

Itzulpen Automatikorako gaztelania-euskara patroiak: lehen urratsak

Uxoia Iñurrieta, Nora Aranberri, Arantza Diaz de Ilarraza, Kepa Sarasola

Laburpena

Lan honetan, adibideetan oinarritutako patroi batzuk sortu ditugu, erregeletan oinarritutako itzulpen-sistema automatiko bat hobetzeko asmoz. Patroirik erabilgarrienak emango zituzten adibideak bakarrik hartzeko, euren erabilera-maiztasunari eta itzulpen automatikoen egokitasunari erreparatu diegu. Ondoren, adibideetako entitate-izenak eta zenbakiak orokortu ditugu, elementu horiek aldatuta ere, patroiak erabili ahal izateko.

1 Sarrera

Erregeletan oinarritutako itzultzaile automatikoen, batzuetan, akatsak egiten dituzte nahiko ohikoak diren esaldiak itzultzean. Izan ere, egitura batzuk ezin dira hitzez hitz itzuli, eta, hortaz, arau gramatikalak ez dira nahikoa emaitza egokiak lortzeko. Guk hobetu nahi dugun sistemak ere, *Matxinek*, arazo hori du, eta honelako itzulpen xelebreak sortzen ditu sarritan:

Gaztelaniazko adibidea: *¡viva Gales!*

Itzulpen automatikoa: *Gales biziduna!*

Horrelako akatsak zuzendu ahal izateko, eskuz egindako itzulpenak adibidetzat hartu, eta patroi batzuk sortu ditugu, itzultzaile automatikoan integrazteko asmoz. Gure helburua da patroiaren bidez itzul daitezkeen segidak detektatzea, itzulgaia analizatzen hasi aurretik; hala, sistemak jakingo du

zati horiek ez dituela erregelen bidez itzuli behar, eta sorburu-hizkuntzako patrioientzat definitutako euskarazko baliokideak jarriko ditu zuzenean.

Bestetik, patrioiak ahalik eta gehien erabili ahal izateko, izen bereziak eta zenbakiak aldagai gisa erabiltzea erabaki dugu. Era horretan, *¡viva Gales!* ez ezik, ondo itzuliko dira *¡viva Francia!*, *¡viva Alemania!* eta antzekoak ere. Honelako itxurako patrioiak sortu ditugu:

Gaztelaniazko patroia: *¡viva ENTI.0!*

Euskarazko patroia: *Gora ENTI.0!*

Hala ere, kontuan hartu behar da patrioiak pisutsuak direla itzulpen-sistementzat eta, hortaz, ez dela komeni gehiegi definitzea. Lehen hurbilpen honetan, adibide sorta bat hartu, eta horietatik erabilgarrienak zein izan litezkeen aztertu dugu, patrioiak sortzeari ekin aurretik.

2 Adibideen aukeraketa

Elhuyar hiztegitik 132 erabilera-adibide erauzi ditugu, eta, horietatik patrioi erabilgarriak sor litezkeen aztertzeke, bi ezaugarriari erreparatu diegu: sistema hobetuko luketen ala ez, eta zenbat erabiliko lirartekeen.

2.1 Hiztegiko adibideen eta itzulpen automatikoen arteko alderaketa

Itzulpen-sistema hobetuko duten adibideak hautatzeko, gaztelaniazko esaldiak *Matxinen* bidez itzuli, eta adibideak sailkatu ditugu, itzulpen automatikoaren egokitasunaren arabera. Taula honetan bildu ditugu gure emaitzak:

	Adibideak	Ehunekoa
Hiztegiko itzulpena makinarena baino hobea	68	% 51,51
Bi itzulpenak berdinak, edo biak egokiak	50	% 37,87
Hiztegikoa hobea, baina makinarena ere onargarria	10	% 6,82
Makinaren itzulpena hiztegikoa baino hobea	4	% 3,03
Guztira	132	% 100

Hala, hiztegiko adibideen % 48,48 alde batera utzi ditugu; izan ere, iruditu zaigu euren itzulpen automatikoak onargarriak zirela eta ez lioketela sistemari hobekuntza handirik ekarriko. Beraz, 68 adibide geratu zaizkigu aztertzen jarraitzeko.

