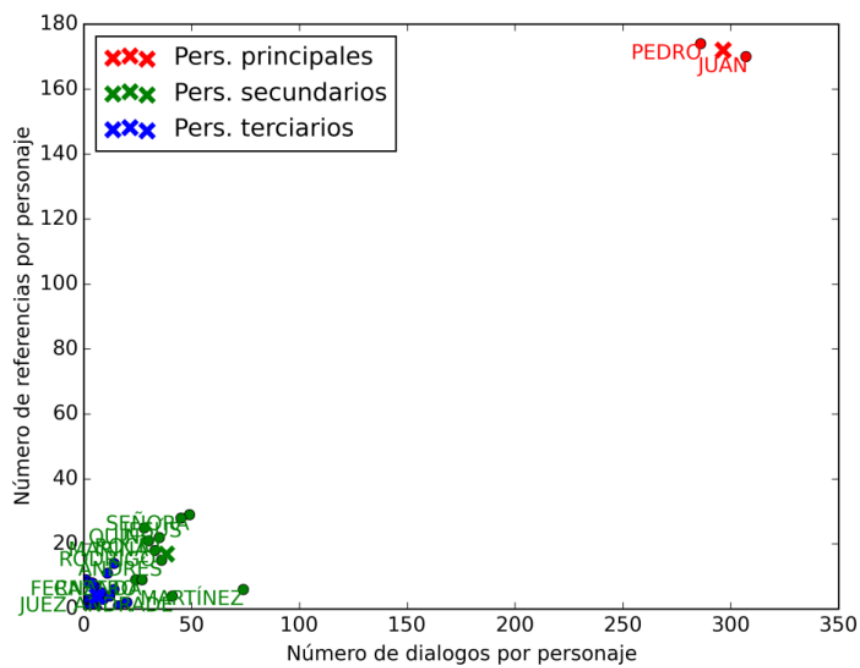


Figura N.1.1: Protagonismo de los personajes en “La isla mínima”.

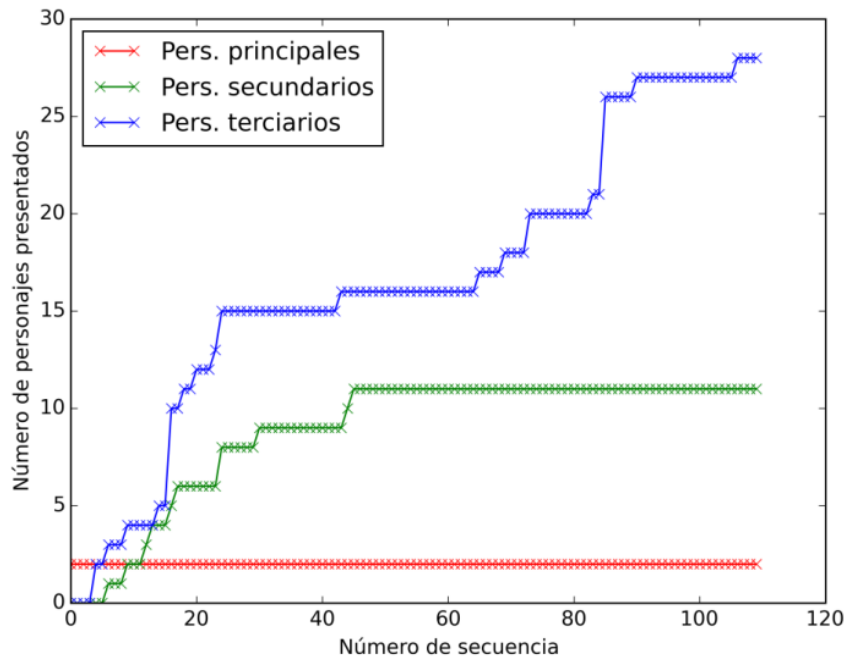


Figura N.1.2: Presentación de los personajes en “La isla mínima”.

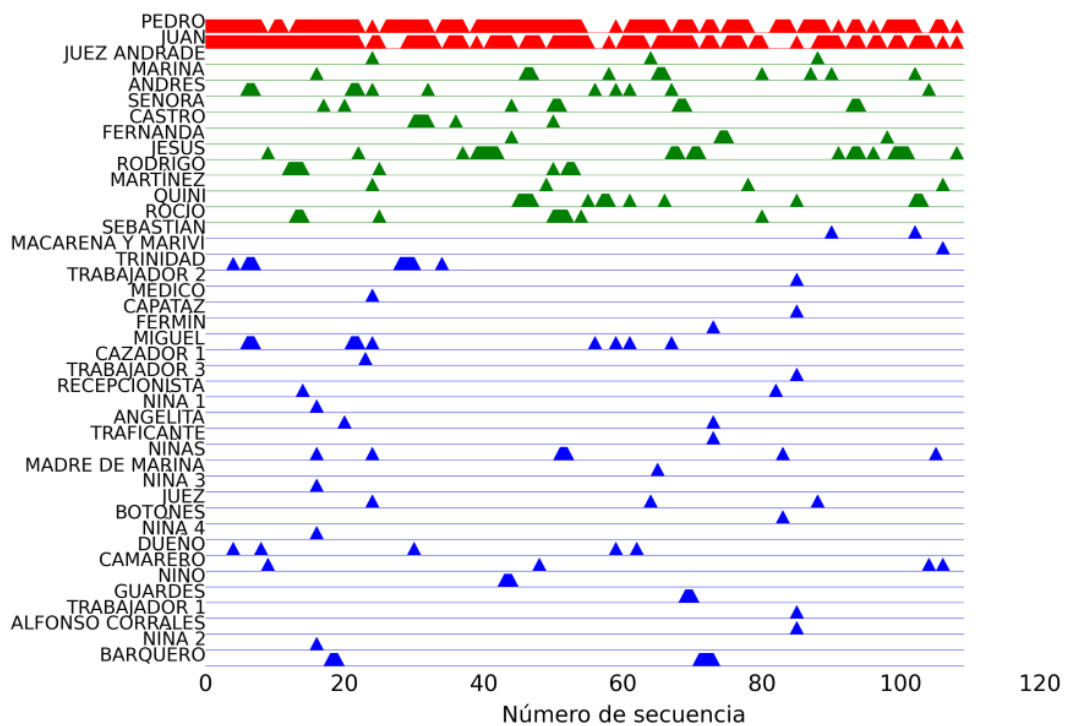


Figura N.1.3: Aparición de los personajes en “La isla mínima”.

N.2 Temas

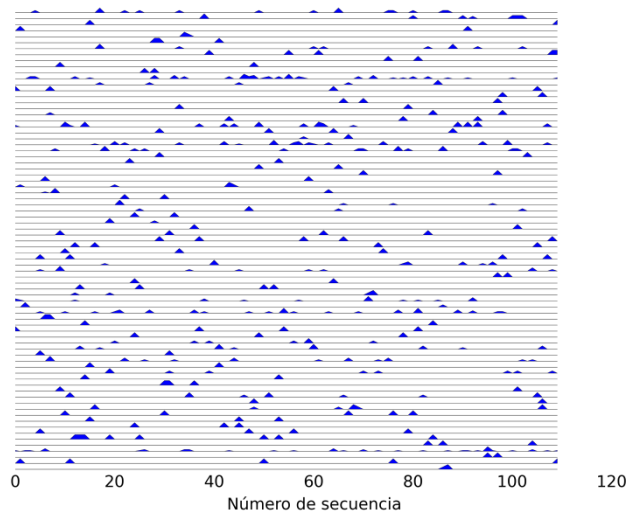


Figura N.2: Distribución de los temas en "La isla mínima".

N.3 Entropía

N.3.1 Entropía temas

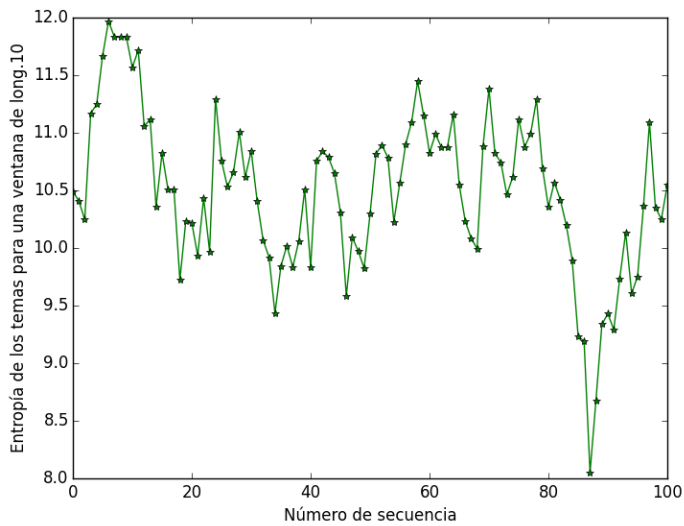


Figura N.3.1.1: Entropía de los temas en "Tesis"

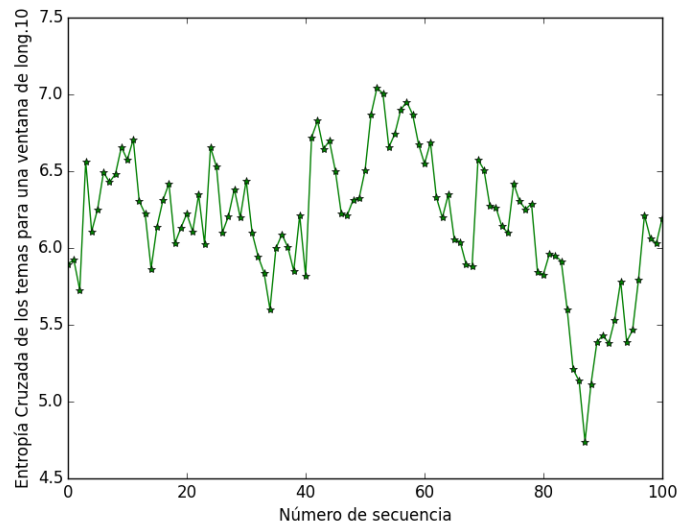


Figura N.3.1.2: Entropía cruzada de los temas en "Tesis"

N.3.2 Entropía personajes

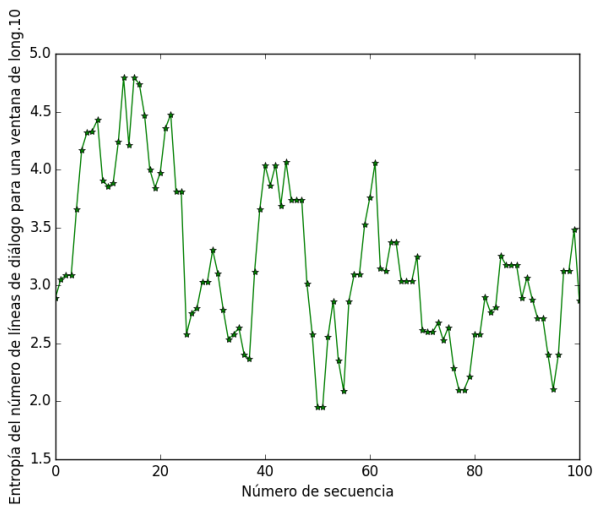


Figura N.3.2.1: Entropía del número de líneas de diálogo en “La isla mínima”

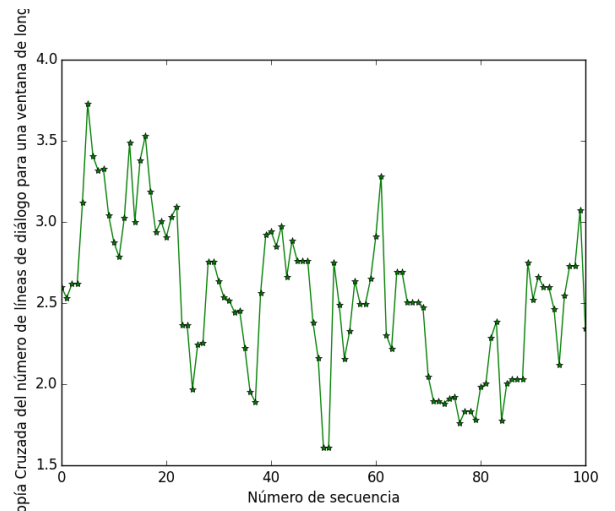


Figura N.3.2.2: Entropía cruzada del número de líneas de diálogo en “La isla mínima”

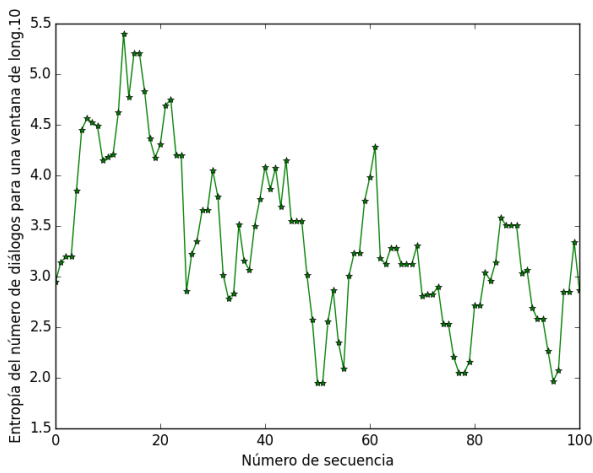


Figura N.3.2.3: Entropía del número de diálogos en “La isla mínima”

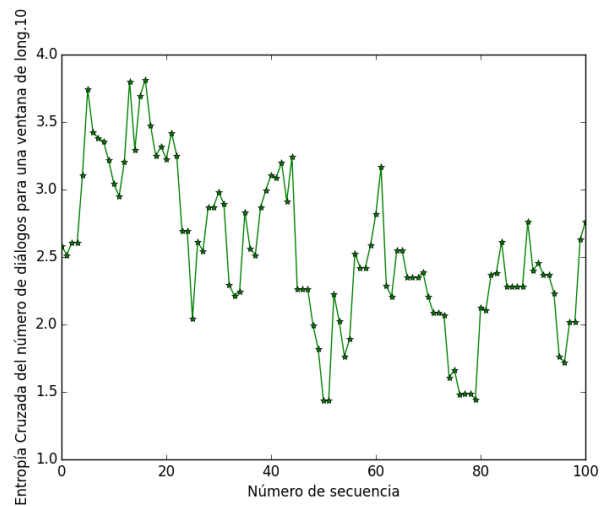


Figura N.3.2.4: Entropía cruzada del número de diálogos en “La isla mínima”

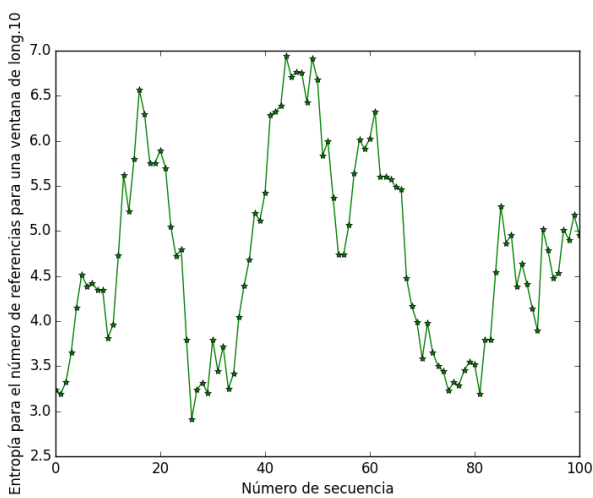


Figura N.3.2.5: Entropía del número de referencias en “La isla mínima”

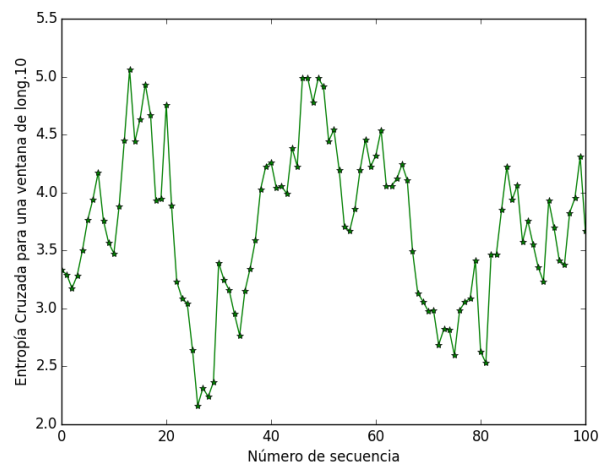


Figura N.3.2.6: Entropía del número de referencias en “La isla mínima”

N.4 Grafos

Se presentan a continuación tres grafos resultados de tres ejecuciones diferentes del programa. Algunos constan de menos o más elementos que otros, pero son en cualquier caso representativos de los resultados, no son casos extremos.

