

Andy Way: "Itzulpen automatikoaren erronka handiena kalitatea da"

Rementeria Argote, Nagore

Elhuyar Zientziaren Komunikazioa

Ingelesek hizkuntza-kontuetan itxi samarrak izateko ospea dute; munduko edozein lekutara joan eta ingelesez komunikatuko direla espero dute. Andy Way multzo horretatik kanpo dago nabarmen. Mundu guztiak hizkuntza bera hitz egitearen ideia erabat zekena iruditzen zaio. "Good morning" esan, eta "egun on" batekin agurtu gaitu; gainerako elkarriketa ingelesez egin behar izan dugu, ezinbestean. Horren gaineko txantxak ezin falta, noski: ingelesa eta euskara automatikoki itzultzeko sistema batek asko erraztuko liguke harremana. Gero, elkarriketaren muinean sartu gara.

Jo dezagun itzulpen automatikoaren jatorrira. Hizkuntza desberdinetako jendea elkarrekin komunikatzeko sortuko zen, noski.

Egia da, bai. Eta neurri handi batean itzulpengintza automatikoa arrazoi politikoengatik ere garatu zen. Garai batean, estatubatuarrek errusiarrek zer esaten zuten jakin nahi zuten, esaterako. Eta, gaur egun, Estatu Batuek diru asko inbertitzen dute arabieraren itzulpen automatikoa garatzen. Alde horretatik, segurtasunarekin lotuta doan kontu bat da.

Beste komunikazio-maila batean, bi lagun arteko komunikazioa dago. Esate baterako, gu bion arteko komunikazioa askoz errazagoa izango litzateke, ni Dublinen nagoenean eta zu hemen Euskal Herrian, zuk niri euskaraz idatzi eta itzulpen automatikoko sistema baten bidez nik ingelesez jasoko banu. Nik ingelesez erantzungo nizuke eta zuk euskaraz jaso.



Dublingo Unibertsitateko irakaslea da Andy Way eta itzulpengintza automatikoaren arloan erreferentziatzeko izena.

N. REMENTERIA

Izan ere, ez du axola ingeles zuzen-zuzena ez bada ere, nik ulertuko dut. Eta gakoa hor dago, euskaraz ez dakien ingeles batentzat hobe da ingeles kaskar bat euskara zuzena baino, eta, alderantziz, ingelesez ez dakien euskaldun batentzat hobe da euskara kaskarra ingeles zuzenena baino. Maila horretan, mezu nagusia jasotzea da garrantzikoena.

Hortaz, itzulpengintza automatikoaren hazkunde-esparruaren oinarria gizabanakoa da. Hizkuntza bera hitz egiten ez duten bi lagun elkarrekin komunikatu ahal izatea. ➔

Internetek egundoko aukera ekarri du komunikazioaren esparruan oro har, itzulpengintza automatikoa ere bai?

Zalantzarik gabe. Interneten badira doako sistemak, Babel Fish, esaterako. Ez dira oso sistema sofistikatuak, baina jendea bere hizkuntzan komunikatzen da horiei esker. Esan bezala, debalde dira, eta ez dago inon izena eman beharrik. Eta egunero milioika aldiz erabiltzen dira, nahiz eta oso kalitate ona ez izan. Beraz, sistema horien kalitatea hobetzen badute, askoz gehiago erabiliko dira.

Kalitateari dagokionez, beste muturrean Europako Batasuna dago. Entzuna dut oso modu eraginkorrean erabiltzen dutela itzulpen automatikoa.

Europar ikaragarritzko aldaketa gertatu da. 80ko hamarkadan bederatzi baino ez ziren Europako Batasuneko hizkuntza ofizialak, eta orain hogeitau baino gehiago dira. Eta Batasunaren begietan hizkuntza horiek guztiak parekoak dira. Beraz, dokumentuak hizkuntza horietara guztietara itzuli behar izaten dituzte, eta 400 hizkuntza-bikote inguru daude.