2.2 Erabilera-maiztasunaren neurketa

Azterketarekin jarraitzeko, hiztegiko gaztelaniazko esaldiak zenbat erabiltzen diren jakin nahi izan dugu, eta, horretarako, *Google* bilatzaileaz baliatu gara. Esan bezala, izen bereziak eta zenbakiak aldagai gisa markatu ditugu patroieta; beraz, bilaketak egitean, elementu horiek alde batera utzi ditugu. Esaterako, *el canal de Panamá* segidaren kasuan, “*el canal de*” bakarrik bilatu dugu.

Adibide bakoitzak emandako emaitza kopuruaren arabera, sarrien agertu diren adibideak bakarrik aukeratu ditugu: 700.000 emaitzatik gorakoak (adibide guztien % 24,24).

3 Patroien sorkuntza

Hiru entitate-detektatzailearen laguntzaz, patroiak automatikoki sortu ditugu. Ondoren, batetik, lan automatikotik sortutako akatsak zuzendu ditugu eta, bestetik, aldaketa gehiago ere egin ditugu, patroien eraginkortasuna hobetze aldera.

3.1 Entitate-detekzioan egindako akatsen zuzenketa

Gaztelaniaren kasuan, erabili dugun analizatzaileak akats bakarra egin du: ez du izen berezi bat hauteman.

Gaztelaniazko adibidea: *El Tajo desemboca en el **Atlántico***

Patroi automatikoa: *El ENTI_0 desemboca en el **Atlántico***

Patroi zuzendua: *El ENTI_0 desemboca en el **ENTI_1***

Euskararen kasuan, berriz, entitateak hautematean egindako 11 akatsez gain, postposizio-informazioa edo kasua gaizki hauteman da beste 9 segidatan (horietako gehienetan, zenbaki bati kasua edo postposizio-informazioa gehitu zaiolako, eta detektatzaileak ez duelako ezagutu). Adibidez:

Euskarazko adibidea: % *6ko* igoera

Patroi automatikoa: % *ENTI_0ko* igoera

Patroi zuzendua: % *ENTI_0+S_GEL* igoera

3.2 Patroien eraginkortasuna hobetzeko proposamenak

Zuzenketa horiez gain, beste aldaketa batzuk ere egin dizkiogu zenbait patroiri, pixka bat moldatuta erabilgarriagoak izan zitezkeelakoan.

Lau kasutan, iruditu zaigu hitz-segida osoa baino, zati bat bakarrik gehiago erabiliko litzatekeela. Beraz, patroia zati horretara mugatu dugu, baldin eta pentsatu badugu horrela ere itzulpen automatikoaren akatsa zuzenduko lukeela. Hona hemen adibide bat:

Gaztelaniazko adibidea: *finalmente me admitieron en la **universidad de Zaragoza***

Euskarazko adibidea: *azkenean **Zaragozako unibertsitatean onartu ninduten***

Gaztelaniazko patroï proposatua: *universidad de ENTI_0*

Euskarazko patroï proposatua: *ENTI_0+S_GEL unibertsitatea*

Bestalde, beste bi patroï laburtzean, egokitzapen gehigarriren bat ere egin behar izan dugu; adibidez:

Gaztelaniazko adibidea: *la puerta tiene 80 cm de ancho*

Euskarazko adibidea: *atea **80 cm zabal** da*

Gaztelaniazko patroï proposatua: *ENTI_0 cm de ancho*

Euskarazko patroï proposatua: ***ENTI_0 cm-ko zabalera***

Patroïak ongi funtzionatuz gero, *Matxinek* honen antzeko zerbait sortuko luke: *ateak 80 cm-ko zabalera du*. Itzulpena hiztegikoa baino literalagoa da, baina, era horretan, kasu gehiagotan erabili ahal izango da, nahiz eta beste subjektu batekin agertu; esaterako: *la ventana tiene 50 cm de ancho*.

Azkenik, beste bi kasutan, hiztegiko adibidea bera iruditu zaigu aldatzekoa. Batean, hitz bat ez zaigulako oso egokia iruditu, eta bestean, hitz jakin bat erabiltzeak arazoak eman litzakeela pentsatu dugulako. Hona hemen horietako bat:

Gaztelaniazko adibidea: *el jorobado de Notre-Dame*

Euskarazko adibidea: *Notre-Dameko **konkorra***

Gaztelaniazko patroï proposatua: *el jorobado de ENTI_0*

Euskarazko patroï proposatua: *ENTI_0+S_GEL **konkorduna***

Gure ustez, gaztelaniazko *jorobado* hitzaren ordain egokiagoa da *konkordun*, eta ez *konkor*. Izan ere, hiztegiek biak jasotzen badituzte ere, Googlen bilatuta, argi ikusten da *Notre-Dameko konkorduna* erabiliagoa dela (17.900 emaitza) *Notre-Dameko konkorra* baino (5 emaitza), ziur asko, haurrentzako film ezagunaren izenburua hala dagoelako itzulita.

