

Euskarazko Itzulpen Automatikoa

NORA ARANBERRI (UPV/EHU), GORKA LABAKA (UPV/EHU)

Itzulpen Automatikoa adimen artifizialaren eta hizkuntzaren prozesamenduaren erronka nagusietako bat izan da hasiera-hasieratik. Ixa ikerketa taldean ere, euskararen prozesamendu automatikoa lantzeko helburuaren harira, lantzen dugu Itzulpen Automatikoa, alor horretako berrikuntzak euskarara ekarriz eta egokituz. Batez ere, gaztelaniazko testuak euskarara itzultzen dituzten sistemak sortu izan ditugu, baina euskaratik gaztelaniara eta ingelesetik euskarara itzultzen dituztenak ere garatu izan ditugu urte hauetan guztietan.

Itzulpen Automatikorako lehenengo sistemak sortu zirenetik gaur egun erabiltzen diren tekniketara, garapen nabaria izan da; Erregeletan Oinarritutako Itzulpen Automatikotik hasi, eta Itzulpen Automatiko Neuronalean bukatu, tartean Itzulpen Automatiko Estatistikotik pasatuta. Artikulu honetan, itzulpen-teknika bakoitzaren funtzionamendua aurkeztuko dugu, onurak eta desabantailak azalduz. Euskararako egindako lanarekin eta lortutako emaitzekin osatuko dugu atal bakoitza.

Bukatzeko, itzulpen automatikoaren erabilgarritasuna neurtzeko egindako saiakera batzuk aurkeztuko ditugu. Hizkuntza nagusienetan, aspalditik erabiltzen da itzulpen automatikoaren postedizioa, giza itzultzaileen lana azkartzeko. Euskararen kasuan, erabilera hori ez da zabaldu, hein handi batean itzulpenen kalitatea baxuagoa delako. Hala ere, gure azterketek erakusten dute hala moduzko itzulpenak ere lagungarriak izan daitezkeela.

Erregeletan Oinarritutako Itzulpen Automatikoa

Itzulpenak sortzeko gai izan ziren lehenengo sistemak, garai hartako joerarekin bat eginez, ezagutzen oinarritzen ziren, eta adituek osatutako hiztegiak eta erregelak erabiltzen zituzten

hizkuntza batean adierazitakoa beste hizkuntza batera itzultzeko. Hurbilpen horri Erregeletan Oinarritutako Itzulpen Automatikoa deritzo, eta, erabiltzen duen informazio linguistikoaren arabera, hainbat multzotan banatu daiteke.

Transferentzian Oinarritutako Sistemek jatorrizko esaldiaren analisi sintaktikoa egiten dute, eta lortutako zuhaitz-egiturei adituek idatzitako erregelak aplikatzen zaizkie, helburu-hizkuntzako esaldia lortzeko asmoz. Erregelek jatorri-hizkuntzako egitura sintaktikoa helburu-hizkuntzara egokitzen dute, lehendabizi. Behin helburu-hizkuntzaren zuhaitz sintaktikoa lortutakoan, beste erregela multzo bat erabiltzen da, bukaerako esaldia sortzeko. Halako sistemek ondo itzultzen dituzte aurreikusitako egitura sintaktikoak (erregeletan modu egokian adierazitakoak), baina arazoak dituzte aurreikusi gabeko egiturak eta hizkuntzaren salbuespenak modu egokian itzultzeko. Gainera, aukeraketa lexikoa arazo bat izaten da halako sistemetarako, ez baita batere erraza erregela jakin batzuk definitzea hitz baten itzulpen posibleen artean testuinguru jakin bati hobekien egokitzen zaiona aukeratzeko.

