



# Hizkuntzaren teknologiak eta HIZKUNTZAREN INDUSTRIAK

**Informazioaren tratamendu automatikoa da informatika. Hasieran zenbakizko kalkuluak egitera mugatzen zen, baina, pixkanaka, beste era bateko informazio batzuk tratatuz joan da: musika, irudia eta abar. Laster hasi zen, 1950eko hamarkadan, bilakaera horretan giza hizkuntzak prozesatzen ere. Hasiera hartatik 50 urtera, bide aldapatsua egin ondoren, heldutasunera eta merkatura iristen ari dira hizkuntzaren prozesamenduan oinarrituriko aplikazioak, eta horiek lantzen dituzten industria batzuk indarra hartzen ari dira.**

Ederra litzateke japoniera ez dakien euskaldun batek eta euskara ez dakien japoniar batek, esate baterako, telefono bidezko elkarrizketa bat eduki ahal izatea, denbora errealean, nor bere hizkuntzan mintzatuz eta besteak dioena norberaren hizkuntzan entzunez. Hori, jakina, amets hutsa da gaur egun, baina horrantz urrats txiki-txikiak egiten ari garela esan genezake.

Askoz sinpleagoa da idatzizko testuak hizkuntza batetik bestera bihurtzea, baina hori bera ere ez da batere erraza. Gai jakinak (tresnen eskuliburuak, eguraldiragarpinak, etab.) eta hizkuntza-bikote jakinak tratatzen dituzten sistema automatikoak dira gizakien zuzenketa handirik gabe emaitza txukunenak lortzen dituztenak.

Halere, itzulpen automatikoaren egoera asko aldatu da azken urteotan, besteak beste egindako itzulpen asko daudelako gordeta euskarri digitalean. Itzulpen automatikoa gramatika-arauetan oinarritzen zen lehen, baina, orain, datu-base handietan oinarrituriko metodo estatistikoak erabil daitezke. Datu-base horiek jatorrizko testuak eta haien itzulpenak

dauzkate, corpus paraleloak osatuz. Itzulpen-sistemak testuen arteko erlazioak ezagutzen ditu, eta antzekoa edo berdina den beste testu-zati bat aurkezten zaionean, gai da hura itzultzeko. Corpus paralelo handiak dauden neurrian, sistema horrek emaitza onak lortzen ditu.

Erregela bidezko metodoak eta metodo estatistikoak konbinatzeko joera nagusitzen ari da gaur egun.

Maila apalago batean, gaur egun oso lagungarriak gertatzen dira itzulpen-memoria deritzenak. Lehen itzuli izan diren pasarteak gordeta dauzkaten datu-baseak dira horiek. Giza itzultzailea lanean ari delarik, sistemak ikusten badu itzuli beharrezko atal bat edo oso antzeko bat lehendik itzulita dagoela, atal horren baliokideak izan daitezkeen batzuk aurkezten dizkio itzultzaileari, eta honek erabakitzen du horietakoren batek balio duen edo aldaketaren bat eginda erabil dezakeen. Oso mesedegarriak dira sistema horiek, lastertasuna eta kontsistentzia lortzen laguntzen baitute.

Hizkuntza bakarreko testuen lanketan laguntzeko, berriz, hainbat tresna daude:

ortografia-akatsak harrapatzen dituzten zuzentzaile ortografikoak, esaldiak gramatikaren araberakoak diren egiaztatzen duten zuzentzaile gramatikalak, dokumentuetan informazioa aurkitzen laguntzen duten bilatzaileak, dokumentuen laburpena automatikoki egiten duten tresnak, hizkuntzazkoak ez diren datuetatik abiatuz hizkuntzazko azalpenak ematen dituztenak (meteorologiako datuetatik iragarpenak, esate baterako).

Gure artean oso ezagunak dira eskaner bidez jasotako testu inprimatuak interpretatzeko programa informatikoak (OCR).

***Inbertsio handiak egin behar dira ikerkuntzan, eta horiek berreskuratzeko aukera handirik eskaintzen ez duten hizkuntzak atzera geratzen ari dira.***

Besterik da eskuizkribuak interpretatzea, hori askoz bihurriagoa baita.

Ahozko hizkuntza tartean denean, oso zailtasun desberdina dute ahozko hizkuntza ulertzeak eta ahozko hizkuntza sortzeak. Gaur egun oso arruntak dira idatzizko testuak ozenki “irakurtzen” dituzten sistemak —esate baterako, idatzizko testuek diotena itsuei entzutea ahalbidetzen dietenak—, baina alderantzizkoa egitea, gizaki batek ahoz dioena sistema batek automatikoki interpretatzea, eta testu idatzi bihurtzea, adibidez, askoz zailagoa da.

Aplikazio horiek guztiak pixkanaka laborategietatik irten eta merkatura doa. Ez hizkuntza guztietan, zoritxarrez. Izan ere, inbertsio handiak egin behar dira ikerkuntzan, eta horiek berreskuratzeko aukera handirik eskaintzen ez duten hizkuntzak atzera geratzen ari dira. Esan gabe doa ingelesa dela nagusi arlo honetan ere, bera delako interes ekonomikoei lotuen dagoena.

Osasun-zerbitzuetan asko ari dira sartzen hizkuntza-teknologiak. Osasun-arloan, orain arte, informazio klinikoak egituratu gabeko testu-masa handietan gorde izan da. Hizkuntza-teknologiari esker, osasun-arloko profesionalek denbora asko aurrez dezakete, eta, gainera, segurtasuna areagotu. Orain arte testu librean idazten zen informazio klinikoaren ordez, diagnostikoen, tratamenduen eta botiken deskribapen estandarizatuak erabiltzeko joera dago, horretarako sistema bereziak baliatuz.

Beste sektore batzuetan ere ari da sartzen hizkuntza-teknologia, esate baterako automobilgintzan, hegazkingintzan eta nazioarteko erakundeetan. Sektore horietan, globalizazioaren eraginez, hizkuntza askotan idatzitako dokumentazioa sortu behar dute, eta kultura eta hizkuntza askotako langileak trebatu behar dituzte.

Hori guztia egiteko, ezinbestekoak dituzte itzulpen automatikoa, terminologia erazteko eta kudeatzeko softwarea, zuzentzaile ortografikoak, dokumentazio eleaniztunaren kudeaketa, eta abar, denbora aurreztu eta emaitzen kontsistentzia bermatu nahi badute. ●



BRAM JANSSENS/123RF