4 Emaitzak

Adibide guztiak aztertu ondoren, hona hemen proposatzen ditugun patroiak biltzen dituen taula:

Gaztelaniazko patroia	Euskarazko patroia
¡viva ENTL0!	gora ENTL0!
con ENTL0 minutos de anticipación	ENTL0 minutu lehenago
el canal de ENTL0	ENTL0+S_GEL kanala
el canal de la ENTL0	ENTL0+S_GEL kanala
el desierto de ENTL0	ENTL0+S_GEL basamortua
el ENTL0 aniversario de la boda	ezkontzaren ENTL0. urtemuga
el ENTL0 desemboca en el ENTL1	ENTL0+S_ABS ENTL1+S_INE itsasoratzen da
el estrecho de ENTL0	ENTL0+S_GEL itsasartea
el jorobado de ENTL0	ENTL0+S_GEL konkorduna
el principio de ENTL0	ENTL0+S_GEN printzipioa
el reino de ENTL0	ENTL0+S_GEL erresuma
ENTL0 cm de ancho	ENTL0 cm-ko zabalera
ENTL0 está en cinta	ENTL0+S_ABS haurdun dago
hora de cierre: ENTL0	itxiera-ordua: ENTL0
la dama de ENTL0	ENTL0+S_GEL dama
la factoría de ENTL0 en ENTL1	ENTL0+S_GEN ENTL1+S_GEL faktoria
la hoguera de ENTL0	ENTL0 sua
la orquesta filarmónica de ENTL0	ENTL0+S_GEL orkestra filarmonikoa
la paga de ENTL0	ENTL0+S_GEL ordainsaria
la selva de ENTL0	ENTL0+S_GEL basoa
la situación en ENTL0	ENTL0+S_GEL egoera
la unidad de los pueblos de ENTL0	ENTL0+S_GEL herrien batasuna
las cataratas de ENTL0	ENTL0+S_GEL ur-jauziak
los menores de ENTL0 años	ENTL0 urtez azpikoak
paso de ENTL0	ENTL0+S_GEL lepoa
reducción de un ENTL0 %	% ENTL0+S_GEL beherapena
un aforo de ENTL0 personas	ENTL0 eserlekuko edukiera

un peso de ENTI_0 kilos	ENTI_0 kiloko pisua
una capacidad de ENTI_0 litros	ENTI_0 litroko edukiera
universidad de ENTI_0	ENTI_0+S_GEL unibertsitatea

5 Ondorioak

Lan honetan, gaztelania-euskara hiztegiko 132 adibide erabili ditugu patroio orokortuak sortzeko. Patroi erabilgarriak bakarrik bildu nahi genituenez, hasteko, adibideen erabilgarritasuna aztertu dugu, bi ezaugarri erreparatuta: ea sistema hobetuko luketen, eta zenbat erabiliko liratekeen. Hiru entitate-detektatzaile erabili ditugu patroiak automatikoki sortzeko, eta, azkenik, hainbat zuzenketa eta egokitzapen eginda, 30 patroio sortu ditugu.

Hurrengo pausoa guk proposatutako patroiak itzulpen automatikoko sisteman integratzea izango da; ondo funtzionatzen ote duten ikusi beharko da, eta, beharbada, egokitzapen batzuk egin beharko zaizkie. Horrez gain, patroietako entitateen mota zehaztea ere komenigarria dela uste dugu, bestela, itzulpen okerrak sor bailitezke kasu batzuetan.

Oro har, ikusi dugu posible dela adibide sorta batetik aukeraketa bat egitea, eta, ondorioz, uste dugu gauza bera egin daitekeela adibide gehiagorekin ere, bai gaztelaniatik eta bai ingelesetik itzulpen-sistemak hobetu ahal izateko.