Euskarari dagokionez, Ixa Taldean, Matxin izeneko Erregeletan Oinarritutako Itzultzaile Automatikoa garatu genuen (Mayor, 2007; Mayor *et al.*, 2011). Matxinen lehenengo bertsioa bakarrik gaztelaniatik euskarara itzultzeko gai bazen ere, zenbait urte geroago, ingelesetik euskarara itzultzeko beharrezko erregelak eta hiztegiak gehitu genizkion (Aranberri *et al.*, 2015). Matxinek, erregeletan oinarritutako sistema bat izanda, hurbilpen hori jarraitzen dioten sistema guztien muga berberak ditu. Aurreikusitako egiturak eta hiztegia itzultzeko gai den arren (ikus 1. adibidea), malgutasuna falta zaio, eta, aurreikusi gabeko egitura edo hitzen bat topatzen duenean, itzulpen kaskarrak sortzen ditu (ikus 2. adibidea). Hala eta guztiz ere, erroreak ulergarri zaizkigu gizakioi, gehienetan, eta oraindik ere beste sistemen aurretik geratu ohi da gaztelania-euskarara itzultzaile automatikoen eskuzko konparaketetan.

- (1) ES: Es fácil perder los estribos en esa situación.
IA: Erraza da egoera horretan nor bere onetik ateratzea.
- (2) ES: Las ingenieras llevaron el proyecto a cabo.
IA: *Ingeniariek proiektua eraman zuten kabora.

Itzulpen Automatiko Estatistikoa

Mende-aldaketarekin batera, Itzulpen Automatikoan, adimen artifizialarekin loturiko hainbat alorretan gertatu zen bezala, metodo probabilistikoak nagusitu ziren. Garai berean sortu zen Itzulpen Automatiko Estatistikoa, azken urteetako paradigma nagusia. Itzulpen Automatiko Estatistikoa dagoeneko itzulita dauden corpus paralelo handiak erabiltzen ditu, agertzen diren hitzen eta hitz-segiden itzulpenak ikasteko. Itzulpen-baliokidetzak horiek batera, bakoitzaren probabilitatea —edo konfiantza-maila— ere ikasten du, eta informazio hori erabiltzen du itzuli beharreko esaldi bakoitzaren itzulpen probableena topatzeko. Halako sistemek, ikasteko datu nahikoa dutenean behintzat, itzulpen naturalak sortzen dituzte, eta, itzulpenak testu errealean

ikusitako hitz-segidez osaturik daudenez, erregeletan oinarritutako sistemek sortutakoak baino naturalagoak dirudite lehenengo irakurketan. Baina, zehaztasun handiagoz begiratzerakoan, nabaritzen da arazoak dituela esaldiaren egitura egokia sortzeko: distantzia luzeko komunztadura mantentzeko arazoak eta hitz ordena desegokia, besteak beste.

Euskararen kasuan, hurbilpenak berezko dituen arazoei hizkuntzaren zailtasuna gehitu behar zaio. Itzulpen Automatiko Estatistikoak itzulpen literalak egiteko joera du, eta, hori ekiditeko, entrenamendu-datu asko behar ditu. Antzeko hizkuntzetan, hori ez da arazo bat, eta tresnak itzulpen egokiak egiten ikas dezake corpus txikia erabili arren. Gaztelaniaren eta euskararen arteko desberdintasun linguistikoak, berriz, corpus handiaren beharra dakar nahi-taez. Baina euskaraz ditugun corpus paraleloak hizkuntza nagusietarako daudenak baino askoz txikiagoak dira.¹ Ixa Taldean, zenbait saiakera egin ditugu Itzulpen Automatiko Estatistikoak gaztelania-euskara hizkuntza bikotera egokitzeko (Labaka, 2010). Lan horietan, bereziki tratatu ditugu gaztelania-euskara Itzulpen Automatiko Estatistikoaren bi arazo-iturri nagusiak: euskararen morfologia aberatsa eta hizkuntzen hitz-ordenan dauden diferentzia handia.

