
RESUMEN



EL PROCESAMIENTO DEL LENGUAJE NATURAL (PLN, o NLP por sus siglas en inglés de *Natural Language Processing*) es un campo de investigación interdisciplinar de las Ciencias de la Computación, Lingüística y Reconocimiento de Patrones que estudia, entre otros, el uso del lenguaje natural humano en la interacción Hombre-Máquina. La mayoría de las tareas de investigación del PLN se pueden aplicar para resolver problemas del mundo real. Este es el caso del reconocimiento y la traducción del lenguaje natural, que se pueden utilizar para construir sistemas automáticos para la transcripción y traducción de documentos.

En cuanto a los documentos manuscritos digitalizados, la transcripción se utiliza para facilitar el acceso digital a los contenidos, ya que la simple digitalización de imágenes sólo proporciona, en la mayoría de los casos, la búsqueda por imagen y no por contenidos lingüísticos (palabras clave, expresiones, categorías sintácticas o semánticas). La transcripción es aún más importante en el caso de los manuscritos históricos, ya que la mayoría de estos documentos son únicos y la preservación de su contenido es crucial por razones culturales e históricas.

La transcripción de manuscritos históricos suele ser realizada por paleógrafos, que son personas expertas en escritura y vocabulario antiguos. Recientemente, los sistemas de Reconocimiento de Escritura (RES, o HTR por sus siglas en inglés de *Handwritten Text Recognition*) se han convertido en una herramienta común para ayudar a los paleógrafos en su tarea, la cual proporciona un borrador de la transcripción que los paleógrafos pueden corregir con métodos más o menos sofisticados. Este borrador de transcripción es útil cuando presenta una tasa de error suficientemente reducida para que el proceso de corrección sea más cómodo que una completa transcripción desde cero. Por lo tanto, la obtención de un borrador de transcripción con una baja tasa de error es crucial para que esta tecnología de PLN sea incorporada en el proceso de transcripción.

El trabajo descrito en esta tesis se centra en la mejora del borrador de transcripción ofrecido por un sistema RES, con el objetivo de reducir el esfuerzo realizado por los paleógrafos para obtener la transcripción de manuscritos históricos digitalizados. Este problema se enfrenta a partir de tres escenarios diferentes, pero complementarios:

- **Multimodalidad:** El uso de sistemas RES permite a los paleógrafos acelerar el proceso de transcripción manual, ya que son capaces de corregir en un borrador de la transcripción. Otra alternativa es obtener el borrador de la transcripción dictando el contenido a un sistema de Reconocimiento Automático de Habla (RAH, o ASR por sus siglas en inglés de *Automatic Speech Recognition*). Cuando ambas fuentes (imagen y habla) están disponibles, una combinación multimodal de las mismas es posible y se puede realizar un proceso iterativo para refinar la hipótesis final.
- **Interactividad:** El uso de tecnologías asistenciales en el proceso de transcripción permite reducir el tiempo y el esfuerzo humano requeridos para obtener la transcripción correcta, gracias a la cooperación entre el sistema asistencial y el paleógrafo para obtener la transcripción perfecta. La realimentación (*feedback* en inglés) multimodal se puede utilizar en el sistema asistencial para proporcionar otras fuentes de información adicionales con señales que representen la misma secuencia de palabras a transcribir (por ejemplo, una imagen de texto, o la señal de habla del dictado del contenido de dicha imagen de texto), o señales que representen sólo una palabra o carácter a corregir (por ejemplo, una palabra manuscrita mediante una pantalla táctil).

- **Crowdsourcing:** La colaboración distribuida y abierta (*crowdsourcing*) surge como una poderosa herramienta para la transcripción masiva a un costo relativamente bajo, ya que el esfuerzo de supervisión de los paleógrafos puede ser drásticamente reducido. La combinación multimodal permite utilizar el dictado del contenido de líneas de texto manuscrito en una plataforma de *crowdsourcing* multimodal, donde los colaboradores pueden proporcionar las muestras de habla utilizando su propio dispositivo móvil en lugar de usar ordenadores de escritorio o portátiles, lo cual permite ampliar el número de colaboradores reclutables.

Estos escenarios son la motivación de los principales objetivos científicos y tecnológicos:

- Estudiar las técnicas de combinación unimodal y multimodal, con el fin de proponer una nueva técnica de combinación multimodal para mejorar la transcripción de imágenes de textos manuscritos históricos utilizando el dictado de los contenidos de dichas imágenes.
- Estudiar el uso de técnicas de combinación multimodal en un sistema de transcripción asistida por ordenador para acelerar el proceso de transcripción interactivo.
- Desarrollar una plataforma multimodal de *crowdsourcing* para la transcripción de manuscritos históricos basada en las técnicas de combinación multimodal estudiadas.

Las aportaciones al estado del arte de esta tesis se pueden resumir en: la evaluación de cómo combinar la salida de diferentes sistemas de reconocimiento de lenguaje natural, la integración de la combinación de diferentes señales en un sistema de transcripción asistida por ordenador, y el desarrollo de una plataforma multimodal de *crowdsourcing* para la transcripción de manuscritos históricos.