Eranskina: aztertutako adibideen taula

Gaztelaniazko adibidea	Euskarazko adibidea
¡girar el timón 10 grados a babor!	lema 10 gradu ababorrera okertu!
¡viva Gales!	gora Gales!
¿has estado alguna vez en Marruecos?	Izan zara egundaino Marokon?
12 dividido por 2 igual a 6	12 zati 2 berdin 6
2 por 2 igual a 4	2 bider 2 4
20 menos 8 igual a 12	20 ken 8 berdin 12
5 más 2 igual a 7	5 gehi 2 berdin 7
60 euros y pico	60 euro eta koska
7 menos 2 igual a 5	7 ken 2 berdin 5
8 por 4 igual a 32	8 bider 4 berdin 32
a orillas del Sena	Sena ertzean
Aquí yace Mikel Goiburu	Hemen datza Mikel Goiburu
Baja Navarra	Nafarroa Beherea
Bellas Artes	Arte Ederrak
Bilbao está bien comunicado	Bilbo ondo komunikaturik dago
Cocer al baño María	Maria bainuan egosi;
	Maria bainutan egosi
Comisión Europea	Europako Batzordea
Compañía de Jesús	Jesusen Lagundia
conocí a Mikel en Bilbao	Bilbon ezagutu nuen Mikel
Creo que Mikel es tonto	Mikel tontoa dela iruditzen zait
crystal de Bohemia	Bohemiako kristala
cuando salí fuera de Francia	Frantziatik kanporatu nintzenean
delta del Nilo	Niloren delta
Departamento de Medio Ambiente	Ingurumen Saila
Dios maldijo la serpiente	Jainkoak sugea madarikatu zuen
el 50 aniversario de la boda	ezkontzaren 50. urtemuga
el canal de la Mancha	Mantxako kanala
el canal de Panamá	Panamako kanala
el cuadrado de 3 es 9	3ren karratua 9 da
el cubo de 2 es 8	2ren kubo 8 da
el desierto de Sáhara	Saharako basamortua
el día de Nochebuena es el 24 de diciembre	Gabon-eguna abenduaren 24a da
el diario de Ana Frank	Ana Frank-en egunkaria
el Espíritu Santo	Espiritu Santua
el estrecho de Gibraltar	Gibraltarko itsasartea
el Gran Cañón del Colorado	Coloradoko Arroila Handia
el jorobado de Notre-Dame	Notre-Dameko konkorra

el ministro de Sanidad	Osasun-ministroa
el norte de Francia	Frantziako iparraldea
el padre Donostia	aita Donostia
el Peñón de Gibraltar	Gibraltarko Haitza
el principio de Arquímedes	Arkimedesen printzipioa
El pueblo más antiguo de Europa	Europako populu zaharrena
el reino de Castilla	Gaztelako erresuma
El símbolo del hierro es Fe	Burdinaren sinboloa Fe da
el Tajo desemboca en el Atlántico	Tajo Atlantikoan itsasoratzen da
el Tratado de Roma	Erromako Ituna
En América amasó una fortuna	Ameriketean dirutza egin zuen
En esta botella caben 5 litros	Botila honek 5 litro hartzen ditu
en la zona de Tolosa	Tolosako partean
En las hermosas llanuras de Mesopotamia	Mesopotamiako ordoki ederretan
en una ocasión fuimos a París	behin Parisera joan ginen
encontrarse en Roma	Erroman elkarrekin gertatu
estado miembro de la Unión Europea	Europar Batasuneko estatua
ETA ha reivindicado el último atentado	ETAk azken atentatua errebindikatu du
Facultad de Derecho	Zuzenbide Fakultatea
finalmente me admitieron en la	azkenean Zaragozako unibertsitatean
universidad de Zaragoza	onartu ninduten
Fondo Monetario Internacional	Nazioarteko Diru Funtza
fórmula 1	1 formula
Fue mécido en Getxo, donde murió. . .	Mediku izan zen Getxon, non hil baitzen
fuera de Euskal Herria	Euskal Herritik landa
Gloria al Padre, al Hijo y al Espíritu Santo	Loria Aitari, eta Semeari, eta Izpiritu Sainduari
Hechos de los Apóstoles	Apostoluen eginak
hora de cierre: 19.00	itxiera-ordua: 19:00
hoy he quedado con Idoia	gaur Idoiarekin elkartzekoa naiz
la Alta Navarra	Nafarroa Garaia
la aparición de la Virgen	Ama Birjinaren agerkundea
la caída del Imperio romano	Erromatar Inperioaren erorialdia
La colonia de los vascos en París	Parisko euskaldunen kolonia
La cuenca del río Deba	Deba ibaiaren arroa
la dama de Amboto	Anbotoko dama
la decadencia del Imperio romano	Erromatar Inperioaren gainbehera
la descendencia de Jacob	Jakoben ondorea
la factoría de Seat en Manresa	Seat-en Manresako faktoria
la habitación mide 2 metros cuadrados	gelak 2 metro karratu ditu
la hoguera de San Juan	San Joan sua
la Luna gira alrededor de la Tierra	Ilargia Lurraren inguruan biratzen da

