

# EZ BURUA HAUTSI, MATXIN!

UXOA IÑURRIETA URMENETA  
ITZIAR ADURIZ AGIRRE  
ARANTZA DÍAZ DE ILARRAZA SÁNCHEZ  
GORKA LABAKA INTXAUSPE  
KEPA SARASOLA GABIOLA  
EHUko IXA taldeko ikertzaileak

**G**izarte elebidunetan bizi garenontzat, oso ohikoak dira hizkuntza batean pentsatu eta bestean hitz egiteagatik egingandako hanka-sartzeak. Gutako asko gorrituko ginen noizbait, adibidez, txikitari gurasoak barrez ikusita, eskerrak gaztelaniaz eman zizkigunari ¡No por eso! erantzun geniolako lasai-lasai. Kalean erori den haurrari norbait *Min egin zara?* galdetuko balio ere, euskaldun zahar gehienok ez ginateke hainbeste harrutuko, esaldia arrotza iruditu arren, berehala pentsatuko baikenuke pertsona hori, ziur asko, euskaldun berria dela. Izan ere, hizkuntza bat baino gehiago hitz egiten ditugunok badakigu, esperientziaz, batean ikasitakoak ez duela beti besterako balio izaten: ez horregatik ez da *no por eso* gaztelaniaz, baizik eta *de nada*; eta *te has hecho daño*, berriaz, ez da *min egin zara* euskaraz, baizik eta *min hartu duzu*.

Horrelakoetan, irakasleek, gurasoek edo lagunek zuzentzen dizkigute akatsak, hurrengorako forma zuzena zein den ikas dezagun. Bada, hemen ere ikasle-irakasle ba-

tzuk izango ditugu hizpide, baina ez nolana hikoak. Ikasleak hamaika urte pasatxo ditu, Matxin du izena, eta ez du hezurrik ez haragirik; itzultzaile automatiko bat da. Arau-mordoxka bat erabiltzen du hizkuntzak ikasi eta erdaraz irakurritakoa euskaratzeko, baina askotan ematen dizkiote arau horietatik kanpo geratzen diren itzulgaiak ere, eta irakaslearen lana horrelakoetan laguntzea da, ez dezan *burua hautsi*, ez dezan hain sarri *hanka sartu*.

## MATXIN, ITZULTZAILE

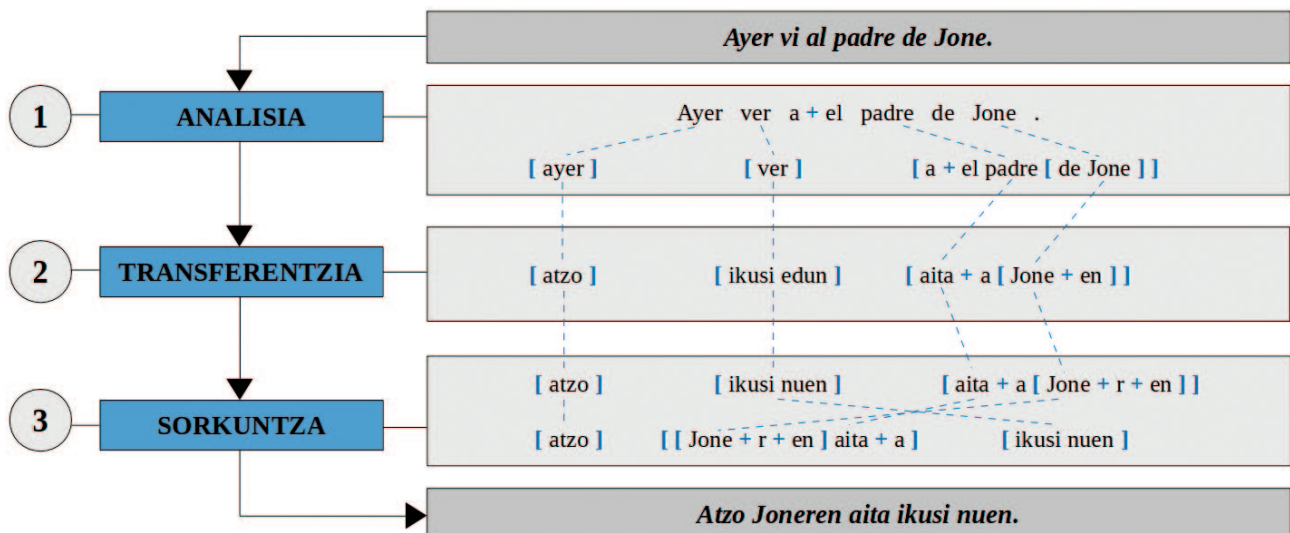
Matxinek, gaur egun, gaztelaniatik eta ingelestik euskarara itzultzen du. Gramatika-arau sorta bat eta bi hiztegi elebidun ditu oinarrian, eta horietatik lortzen du informazioa testuak hizkuntza batetik bestera pasatzeko. Hiru fasetan egiten du bere lana: analisisa, transferentzia eta sorkuntza (Mayor et al., 2009).