Los nodos con letras corresponden a personajes, mientras que los marcados con letras corresponden a temas. Los nodos unidos mediante aristas de trazo continuo han obtenido un alto valor de correlación de Información Mutua –ver Apartado 3.6.3-. Los nodos pintados del mismo color son aquellos que han resultado agrupados según esta métrica. Cuanto mayor correlación más grueso el trazo de la arista. Por otra parte, los nodos unidos mediante trazos discontinuos comparten alguna de las ideas que conforman los temas.

Lo primero que llama la atención es la cantidad de subgrafos distintos que aparecen y lo poco conectados que están entre ellos. Además los personajes principales prácticamente no aparecen, tan sólo en el último caso aparece Pedro, y con una relación muy poco relevante. No es fallo del programa que no aparezcan estos personajes, sino de la poca correlación que tienen con los temas.

Rodrigo, el padre de la chica asesinada, con “Familia” y ”Reproducción”. Martínez, el periodista, con “Periódico”. Andrés, el mayor de los guardia civiles, con “Enfermedades y accidentes – Curación”, “Mueble” y “Plaza”. Castro, el ex novio de la chica asesinada, con “Reproducción – Sexualidad”. Fernanda, la abuela, con “Provincia – Provincial”. Y finalmente Pedro, uno de los dos protagonistas, con “Cuerpo de ser vivo – Tórax”.

En cuanto a las relaciones entre temas, destacan “Agricultura” con “Comarca”, “Construcción” con “Idea, pensamiento”, “Cantidad (número) – Unir” con “Deporte ecuestre” y “Obra escultórica”- con “Sexualidad”.

Las dos únicas relaciones temáticas que se dan entre subgrafos, en los dos últimos ejemplos, se establecen mediante la idea de “Reproducción”. Quedan entonces relacionados en un caso Castro y Rodrigo y en otro Castro con “Obra escultórica”. Unas relaciones muy poco significativas.

Como ya se decía en el Apartado 4.3 el grafo muestra una pobre estructuración del guión en comparación con los dos anteriores.

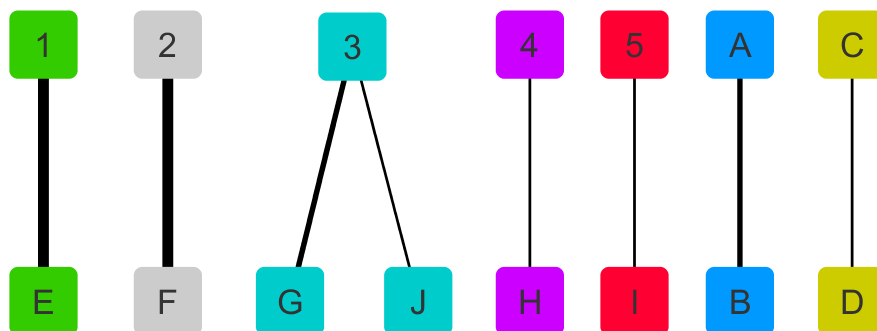


Figura N.4.1: Grafo de relaciones entre temas y personajes para “La isla mínima”

- TEMAS -

A	Cantidad (número) – Unir
B	Deporte ecuestre
C	Ser humano
D	Sentimientos (en general) – Optimismo
E	Familia (en general) – Familia nuclear
F	Periódico
G	Enfermedades y accidentes – Curación
H	Reproducción (biología) – Sexualidad
I	Provincia – Provincial
J	Plaza - Plaza

- PERSONAJES -

1	RODRIGO
2	MARTÍNEZ
3	ANDRÉS
4	CASTRO
5	FERNANDA

-ARISTAS-

—————	Distribución relacionada (MI)
-----	Temática relacionada (DIC)

Tabla N.4.1: Leyenda de temas y personajes de la Figura N.4.1

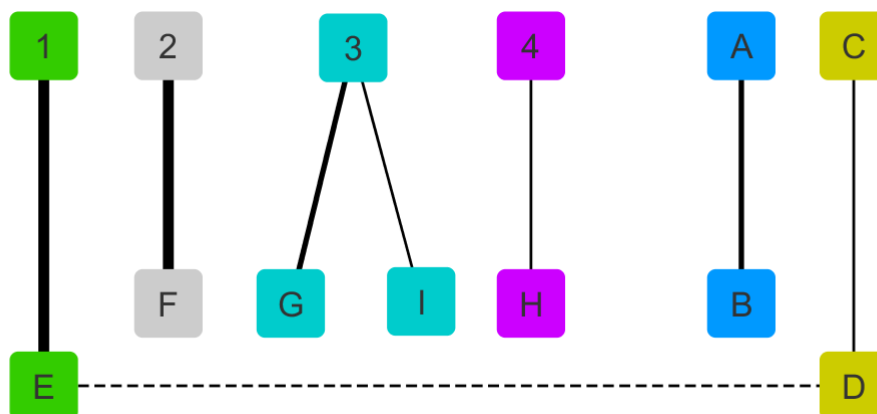


Figura N.4.2: Grafo de relaciones entre temas y personajes para “La isla mínima”

- TEMAS -

A	Cantidad (número) – Unir
B	Deporte ecuestre
C	Artes plásticas – Obra escultórica
D	Reproducción (biología) – Sexualidad
E	Reproducción (biología) – Reproducirse (fisiología)
F	Periódico
G	Enfermedades y accidentes – Curación
H	Provincia – Provincial
I	Mueble

- PERSONAJES -

1	RODRIGO
2	MARTÍNEZ
3	ANDRÉS
4	FERNANDA

-ARISTAS-

—————	Distribución relacionada (MI)
- - - - -	Temática relacionada (DIC)

Tabla N.4.2: Leyenda de temas y personajes de la Figura N.4.2

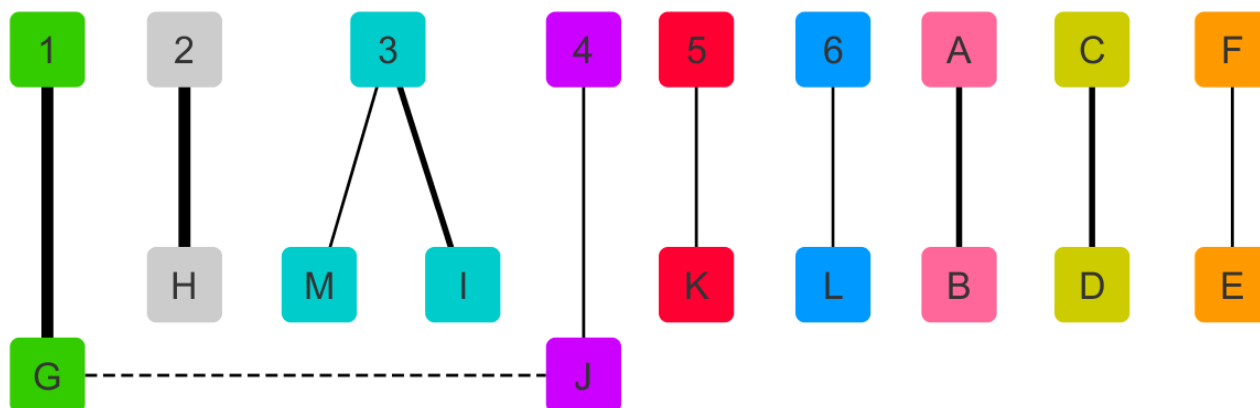


Figura N.4.3: Grafo de relaciones entre temas y personajes para “La isla mínima”

- TEMAS -	
A	Cantidad (número) – Unir
B	Deporte ecuestre
C	Construcción (en general) – Construir (edificio)
D	Idea, pensamiento
E	Comarca – Comarca
F	Agricultura – Labor agrícola
G	Reproducción (biología) – Reproducirse (fisiología)
H	Periódico
I	Enfermedades y accidentes - Curación
J	Reproducción (biología) – Sexualidad
K	Cuerpo de ser vivo – Tórax
L	Provincia – Provincial
M	Plaza - Plaza
- PERSONAJES -	
1	RODRIGO
2	MARTÍNEZ
3	ANDRÉS
4	CASTRI
5	PEDRO
6	FERNANDA

-ARISTAS-	
—————	Distribución relacionada (MI)
-----	Temática relacionada (DIC)

Tabla N.4.3: Leyenda de temas y personajes de la Figura N.4.3

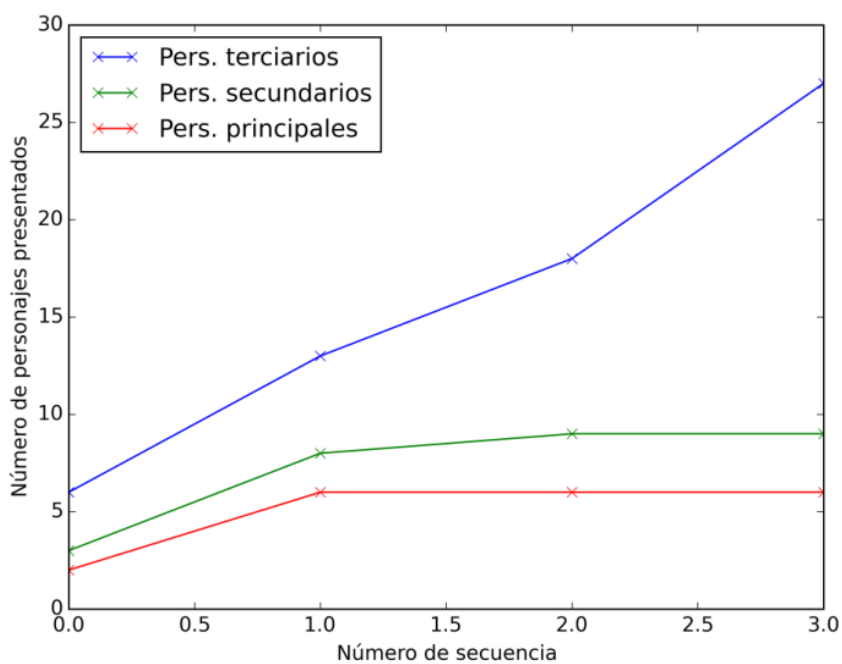


Figura Ñ.1.2: Presentación de los personajes para las Actas, estructura original.

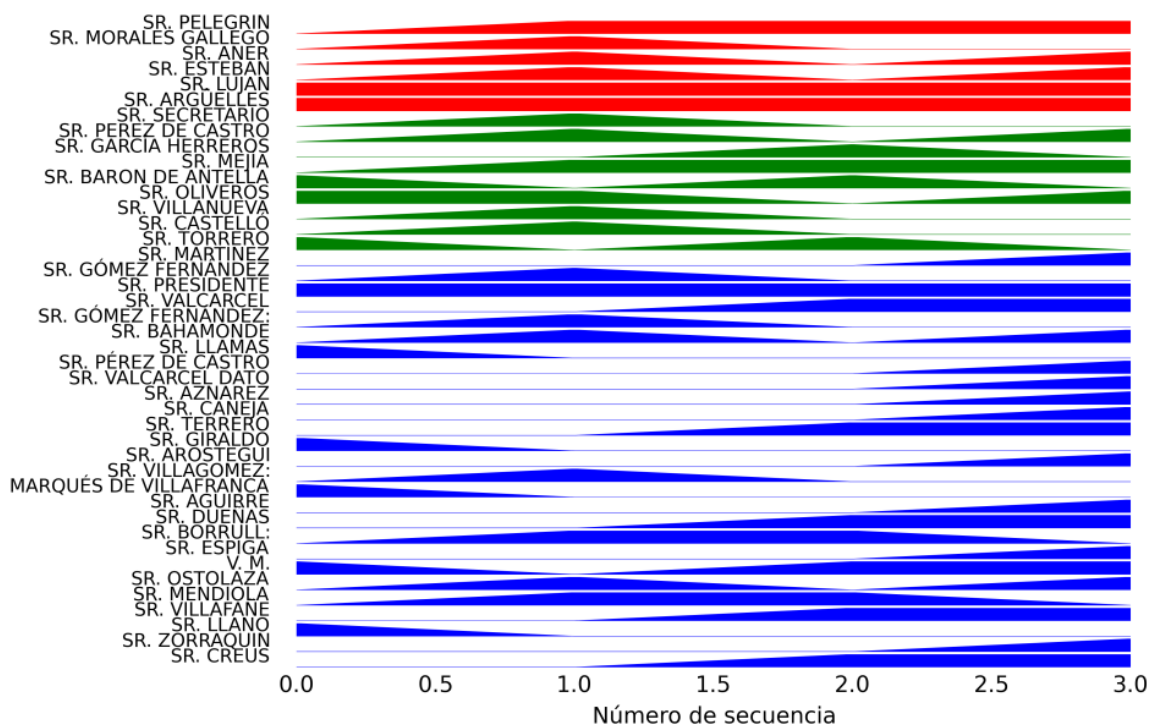


Figura Ñ.1.3: Aparición de los personajes para las Actas, estructura original.