la ONU enbarrá observadores	NBEk begiraleak bidaliko ditu
la orquesta filarmónica de Boston	Bostongo Orkestra Filarmonikoa
la paga de Navidad	Gabonetako ordainsaria
la puerta tiene 80 cm de ancho	atea 80 cm zabal da
La Rioja es famosa por la excelencia de sus vinos	Errioxa famatua da bere ardoen bikaintasuna dela medio
la sala tiene un aforo de 300 personas	aretoaren edukiera 300 eserlekukoa da
la selva de Irati	Iratiko basoa
la situación en Iraq	Irakeko egoera
La temperatura en verano oscila entre los 20 y los 35 grados	Udan tenperatura 20 eta 35 gradu artean dabil
la unidad de los pueblos de América	Ameriketako herrien batasuna
Las aventuras de Tom Sawyer	Tom Sawyer-en abenturak
las cataratas de Iguazú	Iguazuko ur-jauziak
las obras de Mitxelena	Mitxelenaren lanak
le dicen Manolo	Manolo esaten diote
llegué con 5 minutos de anticipación	5 minutu lehenago iritsi nintzen
los Diez Mandamientos de Dios	Jainkoaren Hamar Aginteak
Los festivales musicales de Edimburgo	Edinburgh-eko musika-jaialdiak
los ingleses que se marcharon a América	Ameriketara joan ziren ingelesak
los menores de 14 años	14 urtez azpiko adingabeak
Los quechuas viven en los Andes	Kitxuak Andetan bizi dira
los Reyes Magos	Errege Magoak
Los siete sabios de Grecia	Greziako zazpi jakintsuak
Luisa está en cinta	Luisa haurdun dago
mar Mediterráneo	Mediterraneo itsasoa
Mikel está viendo la televisión	Mikel telebista ikusten dago
Mikel y los demás irán mañana	Mikel eta bihar joango dira
No he estado nunca en París	Ez naiz behin ere Parisen izan
Organización de las Naciones Unidas	Nazio Batuen Erakundea
Oriente Próximo	Ekialde Hurbila
pasando por Toulouse	Tolosan gaindi
Pasé unas vacaciones estupendas en Cuba	Oporraldi ederra egin nuen Kuban
paso de Ibañeta	Ibañetako lepoa
Pedro es un vinagre	Pedro ozpina da
reducción de un 10%	% 10eko beherapena
reproductor de DVD	DVD irakurgailua
San Ignacio	Inazio Donea
Sancho el fuerte	Antso Azkarra
Santa Ana	Santa Ana
Se fue a América, donde amasó una fortuna	Ameriketara joan, eta han aberastu egin zen
Secretaría General	Idazkaritza Nagusia
según Arquímedes...	Arkimedesen arabera...; Arkimedesez dioenez...

sita en Paris
su nombre es Izaskun pero
la llaman Ixaxkun
sucesor de San Pedro
Tendremos que suplir a Martín con otro
todos los precios llevan incluido
el Impuesto sobre el Valor Añadido
trompas de Falopio
un peso de 200 kilos
un tal Stalin
una capacidad de 10 litros
VIII. Congreso de Euskaltzaindia
Zezeaga juega en la zaga

Parisen kokatua
Izaskun deitzen da baina
Ixaxkun deitzen diote
San Pedroren ondorengoa
Martinen ordez besteren bat ipini beharko dugu
salneurri guztiek BEZa barne dute

Falopioren tronpak
200 kiloko pisua
Stalin delako morroi batek
10 litroko edukiera
Euskaltzaindiaren VIII. Biltzarra
Zezeagak atzealdean jokutzen du