Euskararen morfologiaren tratamendua errazteko asmoz, euskarazko hitzen segmentazioa erabili genuen (Diaz de Ilarraza et al., 2009a), hau da, euskarazko hitzaren lema eta atzizkia banatzea (*etxe*ko → *etxe* +*ko*), corpuseko token bakoitzaren agerpen kopurua eta estatistiken esangura handitzeko asmoz. Gainera, aurreprozesamendu horrek bi hizkuntzetan erabiltzen diren token kopurua berdintzen du, eta, hartara, itzultzailearen lana errazten. Noski, itzultzaileak halako testu moldatu bat erabiltzen badu ikasteko, itzulpenek itxura bera izango dute, eta postprozesu bat beharko dugu benetako itzulpena osatu ahal izateko (ikus 3. adibidea).

- (3) ES: El mercado europeo de las tecnologías crecerá el 3,2%
IA-seg: Teknologia +en europa +ko merkatu +a % 3,2 hazi +ko da
IA: Teknologien europako merkatua % 3,2 haziko da

Bestalde, itzultzaile estatistikoak jatorrizko esaldiaren ordena mantentzeko duen joerari aurre egiteko, gaztelaniazko hitzen ordena aldatzen duten erregela multzo bat defini daiteke (Diaz de Ilarraza et al., 2009b). Modu horretan, ordena berria gertuago egongo da euskarazko itzulpenarenetik. Definitutako erregela multzoak esaldi bukaerara mugitzen du gaztelaniazko aditza, edo izen-sintagmaren bukaerara —euskarazko atzizkiaren kokalekura— gaztelaniazko artikulua eta preposizioak, besteak beste. Berrito ere, halako moldaketan helburua Itzulpen Automatiko Estatistikoaren lana erraztea da, itzultzaileak ahalik eta eraldaketa gutxien egiteko beharra izan dezan. Kasu honetan, segmentazioarekin ez bezala, egiten ditugun moldaketak jatorrizko esaldietan egiten ditugu, eta itzultzaileak sortutako konponbidea bere horretan erabili daiteke, ez baitugu inongo postprozesuren beharrik (ikus 4. adibidea).

1. Gaztelania-ingelesa hizkuntza bikotean ez dago arazorik 100 milioi esaldi itzuli topatzeko; aldiz, euskara-gaztelania hizkuntza bikotean, zaila da 10 milioi esaldiko corpus paralelo bat biltzea.

- (4) ES: Cloverfield contó con un presupuesto de 25 millones de dólares.
 ES-ber: Cloverfield 25 millones de dólares de presupuesto un con contó.
 IA: Cloverfield-ek 25 milioi dolarreko aurrekontua izan zuen.

Bi moldaketa horiek konbinatuz, gaztelania-euskara Itzulpen Automatiko Estatistikoaren kalitatea hobetzea lortu genuen, bai eta orain arteko emaitza hoberenak erdietsi ere itzulpenaren kalitatea neurtzen duten metrika automatikoetan. Baina, lehenago esan bezala, ebaluazio automatikoetatik eskuzkoetara pasatzen garenean, oraindik ere Erregeletan Oinarritutako Sistemek lortu ohi dituzte emaitza hoberenak.

Itzulpen Automatiko Neuronal

Azken urteetan, beste hurbilpen bat ari zaie aurreko biei gailentzen: Itzulpen Automatiko Neuronal. Itzultzaile estatistikoak bezalaxe, corpus elebidunak erabiltzen ditu, itzulpena egiten automatikoki ikasteko. Baina, kasu horretan, Sare Neuronal Artifizialak erabiltzen dira itzulpen-prozesua aurrera eramateko. Sare Neuronal Artifizialek zenbakizko errepresentazioak erabiltzen dituzte, eta jatorri-hizkuntzako esaldiak zenbakizko errepresentazio horretara pasatzea eskatzen du. Behin zenbakizko errepresentazio horiek izanda, gizakion garunaren funtzionamenduan inspiratutako Sare Neuronal Artifizialek errepresentazio hori moldatzen dute, helburu-hizkuntzako errepresentazioa lortu arte. Halako itzultzaileek oso itzulpen naturalak lortzen dituzte, baina arazo larriena itzultzailearen funtzionamendua interpretatzeko zailtasunak dira. Itzulpen-prozesu osoa zenbaki-bektoreen gaineko eragiketa aritmetikoen bitartez egiten da, eta, erroren bat sortzen denean, oso zaila da erroren jatorria topatzea, eta are zailagoa irtenbide bat proposatzea.