Ñ.2 Temas

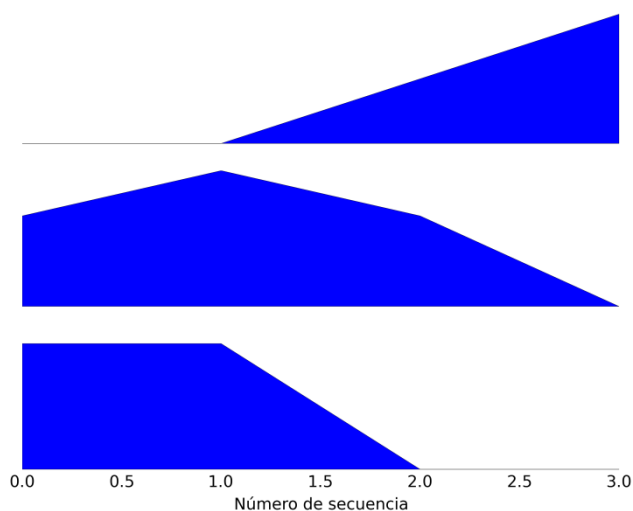


Figura Ñ.2: Distribución de los temas para las Actas, estructura original.

Ñ.3 Entropía

No hay suficientes secuencias como para poder realizar las medidas de entropía.

Ñ.4 Grafos

Se presenta a continuación el grafo resultante de una ejecución del programa. Otras ejecuciones arrojan resultados de menos o más elementos, pero no se alejan en cualquier caso del ejemplo aquí mostrado.

Los nodos con letras corresponden a personajes, mientras que los marcados con letras corresponden a temas. Los nodos unidos mediante aristas de trazo continuo han obtenido un alto valor de correlación de Información Mutua –ver Apartado 3.6.3-. Los nodos pintados del mismo color son aquellos que han resultado agrupados según esta métrica. Cuanto mayor correlación más grueso el trazo de la arista. Por otra parte, los nodos unidos mediante trazos discontinuos comparten alguna de las ideas que conforman los temas.

Lo primero que llama la atención es que todo el grafo se encuentra fuertemente conectado y, además, que lo hace siguiendo una topología en estrella que no hemos visto en los casos de los guiones.

La centralidad está ocupada por dos ideas -“Actividades y estados físicos” y “Movilidad, inmovilidad”- a las que se refieren todos los personajes.

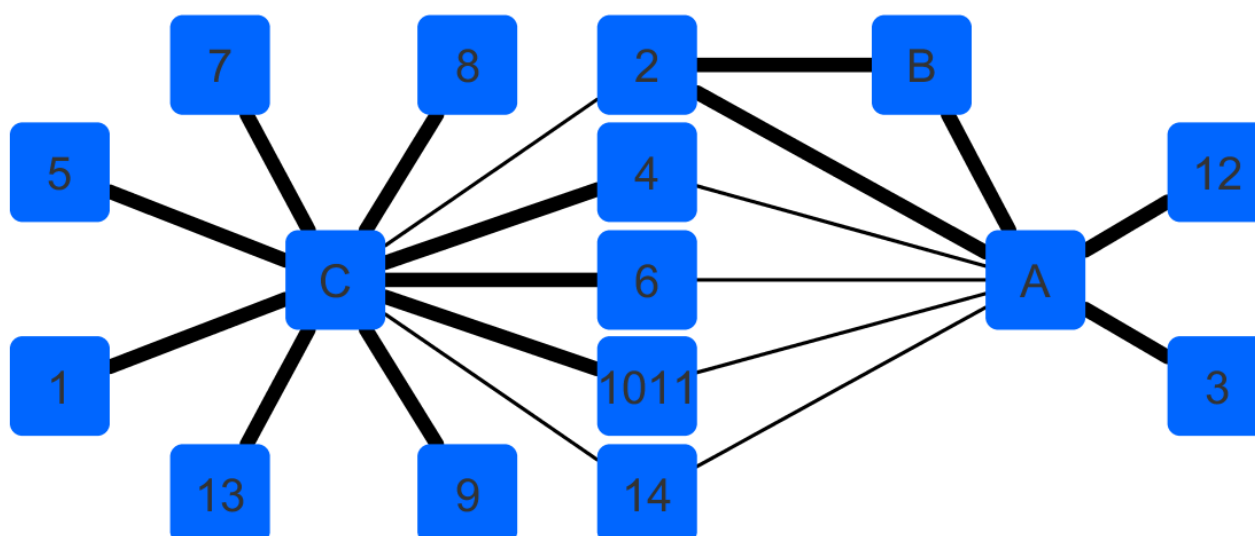


Figura Ñ.4.1: Grafo de relaciones entre temas y personajes para las Actas, estructura original

		- TEMAS -
A	Actividades y estados físicos	
B	Movilidad, inmovilidad	
C	Idea, pensamiento	
		- PERSONAJES -
1	SR. SECRETARIO	
2	SR. LUJAN	
3	SR. ARGÜELLES	
4	SR. MEJÍA	
5	SR. CASTELLÓ	
6	SR. ANER	
7	SR. VILLANUEVA	
8	SR. MORALES GALLEGO	
9	SR. ESTEBAN	
12	SR. GARCÍA HERREROS	
13	SR. PEREZ DE CASTRO	
14	SR. TORRERO	
1011	SR. PELEGRIN	
		-ARISTAS-
—————	Distribución relacionada (MI)	
- - - - -	Temática relacionada (DIC)	

Tabla Ñ.4.1: Leyenda de temas y personajes de la Figura Ñ.4.1

O

Actas, estructura uniforme

O.1 Personajes

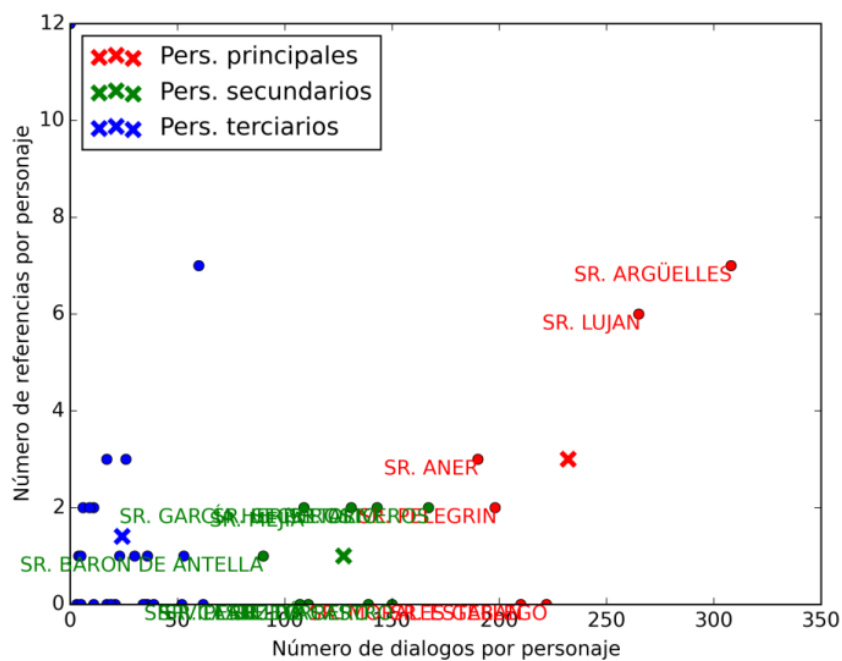


Figura O.1.1: Protagonismo de los personajes para las Actas, estructura uniforme.

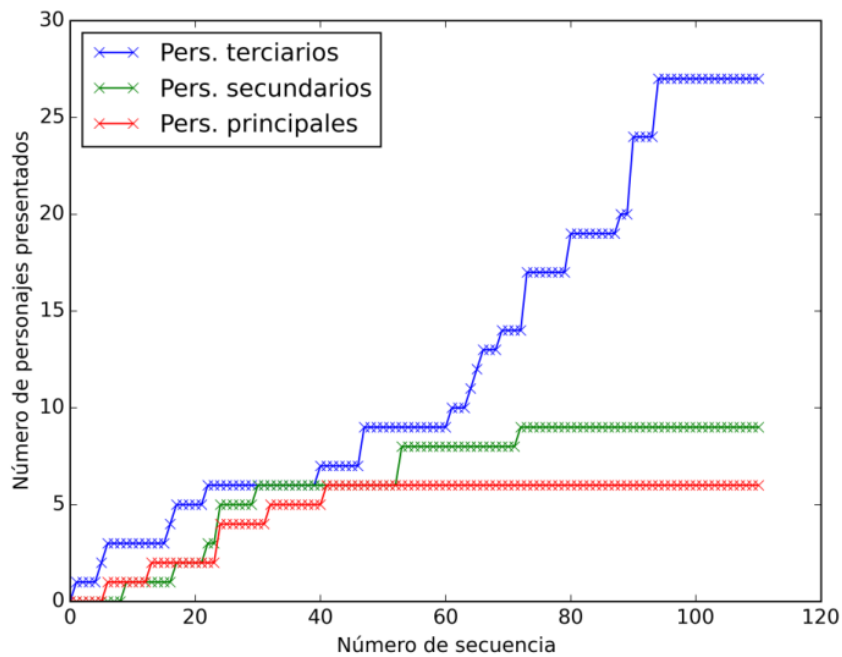


Figura O.1.2: Presentación de los personajes para las Actas, estructura uniforme

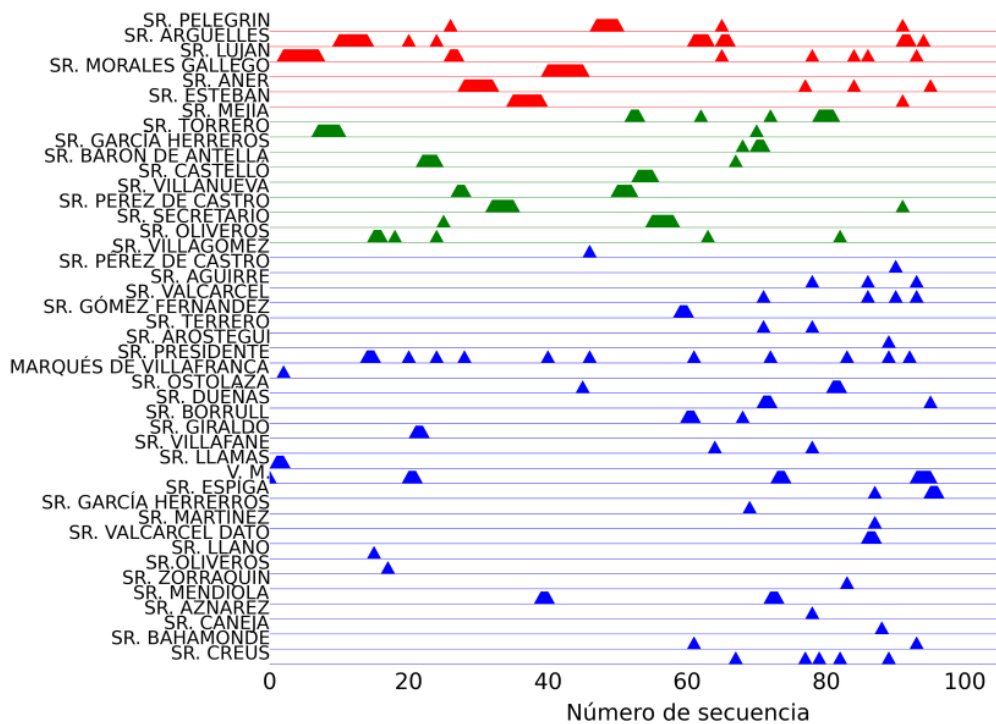


Figura O.1.3: Aparición de los personajes para las Actas, estructura uniforme

O.2 Temas

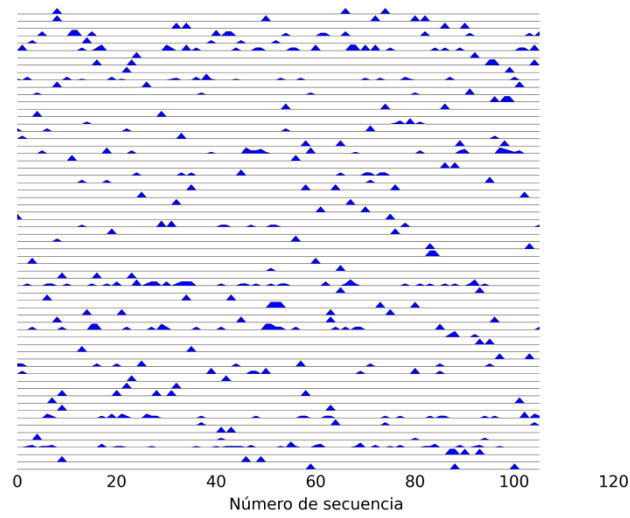


Figura O.2: Distribución de los temas para las Actas, estructura uniforme.