Lehen irudian ikus daitekeenez, lehenik, gaztelaniazko —edo ingelesezko— testua analizatzen du, morfologikoki eta sintakti-

koki: hitz bakoitzaren lema (adibidez: *vi* → *ver*), kategoria (aditza, izena, adberbioa...), funtzio sintaktikoa (subjektua, objektu zuzena, modifikatzailea...), eta beste hainbat ezaugarri. Ondoren, transferentzia-fasean, jatorrizko esaldiko hitzei ordainak ematen dizkie, eta informazio gramatikala egokitzen du. Eta, azkenik, sorkuntza-fasean, euskarazko testua sortzen du, transferentzian lortutako informazioa baliatuta: lema bakoitzari dagokion forma sortu (adibidez: *ikusi edun* → *ikusi nuen*), eta hitzen ordena egokitzen du.

Hala ere, ezin dira itzulgai guztiak modu egokian itzuli arau eta hiztegi orokor horien bidez, eta horrek arazoak ematen dizkio Matxini. Ikusi, bestela, zer-nolako esaldiak sortzen dituen batzuetan:

- (1) ES: **Surtió un gran efecto.**  
EU (Matxin): **Efektu handi bat hornitu zuen.**  
EU (zuzena): **Eragin handia izan zuen.**



1. irudia. Matxinen itzulpen-prozesuaren adibide bat (2016-02-01).



2. irudia. Lepo hitza duten hiru Unitate Fraseologiko (UF), Antton Olariagak 2016-01-15ean *Berrian* argitaratutako zintan.

### UNITATE FRASEOLOGIKOAK (UF), MATXINEN GRAMATIKA-ARAU OROKORRETATIK KANPO

Hain zuzen ere, badaude hitz-konbinazio batzuk hizkuntzen arau orokorretatik kanpo geratzen direnak, eta horien artean daude Unitate Fraseologikoak (Gurrutxaga, 2016). Corpasek (1997) eta Urizarrek (2011), besteak beste, hiru multzotan sailkatu dituzte:

1. **Enuntziatu fraseologikoak:** beren horretan erabil daitezke, esaldi baten barruan sartu gabe, eta egoera jakin batzuetan bakarrik. Multzo honetakoak dira eskerrik asko eta ez horregatik bezalakoak.
2. **Lokuzioak:** ez dira esaldi osoak, eta konbinazioaren esanahia ez da konbinazioko hitzen esanahien batura. *Esku harturen* esanahia, adibidez, ez dator bat eskuk eta *hartuk* normalean izaten dutenarekin.
3. **Kolokazioak:** hauek ere ez dira esaldi osoak, hitzetako batek behintzat bere esanahia gordetzen du, eta, esanahi jakin bat adierazteko, konbinazioko hitz batekin beste hitz jakin bat erabiltzen da gehienguan, eta ez baliokideak diren beste batzuk. Esate baterako, *arreta jarri* erabiltzen dugu, eta ez *arreta kokatu* edo antzekorik.

Horrez gain, UFak asko aldatzen dira hizkuntza batetik bestera (Sanz, 2015), eta, hortaz, sarritan ez dira itzultzeko errazak izaten. Har ditzagun adibidetzat sailkapenean aipatutakoak:

- Lehenago ere esan dugun bezala, ez horregatik ordaintzat *de nada* erabiltzen dugu gaztelaniaz, eta ez *no por eso*.

<sup>1</sup> <http://ixa2.si.ehu.es/konbitzul>

<sup>2</sup> <http://hiztegiak.elhuyar.eus>

<sup>3</sup> <http://dicesp.com/paginas>

- *Esku hartu* esateko, *tomar parte* edo *participar* erabiltzen da (ez da inon aipatzen *manorik*, *eskurik*).
- *Arreta jarriren* ordaintzat *prestar atención* erabiltzen dugu normalean (*arreta mailegatu*, *alegia*).

Are gehiago, azken adibide hori frantsesera eta ingelesera ere ekarri gero, ikusiko dugu aditzak ez datozela bat ez euskarazkoarekin, ez eta gaztelaniazkoarekin ere: frantsesez, *faire attention* (*arreta egin*); eta ingelesez, *pay attention* (*arreta ordaindu*).