Hizkuntza nagusietara, halako sistemek lortzen dituzte itzulpen hoberenak, eta, hori dela eta, Google-k dagoeneko eskaintzen ditu halako itzultzaileak, 8 hizkuntza bikote handitarako (ingelesa-gaztelania eta ingelesa-frantsesa, besteak beste). Baina, euskara eta halako hizkuntza txikiagoetarako, oraindik ez da nahikoa ikertu, eta ez dago garbi Itzultzaile Automatiko Neuronalekin nolako portaera duten ikasteko datu gutxiago daudenean. Ixa Taldean, jarriak gara dagoeneko horretan lanean: MODELA proiektuan hasi gara euskararako lehenengo Itzultzaile Automatiko Neuronalak sortzen. Lehenengo emaitzak oso itxaropentsuak dira. Itzulpenen kalitatea neurtzeko erabiltzen diren metrika automatikoetan diferentzia handirik ikusi ez arren, lehenengo azterketa subjektiboek adierazten dute kalitate-hobekuntza bat badagoela. Nahiz eta halako sistemek ere erroreak egiten dituzten, itzulpenak beste sistemekin lortutakoak baino naturalagoak dira. 5. adibideko itzulpena zuzena da euskaraz, baina, itzultzaileak oso argi ez duenez nola itzuli *por un golpe de calor*, ezbehar baten ondorioz itzuli du, ezbehar mota zehaztu gabe eta informazio apur bat galduz. Orokortze-maila hori onargarria izan daiteke kasu batzuetan, baina, 6. adibidean ikus daitekeen bezala, itzultzaile neuronalak behar baino gehiago

alda dezake esanahia. Halakoetan, informazioa falta izateaz gain, esaldiaren aditz nagusiaren esanahia (*han sido confiscados*) ere aldatzen du (*ordaindu dituzte*).

- (5) ES: El suceso ha resultado herido un bombero por un golpe de calor y ha sido trasladado al Hospital de Gernika.
IA: Ezbehar baten ondorioz, suhiltzaile bat zauritu da eta Gernikako Ospitalera eramán dute.
- (6) ES: Durante la operación han sido confiscados bienes valorados en unos seis millones de dólares (4,5 millones de euros)
IA: Operazioan, zazpi milioi dolar (4,5 milioi euro) ordaindu dituzte.

MODELA proiektua bukatu eta ebaluazio zehatzago bat egin arte, ezingo dugu jakin zein den halako hurbilpenen benetako kalitatea. Baina, emaitza zehatz horiek gabe ere, dagoeneko esan dezakegu halako metodoen inguruan mugituko dela hurrengo urteetako ikerketa.