O.3 Entropía

O.3.1 Entropía temas

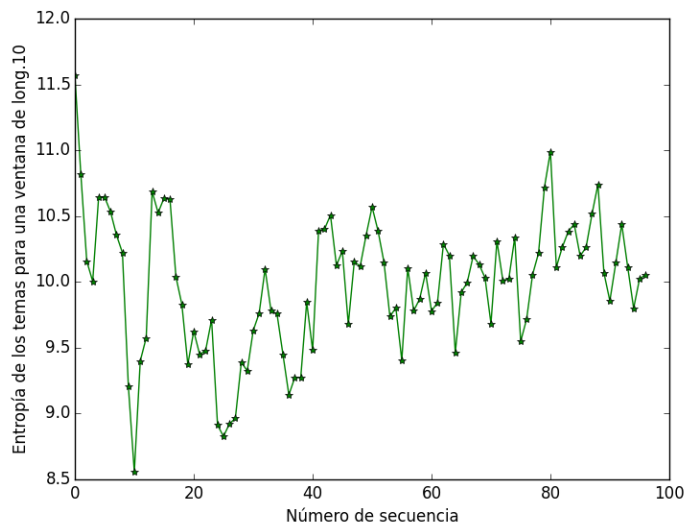


Figura O.3.1.1: Entropía de los temas para las Actas, estructura uniforme

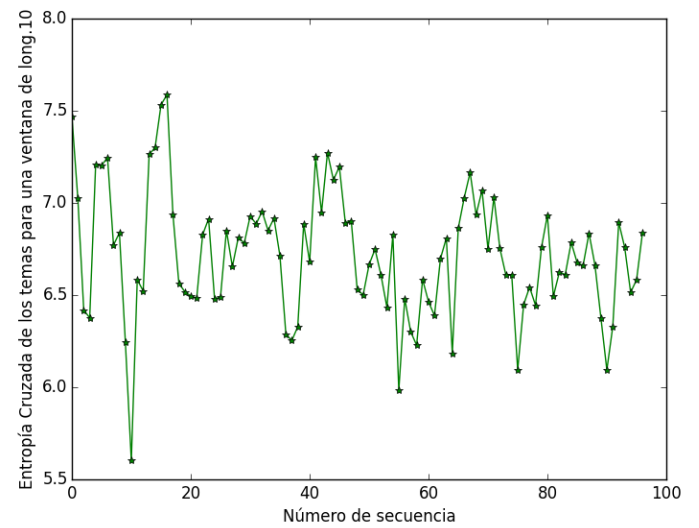


Figura O.3.1.2: Entropía cruzada para las Actas, estructura uniforme

O.3.2 Entropía personajes

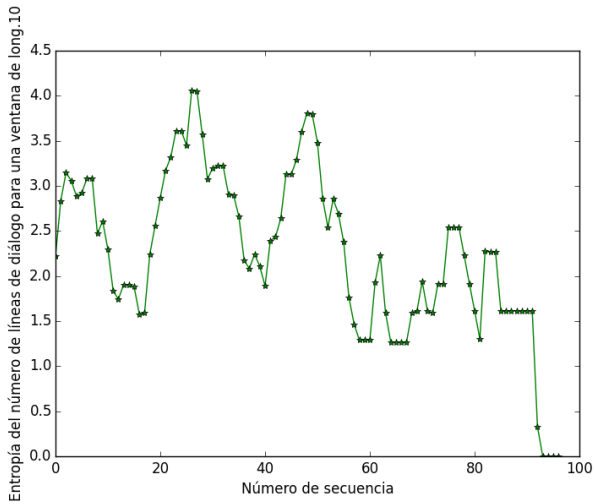


Figura O.3.2.1: Entropía del número de líneas de diálogo para las Actas, estructura uniforme

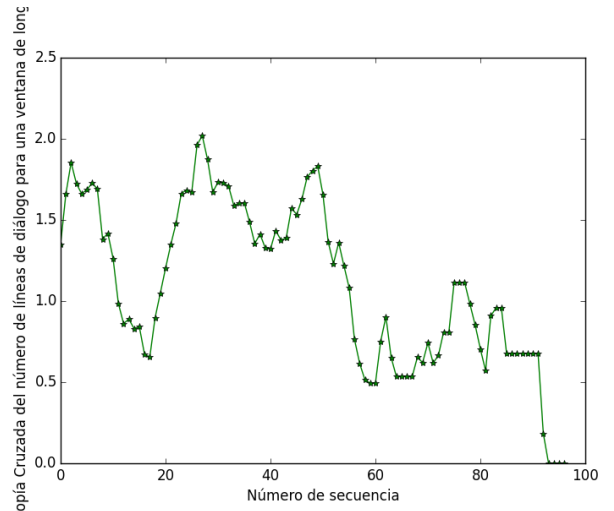


Figura O.3.2.2: Entropía cruzada del número de líneas para las Actas, estructura uniforme

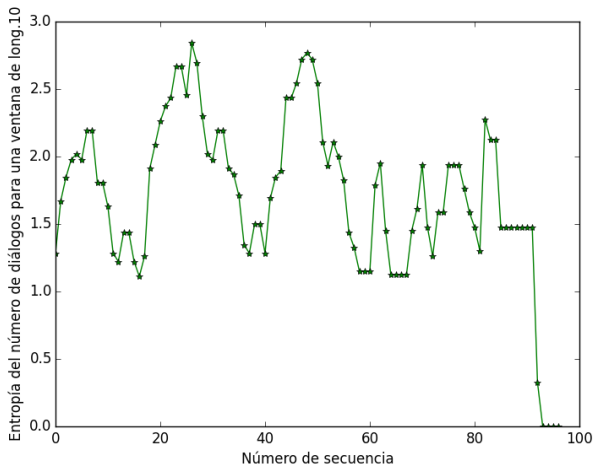


Figura O.3.2.3: Entropía del número de diálogos para las Actas, estructura uniforme

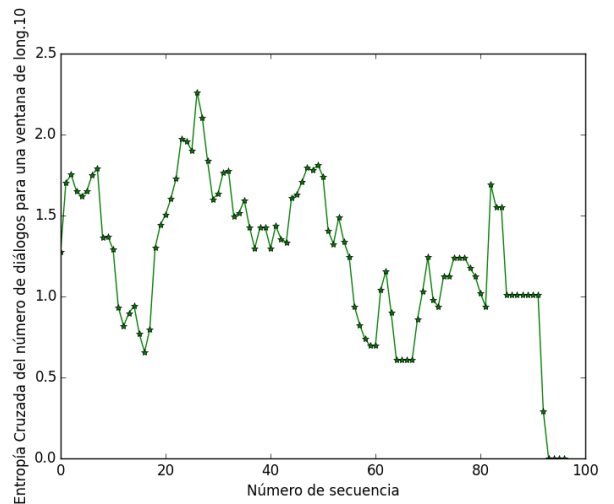


Figura O.3.2.4: Entropía cruzada del número de diálogos para las Actas, estructura uniforme

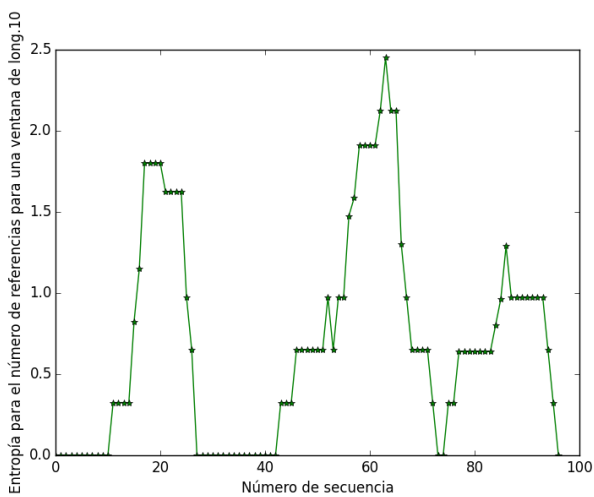


Figura O.3.2.5: Entropía del número de referencias para las Actas, estructura uniforme

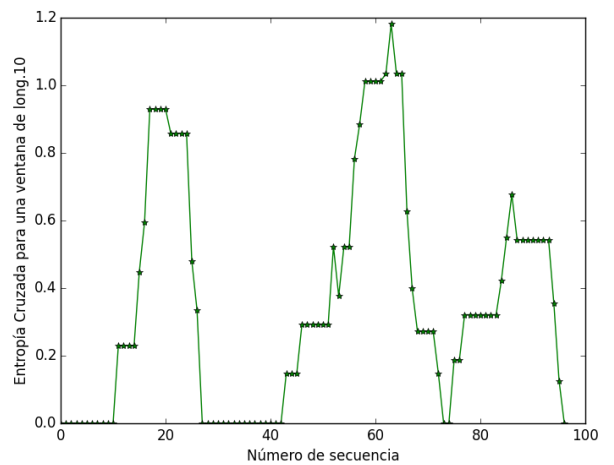


Figura O.3.2.6: Entropía del número de referencias para las Actas, estructura uniforme

O.4 Grafos

Se presenta a continuación el grafo resultante de una ejecución del programa. Otras ejecuciones arrojan resultados de menos o más elementos, pero no se alejan en cualquier caso del ejemplo aquí mostrado.

Los nodos con letras corresponden a personajes, mientras que los marcados con letras corresponden a temas. Los nodos unidos mediante aristas de trazo continuo han obtenido un alto valor de correlación de Información Mutua –ver Apartado 3.6.3-. Los nodos pintados del mismo color son aquellos que han resultado agrupados según esta métrica. Cuanto mayor correlación más grueso el trazo de la arista. Por otra parte, los nodos unidos mediante trazos discontinuos comparten alguna de las ideas que conforman los temas.

Lo primero que llama la atención es la baja cantidad de elementos que presenta. De todos los personajes principales y secundarios, tan sólo dos presentan algún tipo de correlación con los temas.

La distribución uniforme de secuencias impuesta al guión para este caso parece arrojar una estructura prácticamente nula.

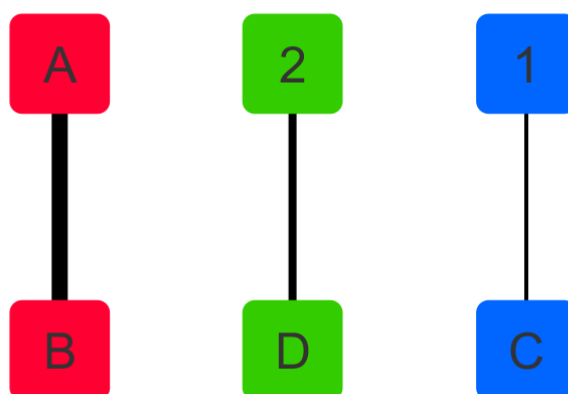


Figura P.4.1: Grafo de relaciones entre temas y personajes para las Actas, estructura uniforme

		- TEMAS -
A	Amistades – Amistad / Enemistad	
B	Idea, pensamiento	
C	Sociología - Libertad	
D	Desarrollo – Desarrollarse (crecer)	
		- PERSONAJES -
1	SR. MORALES GALLEGO	
2	SR. GARCÍA HERREROS	
		-ARISTAS-
—————	Distribución relacionada (MI)	
-----	Temática relacionada (DIC)	

Tabla P.4.1: Leyenda de temas y personajes de la Figura P.4.1

P

Actas, estructura narrativa

P.1 Personajes

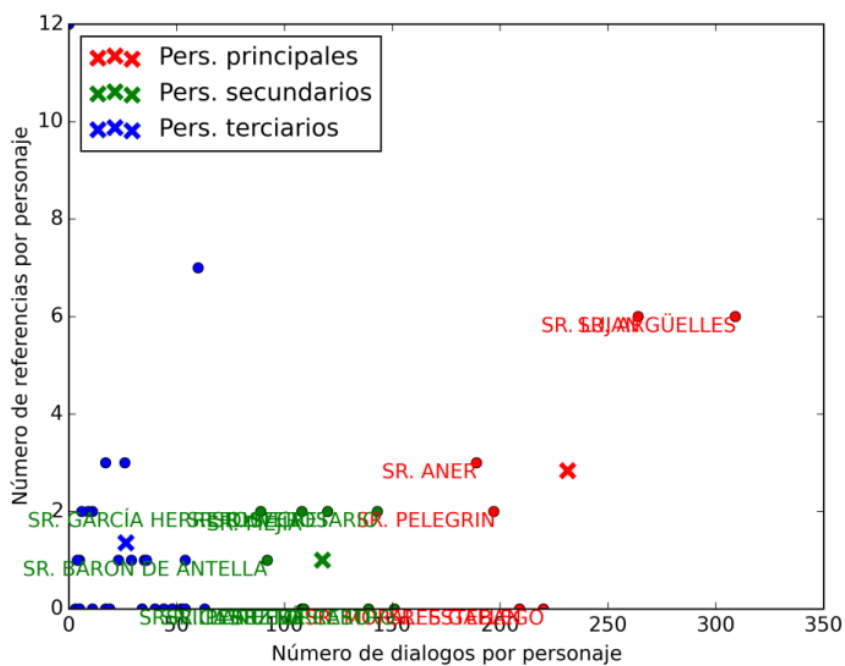


Figura P.1.1: Protagonismo de los personajes para las Actas, estructura narrativa

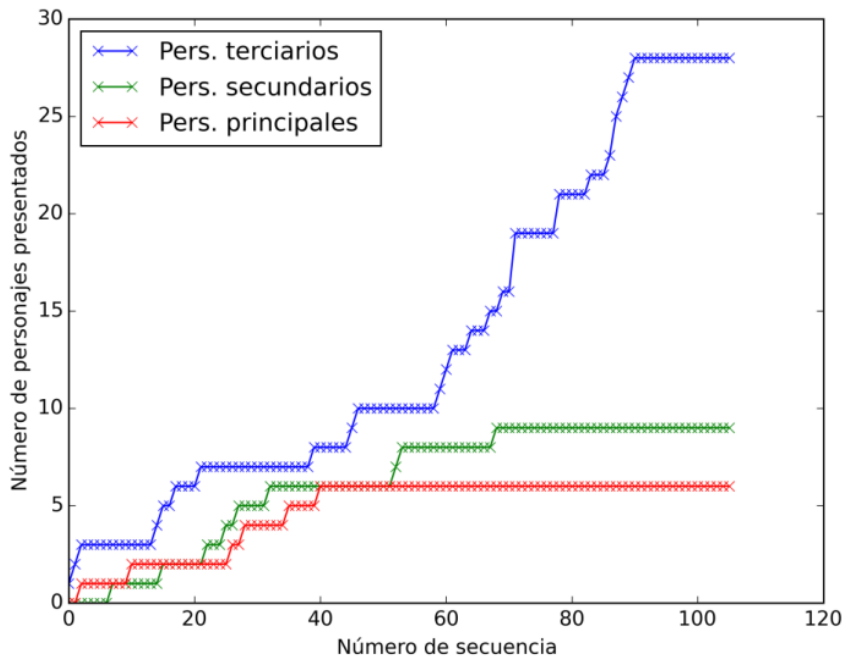


Figura P.1.2: Presentación de los personajes para las Actas, estructura narrativa

