

# Resum

Aquesta tesi presenta una sèrie de contribucions científiques al camp del reconeixement de formes interactiu i multimodal (MISP, de l'anglès, multimodal interactive structured prediction). L'objectiu de MISP és reduir l'esforç humà necessari per corregir l'eixida d'un sistema automàtic, de manera eficaç i ergonòmica. La tesi se centra en els dos aspectes principals de MISP: interacció i multimodalitat. L'objectiu de la interacció és l'estudi d'estratègies de col·laboració home-màquina per produir resultats sense errors. La multimodalitat tracta de com utilitzar modalitats de comunicació amb la màquina més ergonòmiques que les tradicionals (teclat i ratolí).

En la interacció seqüencial (d'esquerra a dreta), l'usuari supervisa l'eixida automàtica seqüencialment, de manera que els errors són corregits en l'ordre d'aparició. En aquesta tesi, el problema de la interacció seqüencial s'estudia baix el marc de la Teoria de la Decisió, baix la qual es defineix un algorisme òptim. Els resultats experimentals sobre diverses tasques suggereixen que l'algorisme òptim aconsegueix millors resultats que l'algorisme estàndard.

En comparació amb la interacció seqüencial, on l'usuari pren la iniciativa, en la interacció activa és el sistema qui decideix el que l'usuari ha de supervisar. D'una banda, es pot reduir l'esforç demanant a l'usuari que revise només les eixides automàtiques susceptibles de tindre errors. Per a això, es defineix una estratègia que demana, en primer lloc, les eixides amb un major error esperat. A continuació, es demostra i valida experimentalment que aquesta estratègia és òptima baix certes condicions. D'altra banda, si l'objectiu és reduir el nombre de correccions, la interacció activa funciona seleccionant, element a element, quin element de la eixida hauria de ser supervisada per l'usuari, per exemple, una paraula. En aquest cas, es comparen diverses estratègies. Al contrari que en el cas anterior, l'estratègia que funciona millor ací és la d'escollir l'element en el qual tenim una major confiança, cosa que coincideix amb els resultats de l'algorisme òptim per a la interacció seqüencial. Això suggereix que minimitzar l'esforç i la supervisió són objectius que van en direcció oposada.

Respecte a la multimodalitat, aquesta tesi aprofundeix en tècniques per aconseguir sistemes multimodals més robustos. Per això, els sistemes multimodals es milloren de manera que puguin acceptar informació contextual de l'aplicació

---

que s'està utilitzant. Primer, s'estudia com integrar el llapis electrònic en una tasca de traducció automàtica. Els models de llenguatge es milloren aprofitant la informació de la frase origen. Es comparen diverses estratègies que es poden agrupar en les següents aproximacions: una inspirada en els models de traducció basats en paraules i una altra basada en sistemes de traducció basats en segments. Els resultats demostren una millora notable respecte a les estratègies habituals que no utilitzen informació contextual. La veu també pot ser una modalitat d'entrada interessant, atès que deixa les mans lliures per a altres tasques. En aquest sentit, es proven experiments similars als del llapis electrònic però per a una interfície que permet interaccionar amb la veu. Les millores sobre l'aproximació estàndard són importants. Quan es compara amb les estratègies del llapis electrònic, els models basats en frases aconsegueixen millor rendiment que models basats en paraules. Aquest fet és atribuïble a que els models acústics estan probablement pitjor estimats que els models morfològics i, per això, es beneficien més del model de llenguatge. Finalment, es proposen tècniques similars per al dictat de documents manuscrits. Els resultats demostren que el reconeixement de veu i el reconeixement de texts manuscrits es poden combinar d'una manera eficaç.

Finalment, la tesi presenta una avaluació amb usuaris reals d'un prototip de traducció interactiva que es compara amb un sistema de post-edició de la traducció. Els resultats de l'estudi revelen que els usuaris són molt sensibles a aspectes d'usabilitat de la interfície d'usuari. Per tant, la usabilitat és un aspecte crucial a considerar en una avaluació humana ja que pot impedir que els usuaris aprecien els beneficis reals de la tecnologia. Per sort, si es corregeixen els problemes d'usabilitat, l'estudi indica que els usuaris prefereixen treballar amb el sistema de traducció interactiu a utilitzar el sistema de post-edició.