Har ditzagun letoniera eta greziera. Zenbat itzultzaile dira bi hizkuntza horien arteko itzulpenak egiteko gai? Asko ez. Beraz, ez dago nahiko itzultzaile, eta testu pila bat dituzte itzultzeko. Bada, itzulpen automatikoa oso lagungarri zaie itzultzaile horiei.

Europako Batasunean etxe barrurako bereziki prestatutako itzulpen automatikoko sistema bat dute, SYSTRAN deiturikoa. Ez da dendetan salgai dagoen SYSTRAN sis-

tema bera; barruan erabiltzeko diseinatu eta egokitu zuten, eta itzulpenaren lehenengo zirriborro bat egiteko erabiltzen dute. Ondoren, itzultzaileek editatu eta zuzendu egin behar izaten dute. Izan ere, bezeroei edo jendeari bidaltzeko agiri batek akatsik gabea izan behar du.

Makinak erabiltzen dira azken finean, eta horrek bere alde ona du: 24 orduz egin dezakete lan, eta itzultzaileek ez. Tresna horien abantaila da askoz azkarrago egin dezaketela lan, baina kalitate eskasagoan.

Hortaz, jendeak ez luke ikusi behar giza itzultzailearen lana ordezkatzera datorren jarduera baten gisan; beste edozein tresna bezalakoa da, telefonoa, txigorgailua edo autoa bezalakoa. Laguntza ematen digun tresna bat baino ez da.

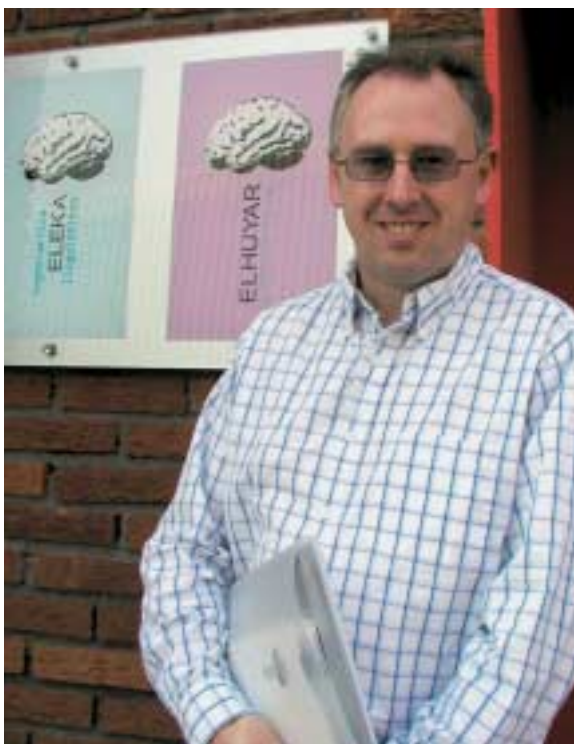
“itzultzaile automatikoen abantaila da askoz azkarrago egin dezaketela lan, baina kalitate eskasagoan”

Mundu osoan daude itzulpen automatikoa hobetzeko lanean diharduten ikerketa-taldeak. Beste talde batzuekin batera lan egitea komeniko zaizue, ezta?

Bai, besteak beste, Euskal Herriko Unibertsitatekoekin harremanetan gaude. Horiek ingelesa-euskara eta gaztelania-euskara bikoteekin egiten dute lan. Eta, pixkanaka, hizkuntza-bikote gehiago batzen goaz. Hala, hainbat hizkuntzarekin lan egiten duten taldeekin dugu harremana, hala nola arabiera, txinera, italiara, frantsesa, alemana, gaztelania, eta, orain, baita euskara ere.

Bestalde, ikasle bat daukagu ingelesaren eta Irlandako zeinu-hizkuntzaren arteko itzulpengintza lantzen, ingurune jakin batzuetarako —adibidez, aireportuetarako—. Izan ere, aireportuetan ez dute informazio guztia idatziz jartzen pantailetan. Hegazkinera igotzeko azken deialdiak eta antzeko abisuak bozgorailuetatik bakarrik ematen dituzte, eta gorrek ez dituzte entzuten. Bada, halako inguruneetan lan egiten ari gara.

