

# Resumen

Esta tesis presenta una serie de contribuciones científicas en el campo del reconocimiento de formas interactivo y multimodal (MISP, del inglés, multimodal interactive structured prediction). El objetivo de MISP es reducir el esfuerzo humano necesario para corregir la salida de un sistema automático, de forma eficaz y ergonómica. La tesis se centra en los dos aspectos principales de MISP: interacción y multimodalidad. El objetivo de la interacción es el estudio de estrategias de colaboración hombre-máquina para producir resultados sin errores. La multimodalidad trata de cómo utilizar modalidades de comunicación con la máquina más ergonómicas que las tradicionales (teclado y ratón).

En la interacción secuencial (de izquierda a derecha), el usuario supervisa la salida automática secuencialmente, de forma que los errores son corregidos en orden de aparición. En esta tesis, el problema de la interacción secuencial se estudia bajo el marco de la Teoría de la Decisión, bajo la cual se define un algoritmo óptimo. Los resultados experimentales sobre varias tareas sugieren que el algoritmo óptimo consigue mejores resultados que el algoritmo estándar.

En comparación con la interacción secuencial, donde el usuario toma la iniciativa, en la interacción activa es el sistema quien decide lo que el usuario tiene que supervisar. Por un lado, se puede reducir el esfuerzo pidiendo al usuario que revise sólo las salidas automáticas susceptibles de tener errores. Para ello, se define una estrategia que pide las salidas con un mayor error esperado en primer lugar. A continuación, se demuestra y valida experimentalmente que esta estrategia es óptima bajo ciertas condiciones. Por otro lado, si el objetivo es reducir el número de correcciones, la interacción activa funciona seleccionando, elemento a elemento, qué elemento de la salida tendría que ser supervisada por el usuario, por ejemplo, una palabra. En este caso, se comparan diversas estrategias. Al contrario que en el caso anterior, la estrategia que funciona mejor aquí es la de escoger el elemento en el que tenemos una confianza mayor, lo que coincide con los resultados del algoritmo óptimo para la interacción secuencial. Esto sugiere que minimizar el esfuerzo y la supervisión son objetivos que van en dirección opuesta.

Con respecto a la multimodalidad, esta tesis profundiza en técnicas para conseguir sistemas multimodales más robustos. Para ello, los sistemas multi-

---

modales se mejoran de forma que puedan aceptar información contextual de la aplicación que se está utilizando. Primero, se estudia cómo integrar el lápiz electrónico en una tarea de traducción automática. Los modelos de lenguaje se mejoran aprovechando la información de la frase origen. Se comparan varias estrategias que se pueden agrupar en las siguientes aproximaciones: una inspirada en los modelos de traducción basados en palabras y otra basada en sistemas de traducción basados en segmentos. Los resultados demuestran una mejoría notable respecto a las estrategias habituales que no utilizan información contextual. La voz también puede ser una modalidad de entrada interesante, dado que deja las manos libres para otras tareas. En este sentido, se prueban experimentos similares a los del lápiz electrónico pero para una interfaz que permite interaccionar con la voz. Las mejoras sobre la aproximación estándar son importantes. Cuando se compara con las estrategias del lápiz electrónico, los modelos basados en frases consiguen mejor rendimiento que modelos basados en palabras. Este hecho es atribuible a que los modelos acústicos están probablemente peor estimados que los modelos morfológicos y, por eso, se benefician más del modelo de lenguaje. Finalmente, se proponen técnicas similares para el dictado de documentos manuscritos. Los resultados demuestran que el reconocimiento de voz y el reconocimiento de textos manuscritos se pueden combinar de una forma eficaz.

Por último, la tesis presenta una evaluación con usuarios reales de un prototipo de traducción interactiva que se compara con un sistema de post-edición de la traducción. Los resultados del estudio revelan que los usuarios son muy sensibles a aspectos de usabilidad de la interfaz de usuario. Por tanto, la usabilidad es un aspecto crucial a considerar en una evaluación humana ya que puede impedir que los usuarios aprecien los beneficios reales de la tecnología. Por suerte, si se corrigen los problemas de usabilidad, el estudio indica que los usuarios prefieren trabajar con el sistema de traducción interactivo a usar el sistema de post-edición.