

València, 24 de maig de 2012

## Investigadors de la Universitat Politècnica de València desenvolupen un nou traductor automàtic de llengües minoritàries

- El sistema –en fase de prototip– permet superar les limitacions dels traductors automàtics actuals, generant traduccions de llengües poc utilitzades a partir de combinacions amb altres llengües majoritàries de la mateixa família.  
Per a desenvolupar-lo, els investigadors de la UPV es van inspirar en *Blade Runner*, on els personatges parlaven “Cityspeak”, una barreja d'idiomes que incloïa, entre altres, japonès, espanyol i alemany

Investigadors del grup *Pattern Recognition and Human Language Technologies (PRHLT)*, de l'Institut Tecnològic d'Informàtica de la Universitat Politècnica de València han desenvolupat un nou traductor automàtic de llengües minoritàries. El sistema permet superar l'escassetat de recursos lingüístics dels traductors actuals i facilitar la interpretació dels textos, siga quina siga la llengua en què estiguen escrits, aprofitant la similitud d'aquesta amb altres llenguatges.

“El nostre objectiu era ajudar a entendre aquelles llengües que actualment els traductors no són capaces d'abastar, i contribuir així a la preservació de cultures i l'eliminació de barreres lingüístiques”, destaquen Luis Leiva i Vicent Alabau, responsables del nou mètode de traducció.

Actualment s'estima que més del 10% de les llengües del món no poden ser assistides per cap sistema de traducció, ni tan sols emprant llenguatges intermedis per als quals sol haver-hi més recursos. Per exemple, per a traduir de francès a zulu se sol passar per l'anglès (és a dir, es tradueix de francès a anglès i, posteriorment, d'anglès a zulu). Fins i tot per als llenguatges majoritaris en ocasions és complicat abastar tot el coneixement lingüístic, atès el constant creixement i l'enorme evolució. Segons apunten Luis i Vicent, això dona lloc a un fenomen bastant freqüent en els traductors automàtics: quan el sistema no té informació sobre algunes paraules, les deixa sense traduir. “Per tant, en el cas extrem de no tenir dades sobre cap de les paraules en l'idioma origen, el sistema perd tota utilitat i el missatge no es pot transmetre”.

Per a resoldre aquest dèficit, el sistema ideat per els investigadors del PRHLT-ITI incorpora en les traduccions informació de llenguatges relacionats amb l'idioma de destí, aprofitant la similitud gramatical i sintàctica que hi ha en les famílies de llengües. “La idea consisteix a escollir aquelles paraules d'altres idiomes per a les quals es pugui generar una traducció automàtica, de manera que s'assemblen el més possible a l'idioma de destí, que és l'idioma de l'usuari que vol entendre el missatge. Si bé les frases resultants poden semblar estranyes a primera vista, ja que gramaticalment són una barreja de diversos idiomes, en realitat són prou intel·ligibles per a un parlant de la llengua de destí”, destaca Luis Leiva.

Per a explicar el funcionament i la utilitat del nou traductor, els investigadors de la UPV exposen el cas següent: imaginem que una persona que només parle espanyol vulga entendre la frase “*another label with the same name already exists*”, i que l'espanyol siga un idioma per al qual no hi ha traductors automàtics. En aquest cas el sistema cercaria possibles similituds en l'italià, el portuguès i el francès –totes llengües romàniques– per a les

quals sí que pot generar traduccions automàticament. Construeix en primer lloc traduccions de l'anglès a l'italià, el portuguès i el francès. A continuació, les tres traduccions es combinen estadísticament perquè el resultat siga semblant a l'espanyol.

"Així, la frase anterior es tradueix per "*Un'altra étiquette amb mesmo nome existeix déjà*", que no és espanyol però permet a un espanyol entendre el missatge original sense necessitat de dominar cap dels tres idiomes auxiliars; això hauria sigut impossible de realitzar mitjançant els sistemes de traducció automàtica actuals", apunta Vicent Alabau.

Entre els avantatges que té, aquest nou traductor –en el perfeccionament del qual continuen treballant els investigadors de la UPV– facilitaria l'accés a continguts digitals (llibres electrònics, pàgines web, etc.) que no poden ser traduïts a certes llengües; afavoriria també la integració social de parlants monolingües, i ajudaria els usuaris a adquirir desimboltura i a familiaritzar-se amb el vocabulari d'altres llengües.

### **Inspiració en *Blade Runner***

Per a desenvolupar-lo, els investigadors de la UPV es van inspirar en la pel·lícula *Blade Runner*, en la qual alguns personatges utilitzaven un argot anomenat "Cityspeak" ("Interlingua" en la versió espanyola de la pel·lícula), una barreja d'idiomes, que incloïa, entre altres, japonès, espanyol i alemany. Aquest argot sorgia a Los Angeles com a conseqüència d'un creixent entorn multicultural. "Diàriament molta gent incorpora paraules estrangeres en les converses, per familiaritat o freqüència d'ús, però també perquè a voltes un altre idioma té un terme més precís per a definir un concepte", expliquen Luis Leiva i Vicent Alabau. "Això ens va portar a pensar que seria bona idea usar llenguatges semblats als d'un grup de parlants per a suplir la falta de recursos en traducció automàtica; de manera que, al contrari que el Cityspeak, el llenguatge barreja no ha de ser après perquè està orientat al llenguatge d'aquests parlants", conclouen els investigadors de la UPV.

---

#### **Dades de contacte:**

Luis Zurano Conches  
Unitat de Comunicació Científica-CTT  
Universitat Politècnica de València  
ciencia@upv.es  
647422347

#### **- Annexos:**