Sistemen kalitatea eta erabilgarritasuna

Sistemak garatzeaz gain, Ixa Taldea etengabe ari da sistema horien kalitatea neurtzen eta itzulpen automatikoa nola aprobe txta daitekeen aztertzen. Azken finean, sistemak garatzen badira, gizarteratzeko asmoarekin izango da. Horrela, sistemak baliabideekin aberasten diren neurrian eta teknika berriak aplikatzen zaizkien neurrian, horrek kalitatean zer eragin duen aztertzen da. Horretarako, metrika automatikoak erabiltzen dira, hala nola BLEU (Papineni *et al.*, 2002) edo TER (Snover *et al.*, 2006), baina baita pertsonen egindako ebaluazioak ere, non itzulpenen jartotasuna edota doitasuna aztertzen dituzten, esaterako. Izan ere, euskararen kasuan bereziki, emaitza kontrajarriak lortu dira maiz bata edo bestea erabiltzerakoan. Euskararen sintagma-mailako ordena libreak, alde batetik, eta izaera eranskariak, bestetik, maila lexikalean oinarritzen diren ohiko metrika automatikoen mugak azaleratzen dituzte. Adibide gisa, Labakak (2010) eranskaritasunari aurre egiteko tekniken emaitzak aurkeztu zizkigun. Metrika automatikoetan ez zen ikusten aldaketarik kalitatean. Giza ebaluatzaileek, aldiz, eragin positiboa nabaritu zuten itzulpenetan. Sistemen arteko konparazioetan ere antzera gerta daiteke (Labaka *et al.*, 2014): metrikek sistema bat hobea dela esan, eta giza ebaluatzaileentzat beste bat nagusitu. Metrika automatikoak guztiz beharrezkoak eta erabilgarriak zaizkigu garapen jarraitu eta azkar bat bermatzeko, baina Ixa Taldean interes berezia daukagu azken erabiltzaileak ebaluazioetan parte hartzeko (Alegria *et al.*, 2013; Aranberri *et al.*, 2014, 2016).

Neurketak neurketa, ezin esan dezakegu euskarara itzultzen duten sistemek gaur egun kalitate profesionaleko testuak ekoizten dituztenik. Arean ere, zenbait hizkuntza bikotetarako egindako esperiméntuen emaitzak kontu handiz konparatu behar badira ere, ukaezina da euskarara itzultzen duten sistemak hizkuntza handietako sistemek lortzen duten kalitatearen

azpitik daudela oraindik. Baina orain lortzen duten kalitatea nahikoa izan al daiteke onura ateratzen hasteko?

Itzulpen automatikoa erabiltzaile arruntek zein itzultzaile profesionalek balia dezakete. Lehenengo kasuan, batik bat atzerriko hizkuntza batean dagoen testu bat euskarara ekarri eta hobeto ulertzeko edo, zergatik ez, norberaren itzulpen-beharrak asetzeko, edo behintzat laguntzeko izaten da. Bigarren kasuan, aldiz, eguneroko lana eraginkortasun handiagoz egiteko, itzulpena eta postedizioa (itzulpen automatikoaren orrazketa) tartekatuz, gehienetan.

Kalitate ona lortzen duten sistemek erabiltzaile arruntak poz-pozik uzteko moduko adibi-deak ematen dituzte. Nork ez ditu emaitza dotoreak aurkitu ingeleseko testu bat ezin ulertu eta gaztelaniara itzulita, edo ingelesezko testu bat idazteko laguntza lortu gaztelaniatik hasita Google-ren itzultzailearekin? Itzulpen profesionalaren esparruan, erronka handitu egiten da. Hala ere, hainbat enpresa eta neurketak ondorioztatu dute itzultzaile profesionalen produktibitatea handitu egiten dela hizkuntza handiekin eta batez ere sistema espezializatuekin (Schäfer, 2004; Plitt eta Masselot, 2010).

Euskararen kasuan, baina, non kalitatea hala moduzkoa den askotan, zalantzan jartzen da erabilgarritasuna. Hori aztertzeke asmoz, Ixa Taldean lanean ari gara erabiltzaile arrunt eta profesionalek itzulpen automatikoaren nola erabiltzen duten aztertzeke. Esaterako, ENEUS proiektuan, Elhuyarrekina eta UPV/EHUko Informatika Fakultateko zenbait irakaslerekin batera, erabiltzaile arruntak eta itzultzaile profesionalen postedizio- eta itzulpen-lana alderatu zen (Aranberri *et al.*, 2014). Emaitzak erakutsi zuen erabiltzaile arruntak itzulpen automatikoa balioesten duela, kalitate baxukoa izanda ere, bereziki adituaren alorreko testuak itzultzeko orduan.