Beraz, horrelakoak ikastea neketsua bada gizakiontzat ere, pentsa zeinen zaila den ordenagailu batentzat automatikoki ondo itzultzea, kontuan izanik oinarrian dituen hiztegiak mugatuak direla eta gramatika-arauak oso orokorrak.

Matxinek darabilen hiztegian, badaude hitz bat baino gehiagoko sarrera batzuk, eta batzuetan itzulpen zuzenak sortzen dira horiei esker:

- (2) ES: *Unos alumnos hicieron la pelota a la profesora.*  
EU (Matxin): *Ikasle batzuek irakaslea lausengatu zuten.*
- (3) ES: *Acabo de fregar los platos.*  
EU (Matxin): *Harrikoa egin berri dut.*

Tamalez, ez dira sarrera asko, eta ez dira beti ondo erabiltzen. Begira, esaterako, zer gertatzen den 2. eta 3. adibideetako itzulgaiak pixka bat aldatuz gero:

- (4) ES: *La profesora no podía crear la pelota que le estaban haciendo los alumnos.*  
EU (Matxin): *Irakasleak ikasleak egiten ari zitzaizkion pilota ez zuen sinetsi ahal.*  
EU (zuzena): *Irakasleak ezin zuen sinetsi nola lausengatzen zuten ikasleak.*

- (5) ES: *Acabo de fregar el suelo.*

EU (Matxin): *Lurzoria harrikoa egin berri dut.*

EU (zuzena): *Zorua garbitu berri dut.*

Matxinek bi zailtasun nagusi izaten ditu: batetik, gaztelaniazko edo ingelesezko UFak ezagutzea, eta, bestetik, euskarara ekartzea. Gaur egun, beti segidan eta ordena berean dauden hitz-konbinazioak bakarrik detektatzen dira ondo; beraz, konbinazioko hitzak bereizten baditugu edo ordena aldatzen badiegu, hitz solteak balira bezala tratatzen dira, eta ez UF baten parte balira bezala (4. adibidea). Era berean, euskarazko esaldiak sortzeko ere informazioa falta da askotan, sarrera bakoitzari ordain bakarra ematen baitzaio Matxinen hiztegian. Horregatik itzuli da oker 5. adibideko esaldia, ez baita kontuan hartu *fregar* aditzari, *el suelo* izen-sintagmarekin batera doanean, *garbitu* ordaina dagokiola, eta ez *harrikoa egin*.

### KONBITZUL, MATXINEN IRAKASLE BERRIA

Hortaz, orain arteko adibideek agerian uzten duten moduan, Matxinek UFak zuzen itzuliko baditu, behar-beharrezkoa du laguntza, eta horrexetarako sortu da Konbitzul<sup>1</sup>, izen+aditz konbinazioak itzultzen irakatsiko dion tresna.

Konbitzul datu-base publikoa da, eta azterketa linguistiko batetik lortutako informazioa jasotzen du. Izenez eta aditzez osatutako hitz-konbinazioen —eta haien ordainen— ezaugarriei buruzko datuak gordezen ditu, oraingoz gaztelania-euskara hizkuntza-bikotean. Hiru iturritatik bildu dira azterketa horretako izen+aditz konbinazioak: Elhuyar hiztegi elebidunetik<sup>2</sup>, eskuz egindako itzulpen-multzo erraldoietatik eta gaztelaniazko kolokazioen DiCE hiztegitik<sup>3</sup> (Alonso, 2004).



Landutako informazio gehiena eskuragarri dago Interneten, eta falta dena ere laster egongo da erabiltzaileen eskura. Izan ere, datu-basearen interfazea bilatzaile itxurakoa da, eta edozein erabiltzailek egin ditzake bilaketak oso erraz: bilatu nahi dena idatzi, eta idatzitakoarekin bat datozen konbinazioen zerrenda bat erakusten da, ordainekin batera. Ondoren, ordainetan klikatuz gero, informazio linguistikoko gehiago ere ikus daiteke (3. irudia).

Dena den, lehenago esan bezala, Konbitzulen zeregin nagusia Matxini laguntzea da, bi erronkari aurre egiteko: batetik, sorburu-hizkuntzako UFak ezagutzeko, eta, bestetik, euskaratzeko. Demagun honako esaldi hauek ematen dizkiotela:

- (6) El tema **despertó interés** en los oyentes.
- (7) El tema **despertó un gran interés** en los oyentes.
- (8) Quedó claro el gran **interés** que el tema **despierta** en los oyentes.