Aplikazio-eremua murrizten bada, hizkuntza-bikotea dena dela ere, askoz errazagoa da itzulpena. Hala, aireportuaren esparrura mugatzen bagara, itzulpen orokorragoetan izaten diren arazo gehienak gaindituta daude.



N. REMENTERIA



EHUko Informatika Fakultateko IXA taldearekin, Elekarekin eta Elhuyar Fundazioarekin ari da lanean, besteak beste. Argazkian, Way EHUko Kepa Sarasola irakaslearekin.

N. REMENTERIA

Hizkuntzari dagokionez, arazo handiagoak dituzte hiztun gutxiko hizkuntzek. Zer alde dago hiztun asko eta hiztun gutxi dituzten hizkuntzen artean?

Gaur egun itzulpen automatikoa lan egiten duten iker-tzaile gehienek corpusetan¹ oinarritutako itzulpen automatikoa dihardute. Beraz, corpus bat behar da; batez ere corpus paraleloa behar izaten da, hau da, esaldi honi beste hizkuntzan beste esaldi hori dagokio. Eta hiztun askoko hizkuntza batzuetarako corpus paralelo handia dago. Ingelesaren eta frantsesaren artekoa badago, Kanadako parlamentuko aktak, esate baterako; eta, ingeles-txinerarako, Hong Kong-eko parlamentukoak.

“itzulpen automatikoko teknikek edozein hizkuntza-bikotetarako balio dute berez, baina hiztun gutxiko hizkuntzetarako testu paraleloak falta dira”

Eta ingelesa-euskara bikoterako? Nondik lortu corpusa?

Gaur egun erabiltzen diren itzulpen automatikorako teknikek edozein hizkuntza-bikotetarako balio dute berez, baina, oro har, hiztun gutxiko hizkuntzetarako corpus paraleloak falta dira. Hiztun askoko hizkuntzetan askoz testu gehiago daukagu, eta itzulpen gehiago, gaztelaniaz, ingelesez, frantsesez... Hori da hiztun gutxiko hizkuntzen arazo handiena. Gaelikoaren eta euskararen artean itzulpena egiteko, adibidez, ez daukagu testu paralelorik. Eta hori zailtasun handia da.

Hortaz, zein da gaur egun itzulpen automatikoaren erronka handiena?

Erronka handiena, zalantzarik gabe, kalitatea da, oro har ez baita oso ona oraindik; eta, noski, aipatu berri dugunez, hizkuntza-bikote jakin batzuetarako ezin dela corpus bidezko hurbiltzerik egin, corpusa falta delako. Beraz, itzul-penerako arazoetako bat lehenengo urratsean bertan dago: dauden baliabideak bideratzean.

Nik uste dut itzulpen automatikoaren erronka handienetako bat jendearen etxeetara iristea dela. Izan ere, unibertsitatean oso arazo zailak konpontzen saiatzen gara normalean, baina badira soluzio simple samar batzuk jendearen eguneroko bizitzan lagun dezaketenak, zeinu-hizkuntzan, esaterako.

Eta kaleko jendeak argi dauka itzulpen automatikoa beharrezkoa dela, eta oso erabilgarria. Hizkuntzalaritza konputazionalerako beste arazo batzuk oso zailak dira ulertzeko. Baina edonork daki itzulpengintza zer den eta ordenagailu bat zer den; eta itzulpen automatikoak komunikazioa errazten duela; eta, ondorioz, beharrezkoa dela.

Elkarrizketa amaitzear zegoela ikusita, zerbait gehitu nahi izan zuen Andy Way-k: “Hasieran komunikazioaz hizketan jardun gara. Nik uste dut hizketaren itzulpena laster etorriko dela. Urte batzuk barru, zuk zure hizkuntzan egingo dizkizudazu galderak, euskaraz, eta nik ingelesez entzungo ditut. Hortaz, hizketaren ezagutza (speech recognition) erabiliko dugu, eta zuk euskaraz hizketan entzungo nauzu zure ordenagailuaren bidez. Laster izango da hori, eta aurrerapen handia izango da, hizketa komunikazio idatzia baino askoz naturalagoa baita”.

¹Testu- eta agiri-bilduma bat.