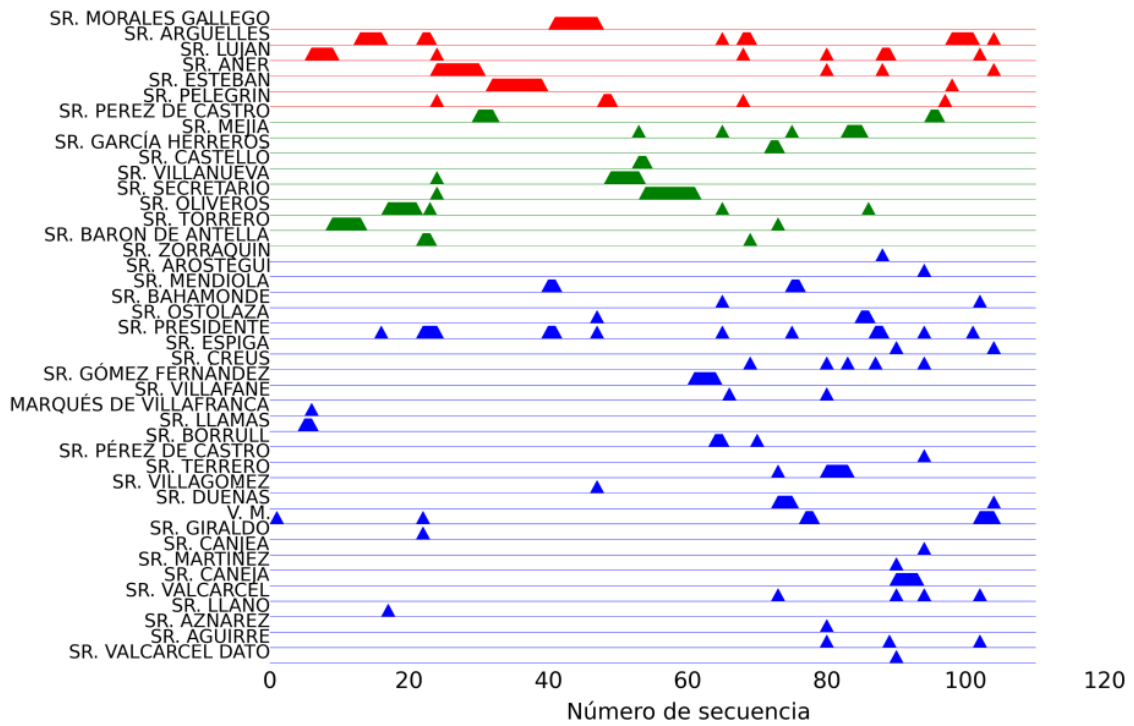


Figura P.1.3: Aparición de los personajes para las Actas, estructura narrativa

P.2 Temas

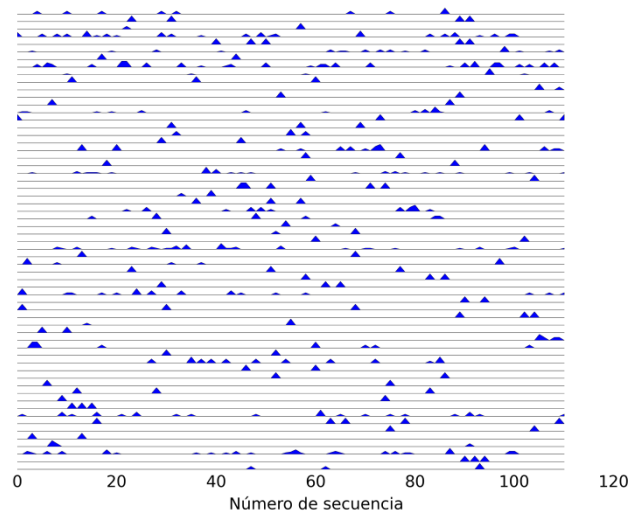


Figura P.2: Distribución de los temas para las Actas, estructura narrativa

P.3 Entropía

P.3.1 Entropía temas

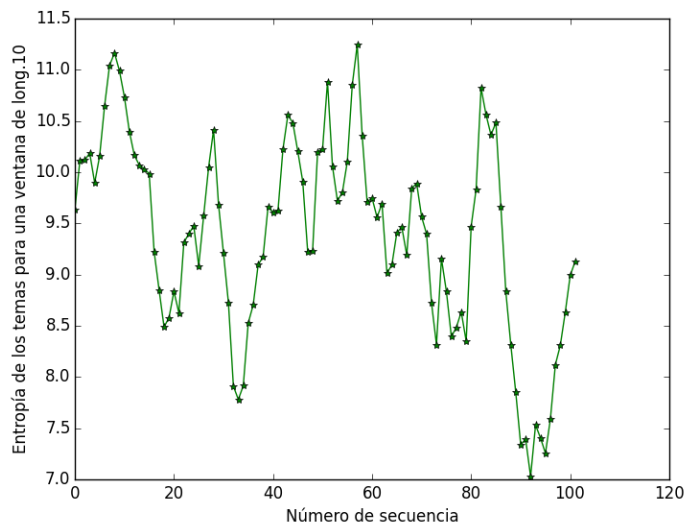


Figura P.3.1.1: Entropía de los temas para las Actas, estructura narrativa

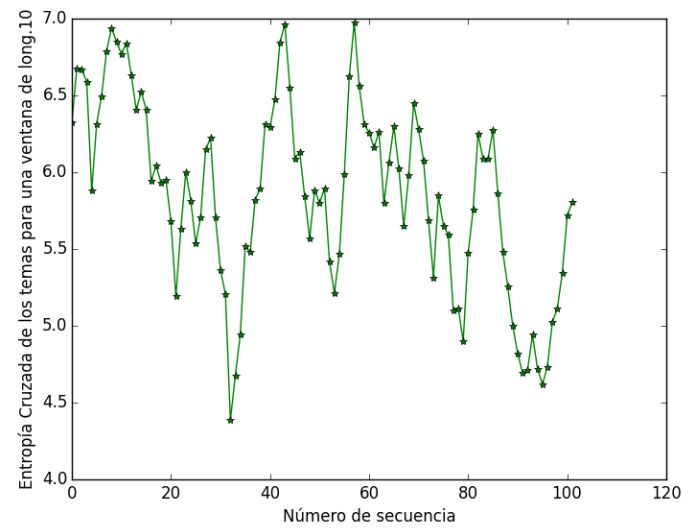


Figura P.3.1.2: Entropía cruzada de los temas para las Actas, estructura narrativa

P.3.2 Entropía personajes

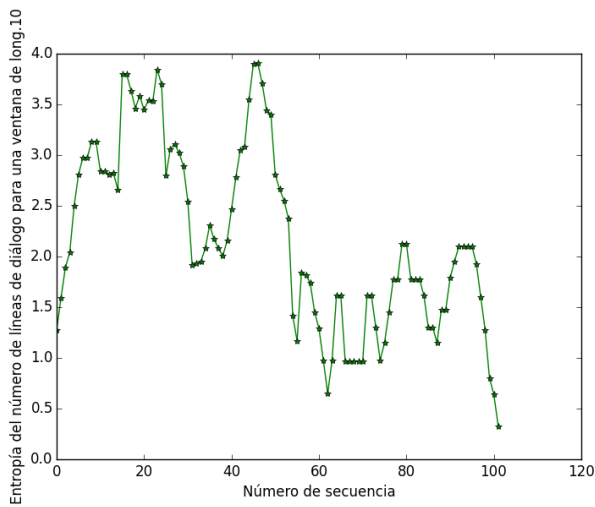


Figura P.3.2.1: Entropía del número de líneas de diálogo para las Actas, estructura narrativa

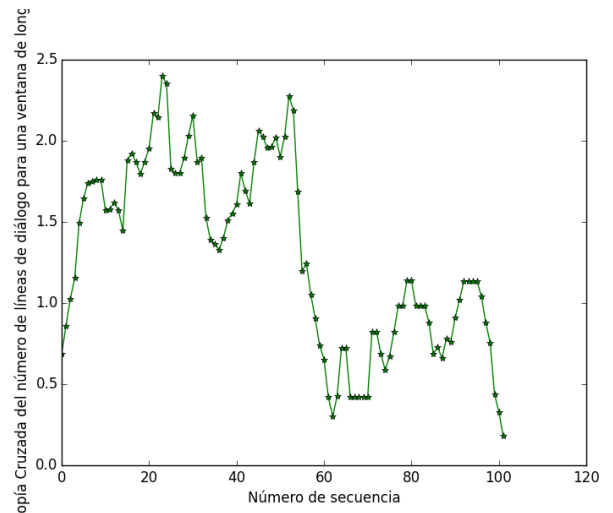


Figura P.3.2.2: Entropía cruzada del número de líneas de diálogo para las Actas, estructura narrativa

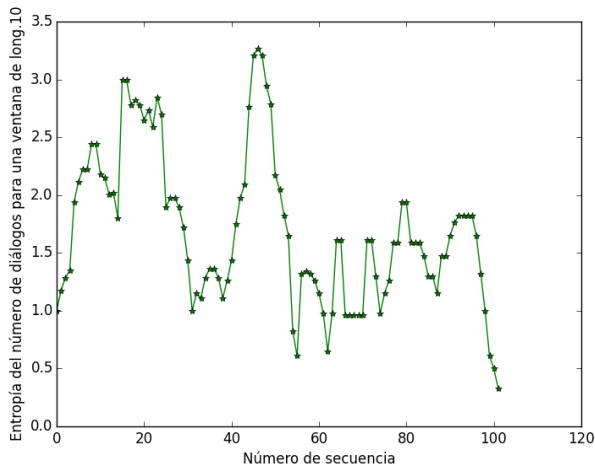


Figura P.3.2.3: Entropía del número de diálogos para las Actas, estructura narrativa

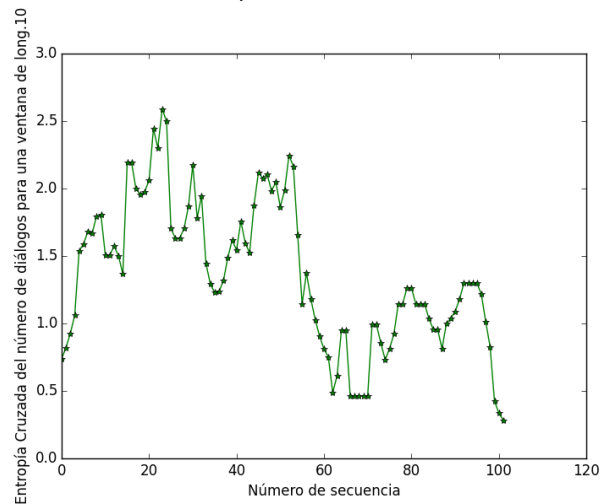


Figura P.3.2.4: Entropía cruzada del número de diálogos para las Actas, estructura narrativa

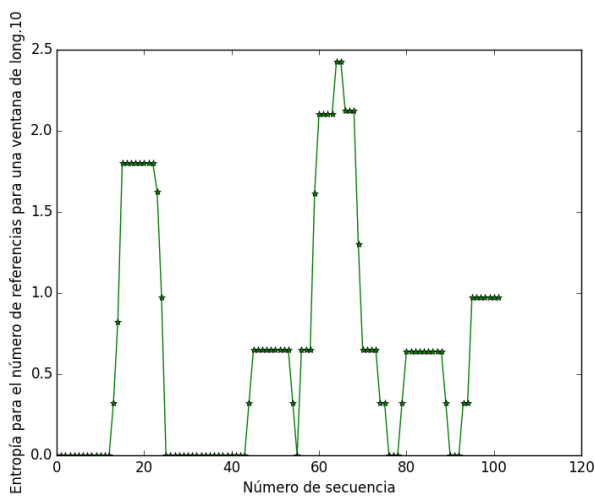


Figura P.3.2.5: Entropía del número de referencias para las Actas, estructura narrativa

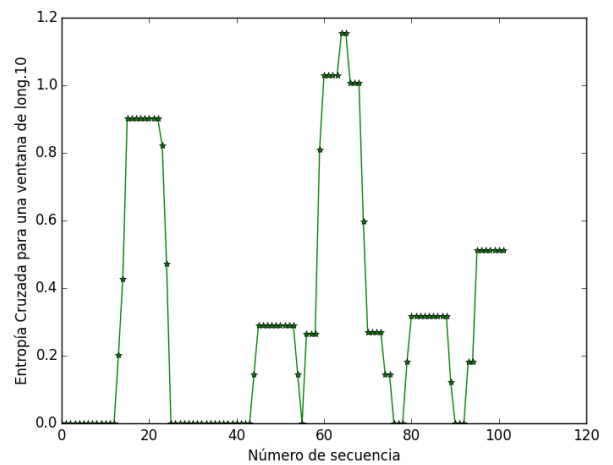


Figura P.3.2.6: Entropía del número de referencias para las Actas, estructura narrativa

P.4 Grafos

Se presenta a continuación el grafo resultante de una ejecución del programa. Otras ejecuciones arrojan resultados de menos o más elementos, pero no se alejan en cualquier caso del ejemplo aquí mostrado.

Los nodos con letras corresponden a personajes, mientras que los marcados con letras corresponden a temas. Los nodos unidos mediante aristas de trazo continuo han obtenido un alto valor de correlación de Información Mutua –ver Apartado 3.6.3-. Los nodos pintados del mismo color son aquellos que han resultado agrupados según esta métrica. Cuanto mayor correlación más grueso el trazo de la arista. Por otra parte, los nodos unidos mediante trazos discontinuos comparten alguna de las ideas que conforman los temas.

El número de elementos aumenta, pasando de la estructura prácticamente inexistente del caso anterior a otro más parecido al de “La isla mínima”, del que toma su estructura.

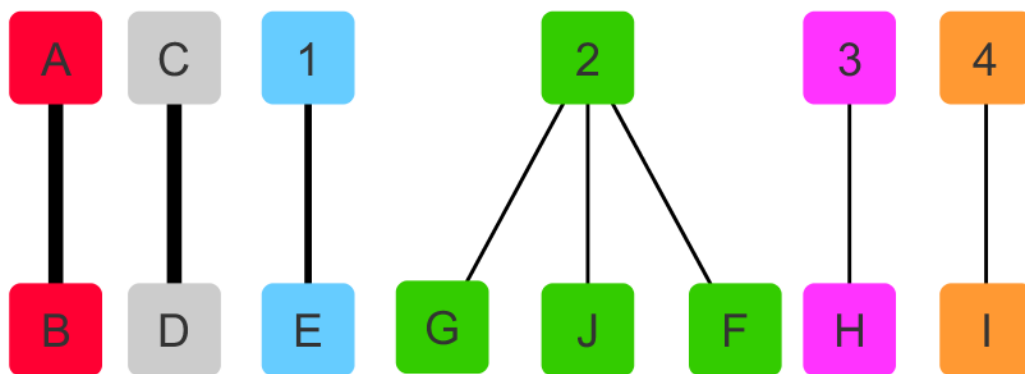


Figura P.4.1: Grafo de relaciones entre temas y personajes para las Actas, estructura narrativa

	- TEMAS -
A	Provincia
B	Cantidad (número) – Unir
C	Trabajadora (persona que trabaja)
D	Oficina
E	Comunicación y lenguaje
F	Tecnología (Física y Química aplicadas) – Mecánica
G	Industria alimentaria – Elaboración de alimentos
H	Actitud (sicología) – Irresponsabilidad (no reflexiv@)
J	Espacio
J	Sicología – Carácter (personalidad)
	- PERSONAJES -
1	SR. GARCÍA HERREROS
2	SR. TORRERO
3	SR. BARÓN DE ANTELLA
4	SR. CASTELLÓ
	-ARISTAS-
—————	Distribución relacionada (MI)
-----	Temática relacionada (DIC)