Hiru adibideotan ageri den UFa —*despertar interés*— ez dago Matxinen hiztegiaren oraindik; beraz, gaur egun, ez du hitz-konbinazio hori UFTzat tratatzen (4. irudia). Baina, Matxinen hiztegiaren egongo balitz ere, orain arte erabili izan duen metodoaren bidez, 6. adibidean bakarrik ezagutuko luke, esaldi horretan bakarrik agertzen baitira bi hitzak ordena berean eta tartean beste elementurik gabe.

Konbitzulen laguntzaz, ordea, jakingo du *despertar interés* UF bat dela, eta, gainera, konbinazio malgua dela, hau da:

- Konbinazioko bi osagaien artean beste hitz batzuk ere ager daitezkeela.
- Hitzen ordena aldakorra dela.

Hala, sorburu-hizkuntzako esaldia analizatzean, informazio hori guztia kontuan hartu eta gai izango da 7. eta 8. adibideetan ere UF bat dagoela jakiteko. Izan ere, esperimentu batek erakutsi duenez, datu-baseko informazioari esker, lehenago metodoarekin baino ia % 30 UF gehiago ezagutzen dira.

Bestalde, behin konbinazioak detektatutakoan, euskarara ekarri behar dira, eta horretarako ere informazio gehigarria beharko du Matxinek. Oraingoan ere, Konbitzulek argituko dizkio zalantzak:

- *Despertar* aditzaren ordaintzat, erabili piztu (eta ez esnatu).

epea amaitu		
	vencer el plazo	epea amaitu
Marka/egitura morfo.	adi + det + ize	ize (abs) + adi
Mugatasuna/numeroa	s	s
Baliokidetzak	Izenak ordaintzat hiztegiaren? BAI Aditzak ordaintzat hiztegiaren? EZ	
Iturria	Elhuyar eu > es	
epea bete +		
epea bukatu +		

3. irudia. Konbitzulen interfazearen itxura.

**matxin** Matxin 2.0 - Euskarazko itzultzaile automatikoaren bertsio berria

Testuen itzulpena Gaztelania - Euskara Itzuli

El tema despertó interés en los oyentes.

Gaiak interesa esnatu zuen entzuleetan.

Entzun emailza gure ahots sintesiko sistema erabiliz

4. irudia. Matxinen interfazea eta adibide bat (2016-02-02).

- *Interés* izenaren ordaintzat, erabili *interes*, eta jarri mugatuan.

Horrela, *Interes handi bat esnatu zuen* bezalako esaldiak sortu beharrean, *Interes handia piztu zuen* bezalakoak sortu ahal izango ditu. Bigarren ataza honetarako informazioa ez da oraindik sistemari integratu, baina azterketa linguistikoa eginda dago; beraz, pentsatzekoa da emaitzak laster ikusi ahal izango ditugula sarean.

#### JAKIN-MINA ASETZEKO, ZAKUA BETETZEN

Dena dela, lana ez da hor bukatzen, noski, Matxin jakin-min handiko ikaslea baita! Hurrengo pausoa ingelesezko UFak itzultzeko informazioa biltzea izango da, eta, hortik aurrera ere, Konbitzulek datuak biltzen jarraitu beharko du, pixkanaka zakua betetzen, ikaslea, hazi ahala, gero eta hobe izan da-din itzultzaile-lanetan. ●

#### BIBLIOGRAFIA

- CORPAS, G.: *Manual de fraseología española*. Editorial Gredos (1997).
- GURRUTXAGA, A.; ALEGRIA, I. & ARTOLA, X.: "Idiomatikotasunaren karakterizazio automatikoa: izena+ aditza konbinazioak". *EKAIA Euskal Herriko Unibertsitateko Zientzi eta Teknologia Aldizkaria* (2016).
- MAYOR, A.; ALEGRIA, I.; DÍAZ DE ILARRAZA, A.; LABAKA, G.; LERSUNDI, M. & SARASOLA, K.: "Matxin, euskararako lehenengo itzultzaile automatikoa". *Senez: itzulpen aldizkaria*, 37 (2009), 197-220.
- SANZ, Z. (2015): "Alemanetik euskaratutako unitate fraseologikoen itzulpen-azterketa. Tesiaren nondik-norakoak". *Senez: itzulpen aldizkaria*, 46 (2009), 211-230.
- URIZAR, R.; ALEGRIA, I.; ODRIOZOLA, J. C. & EZEIZA, N.: "Euskarazko hitz anitzeko unitate lexikalen tratamendu konputazionala". *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*, 43(1-2) (2011), 891-908.