

UNIVERSIDAD DEL VALLE DE GUATEMALA

CC3103 - Procesamiento de Lenguaje Natural

Sección 11

Marcos Antonio Gutierrez Suárez



Proyecto NLP: Rutas de Investigación o Aplicación

Autores

Linda Inés Jiménez Vides, 21169

Diego Alexander Hernández Silvestre, 21270

Mario Antonio Guerra Morales, 21008

Adrian Fulladolsa Palma, 21592

Guatemala, 6 de octubre de 2025

Propuesta inicial

Ruta elegida: Aplicación.

Título tentativo: Modelos antagonistas para la generación y detección de poesía al estilo de autores clásicos del Siglo de Oro español.

1. Objetivo general

El presente proyecto tiene como objetivo diseñar e implementar un sistema de procesamiento de lenguaje natural basado en modelos antagonistas que trabajen de manera complementaria en la que un modelo generador, encargado de producir textos poéticos originales que simulan el estilo, la métrica y el vocabulario característicos de autores clásicos de la literatura española, y un modelo discriminador, destinado a identificar si un poema dado corresponde a un texto auténtico o ha sido generado por el modelo.

A través de esta dinámica competitiva entre ambos modelos, el proyecto busca avanzar en la comprensión y modelado de patrones lingüísticos complejos presentes en obras literarias antiguas, explorando la capacidad de los sistemas de inteligencia artificial para captar características estilísticas profundas y aplicarlas en la generación creativa de texto.

2. Justificación

La generación automática de texto ha alcanzado avances significativos en la última década; sin embargo, reproducir estilos literarios específicos, como aquellos pertenecientes a épocas históricas con estructuras métricas, vocabulario arcaico y características estilísticas particulares de una época pasada, continúa siendo un desafío en el ámbito del procesamiento de lenguaje natural.

Este proyecto busca abordar dicho reto desde un enfoque de procesamiento de lenguaje natural, aplicando el paradigma de modelos antagonistas, inspirado en las redes generativas antagónicas (GAN), al dominio textual. El modelo generador se entrenará para producir poesía que imite el estilo de autores representativos del Siglo de Oro español, mientras que el modelo discriminador aprenderá a distinguir entre textos auténticos y generados. Esta interacción bidireccional promueve la mejora iterativa de ambos modelos, dando lugar a un sistema capaz de aproximarse cada vez más a la producción literaria original.

Más allá del interés técnico, este proyecto posee un valor académico y cultural al contribuir a la preservación, estudio y difusión del patrimonio literario en lengua española, y al mismo tiempo explora aplicaciones potenciales en educación, análisis estilístico asistido por inteligencia artificial y creación literaria automática.

3. Plan tentativo

Semana	Actividades
1	Elaboración de la propuesta inicial y conformación del grupo de trabajo.
2-3	Selección, exploración y preprocesamiento del corpus literario (poemas del Siglo de Oro y textos contemporáneos). Implementación del baseline: un modelo clásico de clasificación (TF-IDF + SVM) para el discriminador y un modelo de generación estadístico basado en n-gramas o prompting sin ajuste fino para el generador.
4-5	Entrenamiento de los modelos avanzados: el generador mediante un modelo Transformer (p. ej., T5 o GPT en español) y el discriminador mediante un modelo de clasificación contextual (p. ej., BETO o BERT fine-tuneado). Documentación detallada de hiperparámetros y procesos de entrenamiento.
6	Evaluación exhaustiva de ambos modelos mediante métricas estándar: BLEU y ROUGE para la generación, y precisión, recall y F1 para la clasificación. Análisis de errores y discusión sobre las limitaciones y posibles mejoras.
7	Elaboración del informe final (6–8 páginas) con introducción, metodología, resultados y discusión. Desarrollo del demo final (notebook interactivo o video demostrativo) y preparación de la presentación final del proyecto.

Referencias

Navarro-Colorado, Borja; Ribes Lafoz, María, and Sánchez, Noelia (2015) "Metrical annotation of a large corpus of Spanish sonnets: representation, scansion and evaluation" 10th edition of the Language Resources and Evaluation Conference 2016 Portorož, Slovenia.