Tabla P.4.1: Leyenda de temas y personajes de la Figura P.4.1

Q

Para Elías

Q.1 Personajes

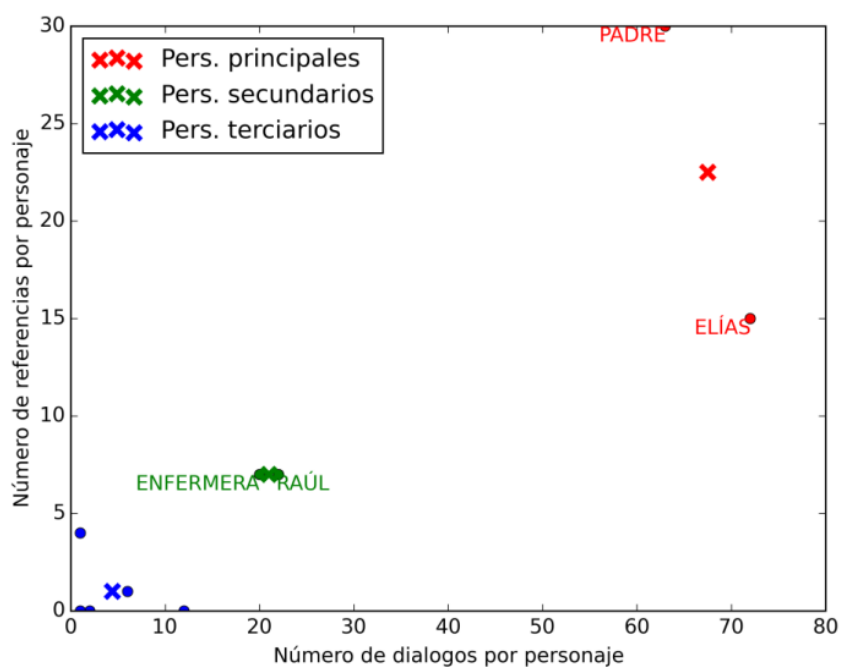


Figura Q.1.1: Protagonismo de los personajes en “Para Elías”

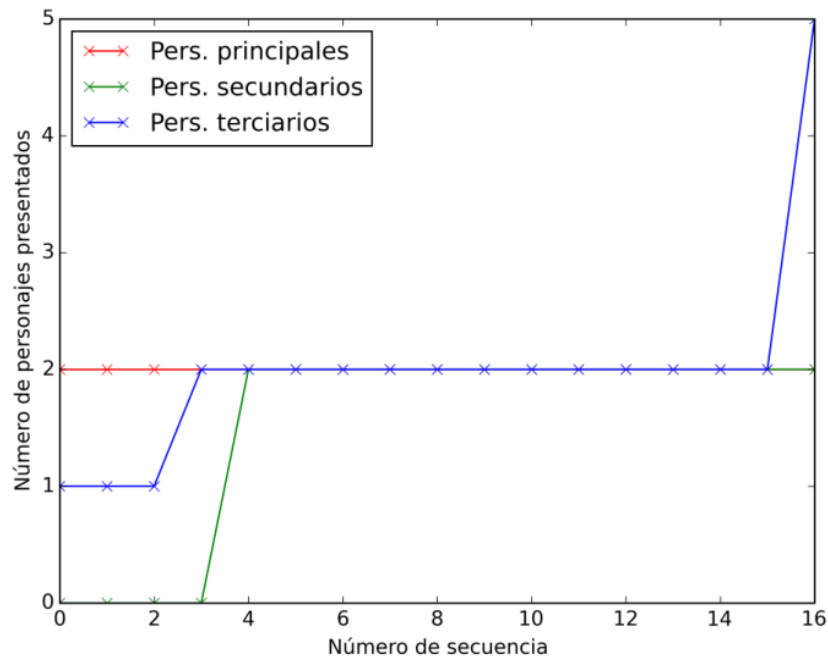


Figura Q.1.2: Presentación de los personajes en "Para Elías"

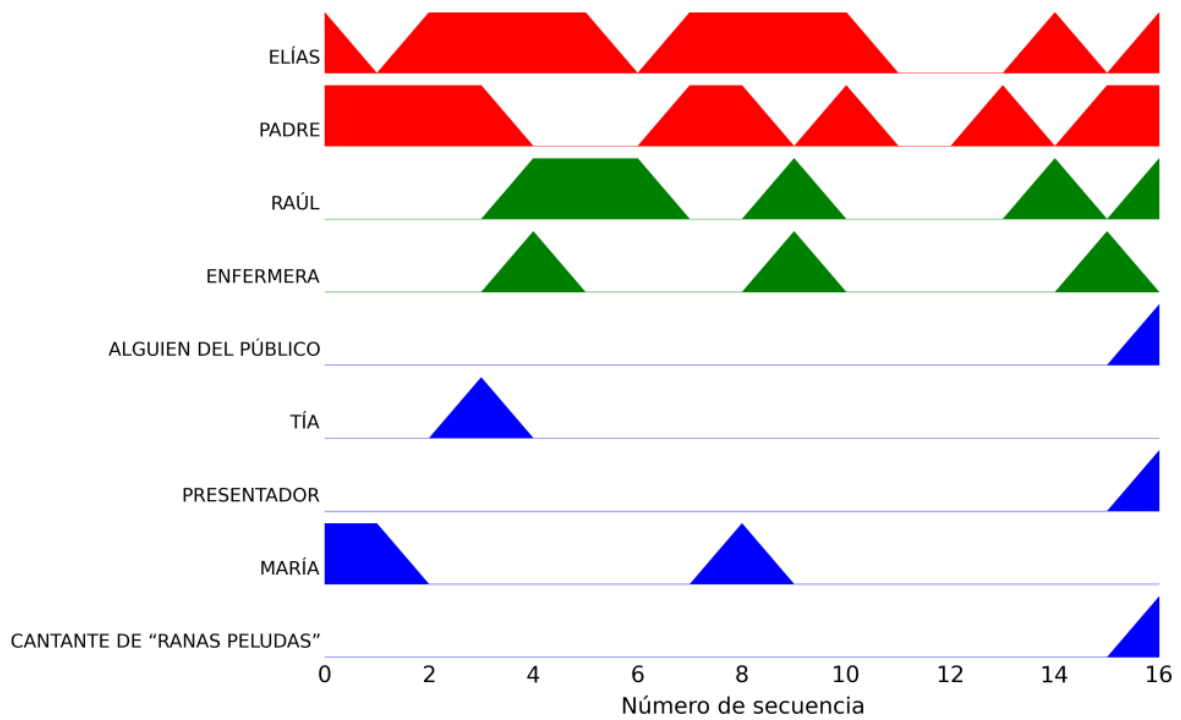


Figura Q.1.3: Aparición de los personajes en "Para Elías"

Q.2 Temas

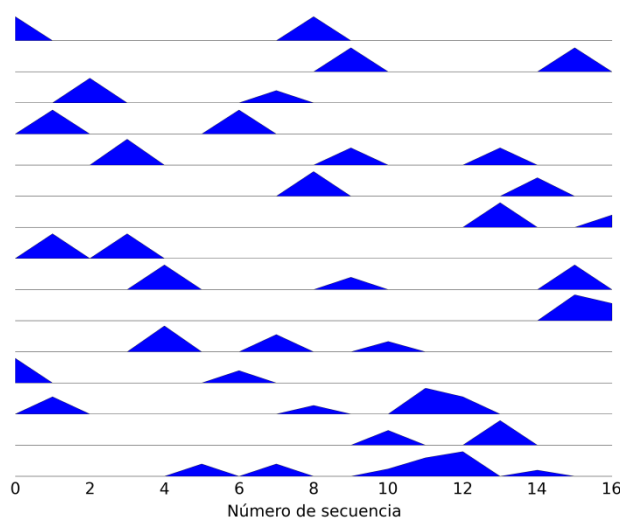


Figura Q.2: Distribución de los temas en “Para Elías”

Q. 3 Entropía

Debido al bajo número de secuencias las gráficas de entropía constan de muy pocos elementos y no resultan nada informativas.

Q.4 Grafos

Se presenta a continuación los grafos resultantes de par de ejecuciones del programa. Otras ejecuciones arrojan distintos resultados, pero no se alejan en cualquier caso los aquí mostrados.

Los nodos con letras corresponden a personajes, mientras que los marcados con letras corresponden a temas. Los nodos unidos mediante aristas de trazo continuo han obtenido un alto valor de correlación de Información Mutua –ver Apartado 3.6.3-. Los nodos pintados del mismo color son aquellos que han resultado agrupados según esta métrica. Cuanto mayor correlación más grueso el trazo de la arista. Por otra parte, los nodos unidos mediante trazos discontinuos comparten alguna de las ideas que conforman los temas.

Suelen aparecer tres subgrafos. Uno principal en el que se relacionan el padre y el hijo, otro en el que está la enfermera y otro con un par de temas. Padre e hijo se relacionan por “Actividades y estados físicos” y “Comunicación – Lenguaje escrito”. La enfermera con “Enfermedades y accidentes – Curación” y “Actitud (sicología”, relaciones sin duda interesantes.. La relación entre temas se da entre “Comunicación y lenguaje (lenguaje escrito)” y “Religión cristiana”. Puede que este tema tenga que ver con las secuencias en el cementerio.

Elías se relaciona también con Raúl mediante “Idea, pensamiento” y “Cuerpo de ser vivo (cara)”, relacionado a su vez con “Tiempo”.

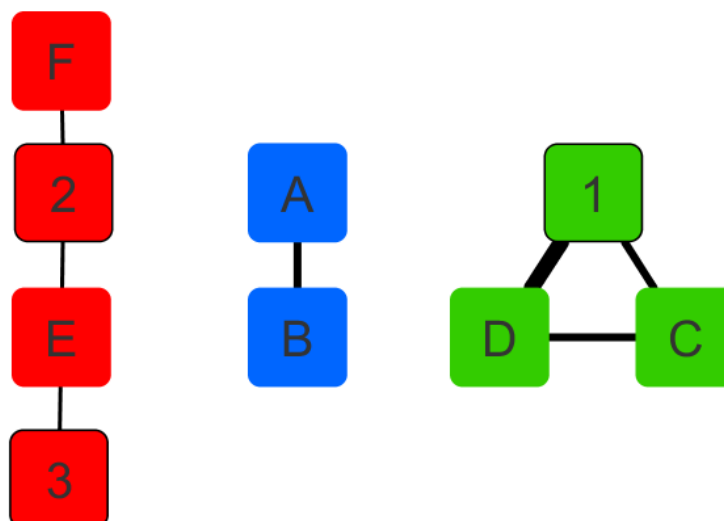


Figura Q.4.1: Grafo de relações entre temas e personagens para “Para Elías”

- TEMAS -	
A	Comunicação y lenguaje – Lenguaje escrito
B	Religión cristiana
C	Enfermedades y accidentes – Curación
D	Actitud (sicología)
E	Actividades y estados físicos
F	Movilidad, inmovilidad
- PERSONAJES -	
1	ENFERMERA
2	ELÍAS
3	PADRE
-ARISTAS-	
—————	Distribución relacionada (MI)
-----	Temática relacionada (DIC)

Tabla Q.4.1: Leyenda de temas y personajes de la Figura Q.4.1

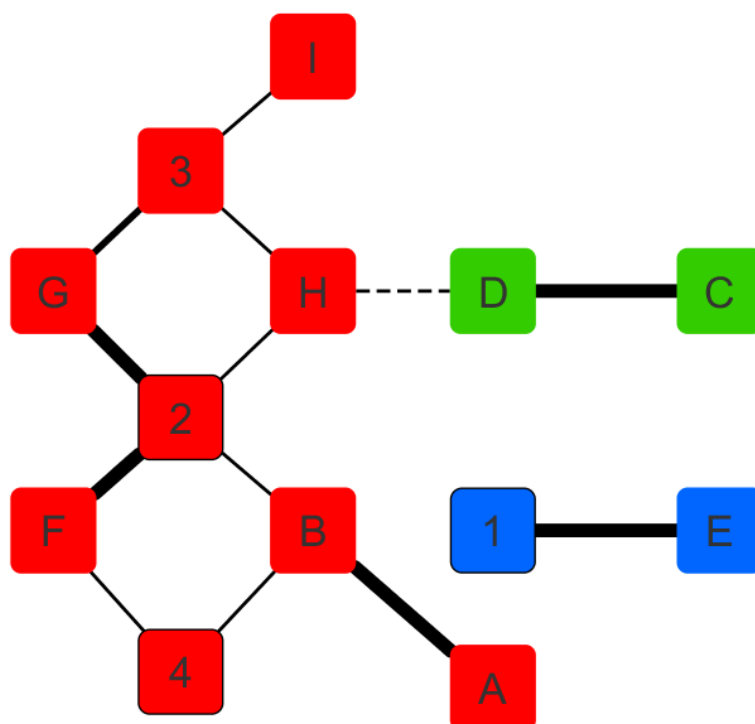


Figura Q.4.2: Gráfico de relações entre temas e personagens para “Para Elías”

- TEMAS -

A	Tiempo (reloj, medida)
B	Cuerpo de ser vivo – Cara (rostro)
C	Religión cristiana
D	Comunicación y lenguaje – Lenguaje escrito
E	Enfermedades y accidentes – Curación
F	Idea, pensamiento
G	Actividades y estados físicos
H	Comunicación y lenguaje – Lenguaje oral
J	Eukariota (Dominio), Animalia (Reino)
J	Cantidad (número)

- PERSONAJES -

1	ENFERMERA
2	ELÍAS
3	PADRE
4	RAÚL

-ARISTAS-

—————	Distribución relacionada (MI)
- - - - -	Temática relacionada (DIC)

Tabla Q.4.2: Leyenda de temas y personajes de la Figura Q.4.2

Q.5 Escenas

ESCENA	PROB. DE CONFORMAR UNA ESCENA	DESCRIPCIÓN DE LA ESCENA
(0, 9)	0.625	Presentación de la familia. Muere la madre. Elías empieza a trabajar en el hospital. Conoce a Raúl y va a verle tocar. Discute con su padre. Decide volver a tocar el piano.
(10, 16)	0.4164	Montaje de secuencia. Padre y Enfermera. Concierto en el bar. Reconciliación y final.

1 Tabla Q.5: Escenas de “Para Elías”. Secuencias, probabilidad y descripción

Q.6 Actos

Q.6.1 Cambio Acto I – II

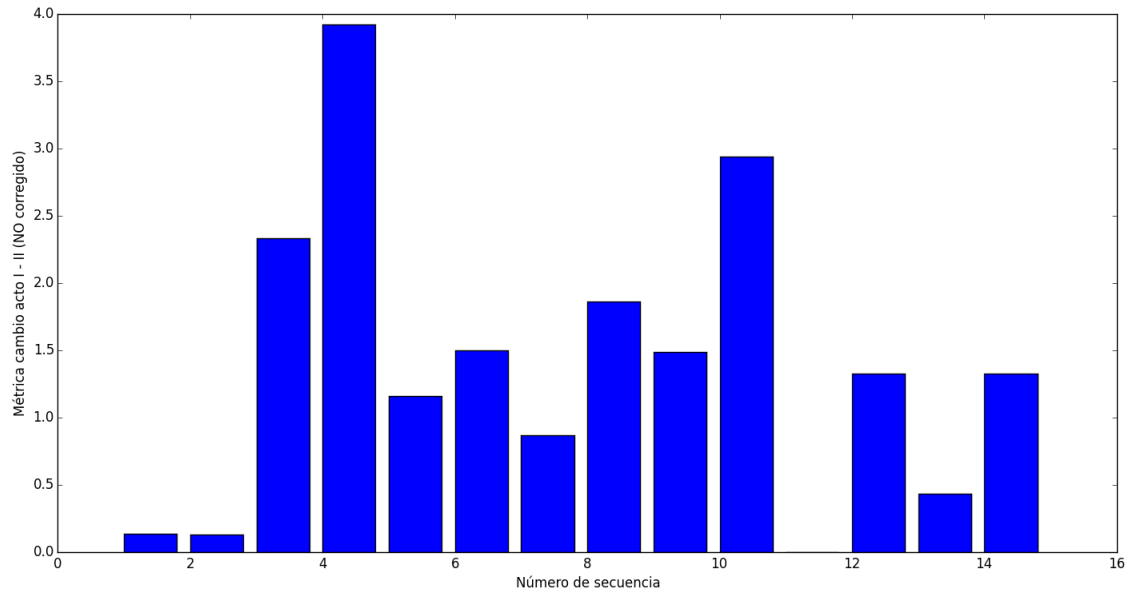


Figura Q.6.1.1: Probabilidad de cambio del I al II Acto sin corrección teórica para “Para Elías”

Cambio de I a II ACTO en la secuencia 4, con valor 3.9219.

Líneas 245 a 289.

Cambio teórico en la secuencia 3 +- 1.

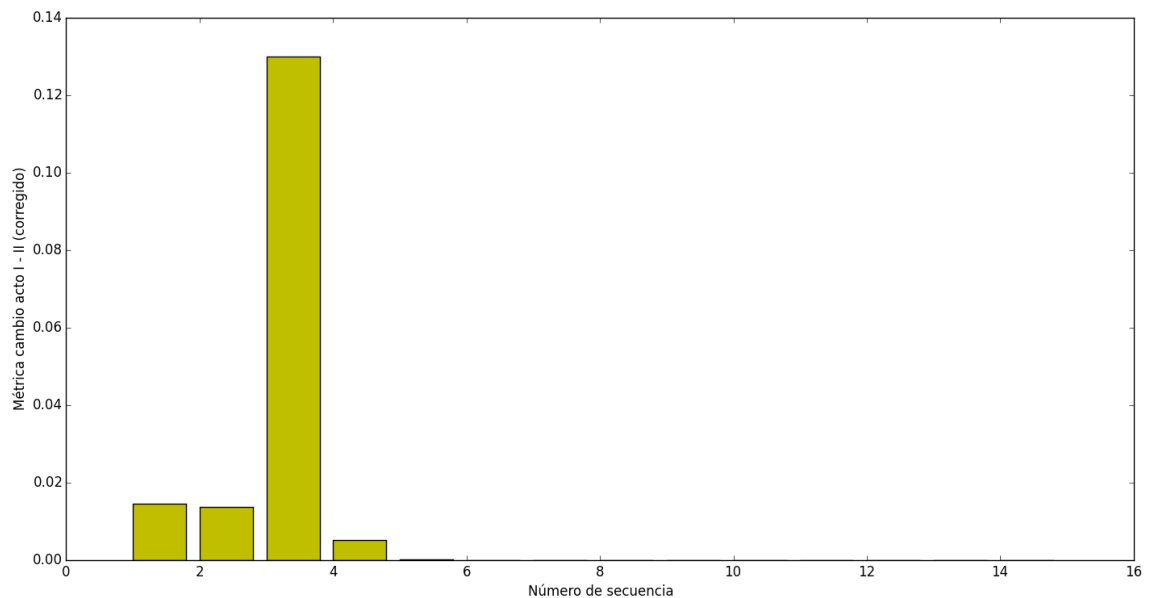


Figura Q.6.1.1: Probabilidad de cambio del I al II Acto con corrección teórica para “Para Elías”

Cambio de I a II ACTO en la secuencia 4, con valor 3.9219.

Líneas 245 a 289.

Cambio teórico en la secuencia 3 +- 1.

Q.6.2 Cambio Acto II – III

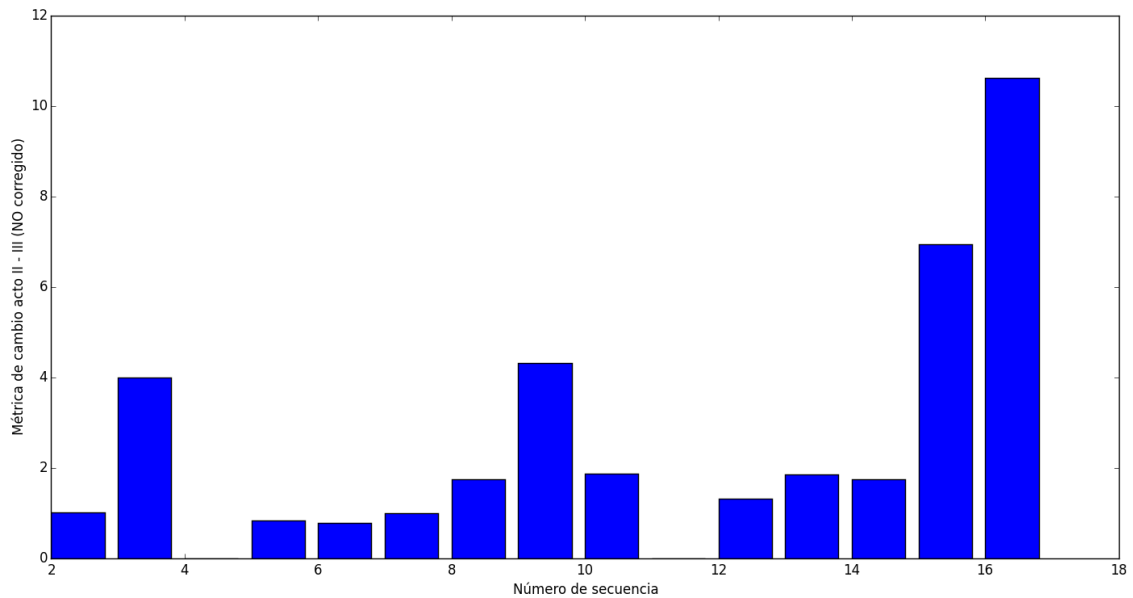


Figura Q.6.2.1: Probabilidad de cambio del II al III Acto sin corrección teórica para “Para Elías”

Cambio de II a III ACTO en la secuencia 16, con valor 10.625.

Líneas 583 a 672.

Cambio teórico en la secuencia 12 +/- 0.

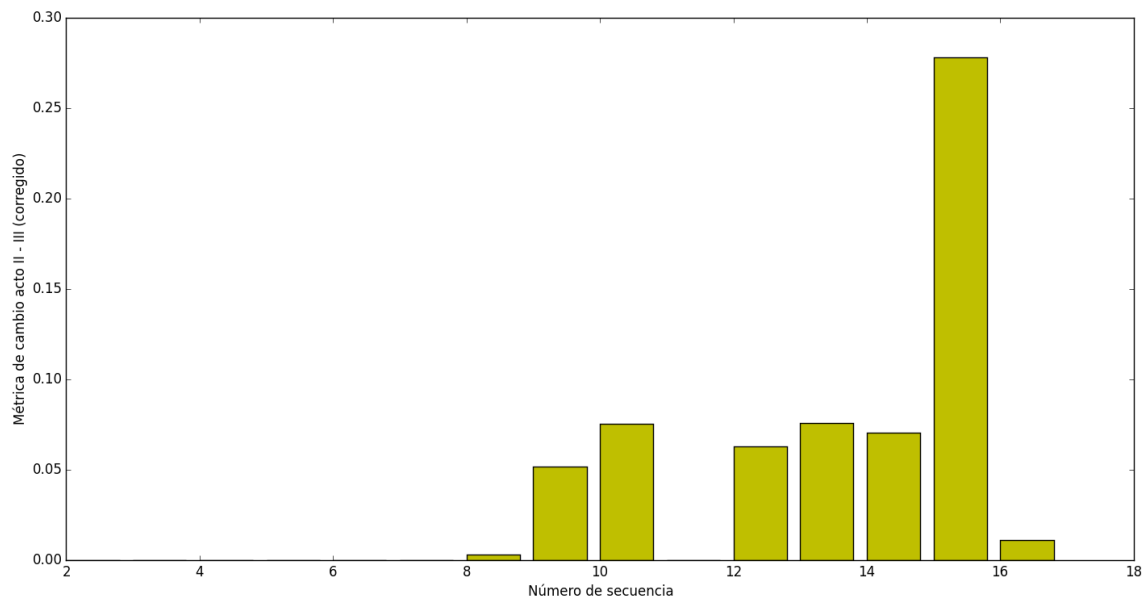


Figura Q6.2.2: Probabilidad de cambio del II al III Acto con corrección teórica para “Tesis”

Cambio de II a III ACTO en la 16, con valor 11.25.

Líneas 583 a 672.

Cambio teórico en la secuencia 12 +/- 0.

Q.7 Guión literario

Se añade a continuación el guión literario de “Para Elías”. Aparte de servir como ejemplo real de una posible entrada del programa desarrollado, el propósito de incluirlo en la memoria es el de poder valorar de primera mano los resultados que arroja el programa teniendo el guión original a la vista. Además, al ser un cortometraje, la lectura tanto del texto como de los resultados es mucho más fácil y manejable.

PARA
ELÍAS

por
Juan López de la Osa

1 INT. SALÓN - DÍA

Montaje de un vídeo casero. Imagen granulada y un poco quemada por el exceso de luz. MARÍA (35), la madre, y ELÍAS (5), el hijo, están sentados frente al piano cada uno en un taburete. Ella intenta enseñarle a él algunas notas, pero el chico anda despistado mirando a la cámara. Su padre, que es el que está grabando la escena, habla desde detrás de la cámara.

PADRE (V.O.)
¡Elías, Elías! Haz caso a mamá.

MARÍA llama cuidadosamente la atención de su hijo sobre el piano y comienza a tocar las primeras notas de LA MELODÍA. ELÍAS da un manotazo sobre el piano, cortando abruptamente la preciosa MELODÍA.

MARÍA
(Suavemente)
Hay que tratar bien al piano,
cariño. Mira...

Coge las manos del niño y las coloca sobre el piano. Y luego coloca las suyas encima. Escuchamos de nuevo comenzar LA MELODÍA, pero esta vez más lentamente. Una nota... otra... y otra más...

PADRE (V.O.)
(En tono cómplice)
Va a ser tan bueno o más que su
madre.

MARÍA se gira sonriente hacia su marido. ELÍAS no despega la vista del piano.

Las notas de LA MELODÍA siguen sonando lenta pero poderosamente...

FUNDIDO A NEGRO

2 INT. HABITACIÓN DE ELÍAS - DÍA

...sigue LA MELODÍA desde la anterior secuencia.

ELÍAS (12) está sentado frente a un teclado con los cascos puestos y los ojos rojos. A su lado una silla vacía.

CONTINÚA:

(CONTINÚA)

Está tocando LA MELODÍA. Aunque la música que se escucha no proviene directamente del teclado, el movimiento de sus manos indica que eso mismo es también lo que él está tocando.

El PADRE entra de pronto en el cuarto y apaga el teclado. LA MELODÍA deja de sonar.

ELÍAS deja de tocar.

El PADRE se va.

ELÍAS, con el teclado todavía apagado, vuelve a tocar con pesadez. Tan sólo escuchamos el sonido mecánico de la pulsación de las teclas, como si fuera una máquina de escribir.

Detrás de él, frente a la ventana, descansan dos fotografías. En una de ellas se ve a una MARÍA joven y sonriente a punto de comenzar a cantar sobre un escenario. La otra es una foto familiar: MARÍA sonríe en el centro, con su marido agarrado a ella a un lado, y su hijo, dándole la mano, al otro.

Cae la lluvia con fuerza sobre la ventana.

FUNDIDO A NEGRO

3 INT. HABITACIÓN DE ELÍAS - DÍA

ELÍAS (23) está en el ordenador con los cascos puestos.

Llaman a la puerta del cuarto.
ELÍAS no responde.

Llaman otra vez más fuerte.

PADRE (V.O.)
¡Elías! ¡Al teléfono!

ELÍAS se quita por fin los cascos.

ELÍAS
¿Qué?

PADRE (V.O.)
Es tu tía. Para felicitarte.

CONTINÚA:

(CONTINÚA)

ELÍAS

Vale. Voy.

ELÍAS se levanta y sale del cuarto.

4 INT. COCINA - DÍA

El PADRE se sienta a la mesa para retomar su plato. Junto a él hay otro plato de comida que parece no haber sido tocado todavía.

Al lado de la puerta de la cocina hay un teléfono. ELÍAS entra y lo coge.

ELÍAS

(TÍA)

Hola.

(¡Hola! ¡Felicidades!)

¡Gracias!

(Ya eres enfermero, eh, ¡qué bien!)

Sí, sí, muy bien.

(Contento, ¿no?)

Sí.

(Qué bien.)

Sí.

(Ahora a celebrarlo, ¿no? Te irás con los amigos y las amigas por ahí a...)

Sí, a ver si hacemos algo.

(Bueno, que...)

Bue...

(...me alegro un montón, eh cariño.

Te voy a dejar que hablo desde un móvil que no es mío y ya sabes que sale muy caro desde aquí. Hablamos a la vuelta ¿vale?)

Bueno, vale.

(Un beso)

Un beso, tía. Adiós.

ELÍAS cuelga el teléfono y se sienta a comer.

PADRE

¿Qué decía la hippie de tu tía?

ELÍAS

Sólo era para darme la enhorabuena.

CONTINÚA:

(CONTINÚA)

PADRE

(Como para sí mismo)

Ya, y de paso restregarnos que esta vez se ha ido a Tailandia. Ya me gustaría a mí también pegarme la mitad del año viajando por el mundo. Pero claro, mejor sin responsabilidades.

ELÍAS

Sólo me ha dicho felicidades.

Silencio.

PADRE

Me ha dicho el jefe de enfermeros de la planta que tienen un contrato para ti para el verano.

ELÍAS deja de comer y se gira hacia su padre. No parece muy contento a primera vista.

ELÍAS

¿Qué me has cogido unas prácticas?

Su PADRE le mira satisfecho con una sonrisa, indiferente a cualquier indicio de enfado.

ELÍAS

Te dije que no me las cogieras tú, que este era el último verano libre que iba a tener y que quería buscármelo por mi cuenta. ¡Voy a ser el enchufado!

PADRE

Nadie te va a tratar así.

ELÍAS

(Sarcástico)

¿Les vas a obligar a que piensen que no soy un enchufado? Eso va justamente...

PADRE

Es sólo un contrato de verano, no es para tanto. Tranquilo.

CONTINÚA:

(CONTINÚA)

ELÍAS

Además quería planificar algo en verano...

PADRE

¿Planificar el qué? ¿Te vas una semana a la playa con tus amigos y luego qué haces el resto del verano? Te pones a trabajar que eso es lo que tienes que hacer, que luego no haces nada. ¡Encima que te ayudo y que eres tú el que tendría que habérselas buscado!

ELÍAS

¡Y lo iba a hacer, pero ahora no!
¡Te lo dije!

PADRE

¡¿Cuándo entonces?! Siempre igual. Luego no haces nada.

ELÍAS

Sí que hago cosas. He acabado la carrera, ¿no? Y...

(Duda un segundo si seguir hablando)

...Y también sé tocar el piano y...

PADRE

(Respondiendo rápidamente)

¡¿Ah, sí?! ¿Y vas a trabajar de tocar el piano?

La respuesta de ELÍAS se superpone a la del padre, que va perdiendo la calma a cada frase

ELÍAS

¡Pues sí!

PADRE

¿Vas a estudiar toda la vida o qué?

ELÍAS

No, trabajaré.

PADRE

¿Cuándo?

CONTIÚA:

(CONTINÚA)

ELÍAS

¡No lo sé!

PADRE

¡Siempre igual, hijo!

ELÍAS se queda sin respuesta. Su PADRE entiende esto como una victoria dialéctica y se calma, sin embargo, en la cara de ELÍAS puede verse que no es así.

PADRE

¿No te puedes mostrar un poco agradecido por que te haya ayudado con las prácticas?

ELÍAS

No entiendes nada.

ELÍAS se levanta de la mesa y se va con cara triste y decepcionada.

ELÍAS

(Como para sí mismo, su padre no le oye)
No sé para qué he estudiado esta mierda de carrera.

5 INT. PASILLO DEL HOSPITAL / PUESTO DE ENFERMERÍA - DÍA

RAÚL (23) y una ENFERMERA (56) van hablando por el pasillo del hospital. Ella lleva puesto el pijama de hospital, él va vestido de calle con colores vivos y lleva una mochila.

ENFERMERA

...y tienes que estar a las 8 aquí.

RAÚL repite lo que le dice la ENFERMERA como para recordarse a sí mismo los datos importantes.

RAÚL

Vale, aquí es la sala de reunión...

ENFERMERA

Eso es, aquí a las 8 ya cambiado y todo, eh. Nos reunimos con los médicos...

(Bajando la voz)

CONTINÚA:

(CONTINÚA)

Huy, mira, este que viene es el hijo del doctor Murillo. Es un sieso.

Se refiere a ELÍAS (27), que viene andando por el pasillo en sentido contrario a ellos. Tiene el gesto más serio y la mirada más pérdida. Alguno diría que da la sensación de persona "más adulta".

RAÚL
(Tendiendo la mano)
Hola, soy...

ELÍAS
(Sin mucha emoción)
Aquí se entra a las 8, incluso si eres nuevo. Hola.

ELÍAS sigue andando. RAÚL se queda cortado.

ENFERMERA
Ya te lo he avisado. No te preocupes, hijo, no es por ti. A algunos les amarga su trabajo. Vamos.

6 INT. ASCENSOR DEL HOSPITAL - DÍA

ELÍAS entra en el ascensor. Está vacío. Se gira y, al levantar la vista, parece ver a alguien. Aparta rápidamente la cabeza haciéndose el loco. Las puertas se están cerrando.

Una mano para las puertas, que vuelven a abrirse. Es RAÚL.

ELÍAS
Llegas tarde y te vas pronto, eh.

RAÚL
Tienes razón. Menudo primer día. Pero es que tengo un concierto ahora y voy muy justo. Están mis compañeros ahí, que los he dejado solos montándolo todo..

ELÍAS se gira para mirarle pero no llega a decir nada.

CONTINÚA:

(CONTINÚA)

Silencio.

ELÍAS
Así que... te gusta la música...

RAÚL
Sí, sí. Tengo un grupo. Vamos,
tocamos en sitios... va bien la
cosa. ¿Tú también tocas?

ELÍAS
Sí, bueno. Tocaba. Ya no toco.
Tocaba con mi madre.

RAÚL
¿El qué?

ELÍAS
El piano.

Silencio.

RAÚL
¿Por qué no te pasas a vernos?

ELÍAS
No...

RAÚL
¡Seguro que te gusta!

ELÍAS
No, no.

RAÚL
Claro, está cerca. Yo voy
directamente ahora hacia allá.

ELÍAS
No, no, no.

7 INT. BAR - DÍA

Es un bar no muy grande con un pequeño escenario al fondo. RAÚL y su banda están tocando. La gente lo está pasando bien.

CONTINÚA:

(CONTINÚA)

ELÍAS se encuentra entre el público. Se fija en los instrumentos y en cómo tocan RAÚL y sus amigos. Una sonrisa se dibuja en su cara.

8 INT. SALÓN - NOCHE

El PADRE se ha quedado dormido en el sofá con la bata puesta. La lámpara de pie que tiene al lado encendida es la única luz que hay.

Se oye el ruido de la puerta principal abriéndose y luego cerrándose. El PADRE se despierta y se despereza un poco.

ELÍAS entra en el comedor y se queda quieto al ver a su padre en el sofá. Casi se asusta, no se lo esperaba.

PADRE

¿Se puede saber qué haces llegando a estas horas?

ELÍAS

¿Me estabas esperando?

PADRE

(Cada vez más serio y despierto)

¿Pero qué es esto de llegar a estas horas entre semana? Mañana hay que trabajar.

ELÍAS

¡Pues vete a dormir! No tienes que esperarme.

PADRE

Claro que tengo que esperarte. ¿Cómo me voy a ir a dormir si no sé dónde estás?

ELÍAS

Joder papá. Soy mayorcito para volver un poco tarde. No pasa nada por dormir un poco menos.

PADRE

Es que no se trata de eso. Se trata de responsabilizarte de las cosas.

CONTINÚA:

(CONTINÚA)

No puedes aparecer aquí como si nada cuando mañana tienes trabajo.

ELÍAS

¡Pero que estaba con un amigo en un concierto! No es nada.

PADRE

En un concierto... ¡Encima de fiesta!

ELÍAS

¡No! ¡Que no!

PADRE

Habrás bebido y todo, ¿no?

ELÍAS

(Perplejo por la situación)

¡No!

PADRE

Mira es que parece mentira...

ELÍAS no deja terminar la frase a su PADRE y se va del salón.

9 INT. HABITACIÓN DE ELÍAS - DÍA

ELÍAS entra en su cuarto agitado y deja la chaqueta sobre la cama. Retira la sábana que tapa el teclado, coge la foto de su madre en la que sale cantando sobre un escenario y la pone en el atril.

Se pone los cascos y empieza a tocar LA MELODÍA. Comienzan a escucharse las primeras notas. ELÍAS cierra los ojos y respira profundamente. Las notas siguen fluyendo mientras su agitación da paso a la calma.

Al abrir los ojos, posa su mirada en la foto de su madre.

La foto cobra movimiento y su madre retoma la actuación desde el momento en el que quedó inmortalizada, al son de la música que toca su hijo.

Madre e hijo tocando juntos su canción, como cuando él era pequeño. ELÍAS sigue también la letra de la canción cantando por lo bajo, murmurando.

CONTINÚA:

(CONTINÚA)

De pronto, un manotazo del PADRE arrancando los cascos a ELÍAS.

LA MELODÍA se interrumpe bruscamente en mitad del momento más bonito de la canción. Hay un pequeño forcejeo en el que los cascos se desconectan del teclado y oímos el sonido cacofónico de las notas que son tocadas por error al estar peleándose al lado del instrumento.

ELÍAS aparta a su padre del teclado y se queda de pie con el rostro compungido mirándole.

ELÍAS

¡¿Qué haces?!

PADRE

¿Qué haces tocando a estas horas?

¡Vas a despertar a los vecinos!

ELÍAS no responde. Mira a su padre con una mezcla de rabia y dolor.

El PADRE se queda también de pie mirándole. En su expresión hay también dolor, cansancio y vergüenza.

La foto de MARÍA, que sigue de pie en el atril, ya no se mueve. Vuelve a estar sobre el escenario, esperando a empezar su actuación con su dulce sonrisa.

Están ahí los tres "mirándose".

El PADRE aparta la vista de su hijo y la fija en la foto de su mujer. Tras una mueca de dolor, aparta de nuevo la vista y sale del cuarto entrecerrando la puerta detrás de sí.

ELÍAS sigue de pie al lado del teclado con la misma expresión en el rostro. Los labios apretados, el ceño fruncido, los ojos vidriosos.

10 INT. PUESTO DE ENFERMERÍA - DÍA

ELÍAS está ordenando unos informes sin prestar mucha atención. Tiene cara de no haber dormido mucho. RAÚL pasa por su lado.

CONTINÚA:

(CONTINÚA)

RAÚL

Ey, Elías.

(Bromeando)

Vaya, veo que eres de los que necesitan dormir sus 8 horas mínimo, eh.

ELÍAS

No he dormido. He estado pensando...
¿Crees que me dejarían tocar a mí en el bar de ayer?

RAÚL sonríe.

Aparece una ENFERMERA.

ENFERMERA

(Poco amigable)

Murillo, ¿tienes ya los informes?

ELÍAS

Sí, claro. Toma.

ELÍAS coge uno de los montones que estaba ordenando y se los entrega a la ENFERMERA prácticamente sin mirarla. Su atención está puesta en la conversación con RAÚL.

ENFERMERA

De las cosas de cada uno en casa, eh. Aquí a trabajar.

ELÍAS

(Sin prestar atención)

Claro.

La ENFERMERA se va.

RAÚL

(Sonriendo)

Te mantengo al tanto.

11 MONTAJE DE ESCENAS

11 A - INT. COCINA - DÍA

ELÍAS y su PADRE comiendo sin hablarse. ELÍAS termina rápido y se va.

11 B - INT. HABITACIÓN DE ELÍAS - DÍA

ELÍAS sentándose a tocar al teclado.

11 C - INT. SALÓN - DÍA

ELÍAS entra en casa y va directamente a su cuarto a tocar el piano.

11 D - INT. SALÓN - NOCHE

El PADRE leyendo solo.

11 E - INT. HABITACIÓN - NOCHE

Mientras, ELÍAS toca el teclado en su cuarto.
El móvil suena y se enciende la pantalla. Se puede leer un mensaje que ha enviado RAÚL: " Tocas el viernes! ;)"
ELÍAS lee el mensaje con satisfacción.

12 INT. PUESTO DE ENFERMERÍA - DÍA

El PADRE, vestido de médico, está hablando con la ENFERMERA.

PADRE

...vale, y le compruebas eso a la de la 7. Por cierto, ¿has visto a mi hijo?

La ENFERMERA se queda un poco parada.

ENFERMERA

Si ha cambiado el turno para lo del concierto ese que lleva dando mal toda la semana.

El PADRE se queda un poco parado.

PADRE

¿Qué concierto?

La ENFERMERA pone de cara haber hablado más de la cuenta.

PADRE

(Insistiendo)

Dora, ¿qué concierto?

CONTINÚA:

(CONTINÚA)

ENFERMERA

Ay, no lo sé. Pues uno que es aquí en un bar de cerca. No sé nada más, eh, yo no sé nada.

13 INT. BAR - DÍA

El PADRE entra en el bar. Al fondo, sobre el escenario, está ELÍAS. A su lado hay un hombre presentando el número.

El PADRE se sienta en la barra.

PADRE

Una caña.

Entre el PADRE y el escenario hay todo un espacio para la pista de baile que está lleno de gente viendo los diferentes números.

PRESENTADOR

Y ahora, un fuerte aplauso para eh... ¡Elías!

El público aplaude. RAÚL, entre el público, silba y grita más alto que el resto.

ELÍAS

(Visiblemente cortado)

Hola. Me llamo Elías y... voy a tocar una canción que se llama "Para Elías".

Se murmulla alguna broma entre el público, que se agita un poco.

ELÍAS

Es una canción que es muy importante para mí. Me la compuso mi madre y solíamos tocarla juntos. Así que quiero dedicársela a ella.

(Una pausa y retoma el discurso con más fuerza.)

Me gustaría que se sentase en un taburete a mi lado como hacía cuando era pequeño, pero bueno,

CONTINÚA:

(CONTINÚA)

sé que desde donde quiera que esté, sigue estando siempre a mi lado.

Y bueno, a mi padre también, que me gustaría que estuviese aquí pero...

El PADRE levanta la mirada perdida y absorta en la cerveza para mirar a su hijo encima del escenario.

ALGUIEN DEL PÚBLICO

¡Empieza ya, que esto no son los Oscar!

ELÍAS se sienta al piano y toca LA MELODÍA a la vez que también canta.

El PADRE va acercándose desde la barra hasta el escenario pasando entre todo el público a lo largo de toda la canción. RAÚL le ve pasar.

La gente asiente al escuchar la música y al ver la entregada actuación de ELÍAS.

Esta vez escuchamos acabar la canción sin problemas.

ELÍAS acaba y, al acercarse al borde del escenario para agradecer los aplausos, ve a su padre ahí al lado. En ese momento, deja de hacer caso al público y se queda mirando a su padre. ÉSTE le devuelve la mirada emocionado.

ELÍAS baja del escenario a la vez que el PRESENTADOR vuelve a escena.

PRESENTADOR

(Haciendo una broma)

Bueno, parece que Elías ha terminado ya lo suyo y se va tal cual.

Padre e hijo no están prestando ya atención al resto del mundo. ELÍAS se queda parado enfrente de su padre, esperando una reacción por su parte.

PRESENTADOR

Y ahora es el turno deeeeeee...
¡Ranas peludas!

El público grita enfurecido.

CONTINÚA:

(CONTINÚA)

PADRE

(A su hijo, con expresión de total sinceridad.)

Ha sido más bonita aún de lo que la recordaba.

CANTANTE DE "RANAS PELUDAS"

(Dirigiéndose al público)

¡Pues nosotros os lo dedicamos a vosotros! ¡Que sois de puta madre!

ELÍAS cambia su gesto de espera por una media sonrisa de agradecimiento por el reconocimiento mostrado.

ELÍAS

Gracias.

El público grita y aplaude a la vez que el grupo comienza a tocar una música bastante heavy y padre e hijo se funden en un abrazo.

FIN.