Itzultzaileekin ere egiten dugu lan. Aspaldi da giza ebaluazioetan parte hartzen dutela, baina, postedizioari dagokionez, UEUren ikastaroetan egin zen lehen urratsa, 2015eko udazkenean (Aranberri, 2016). Zenbait itzultzaile profesionalekin bildu ginen, eta euskarara posteditatzearekin zioten, atazari buruz hausnartzen genuen bitartean. Halako elkarlanek erakusten dute zer-nolako jarrera dagoen teknologia honekin, zer etekin atera dakioken kalitateari dagokionez —badirudi testuinguru batzuetan jada atera dakiokeela—, eta zer lan egiten den den zehazki posteditatzerakoan, itzulpen automatikoaren garapenean lehentasunak jartzeko.

BIBLIOGRAFIA

- Alegria, Iñaki; Cabezón, Unai; Fernandez De Betoño, Unai; Labaka, Gorka; Mayor, Aingeru; Sarasola, Kepa & Zubiaga, Arkaitz (2013). «Reciprocal Enrichment between Basque Wikipedia and Machine Translators», *The People's Web Meets NLP: Collaboratively Constructed Language Resources*, book edited by Iryna Gurevych and Jungi Kim, Springer, Book series “Theory and Applications of Natural Language Processing”, E. Hovy, M. Johnson and G. Hirst (eds.).
- Aranberri, Nora. (2016). «Ba al dago lekurik euskararako postedizioarentzat?», *Senez* 47, 195-203.
- Aranberri, Nora; Labaka, Gorka; Diaz De Illaraza, Arantza & Sarasola Kepa (2014). «Comparison of post-editing productivity between professional translators and lay users», *Proceedings of the Third Workshop on Post-Editing Technology and Practice (WPTP - 3)*, Sharon O'Brien, Michel Simard and Lucia Specia (eds.).
- Aranberri, Nora; Labaka, Gorka; Diaz De Ilaraza, Arantza & Sarasola K. (2015). «Exploiting portability to build an RBMT prototype for a new source language», *Proceedings of the 18th Annual Conference of the European Association for Machine Translation, EAMT-2015*, 3-10.
- Aranberri, Nora; Labaka, Gorka; Diaz De Ilaraza, Arantza & Sarasola, Kepa (2016). «Ebaluatoia: crowd evaluation for English–Basque machine translation», *Language Resources and Evaluation*, 1-32.
- Diaz De Ilaraza, Arantza; Labaka, Gorka & Sarasola, Kepa (2009a). «Relevance of Different Segmentation Options on Spanish-Basque SMT», *Proceedings of the 13th Conference of the European Association for Machine Translation EAMT 2009*, 74-80, Bartzelona, Katalunia.
- Diaz De Ilaraza, Arantza; Labaka, Gorka & Sarasola, Kepa (2009b). «Reordering in Spanish–Basque SMT», *MT Summit XII*. Ottawa, Kanada.
- Mayor, Aingeru (2007). «Matxin: erregeletan oinarritutako itzulpen automatikoko sistema baten eraikuntza estaldura handiko baliabide linguistikoak berrerabiliz», *Euskal Herriko Unibertsitateko Donostiako Informatika Fakultatea*, 2007ko azaroaren 27a.
- Mayor, Aingeru; Alegria, Iñaki; Diaz De Ilaraza, Arantza; Labaka, Gorka; Lersundi, Mikel & Sarasola, Kepa (2011). «Matxin, an open-source rule-based machine translation system for Basque», *Machine Translation Journal*, Volume 25, Issue 1 (2011), 53-82.
- Labaka, Gorka (2010). «EUSMT: Incorporating Linguistic Information into SMT for a Morphologically Rich Language. Its use in SMT-RBMT-EBMT hybridation», *Lengoaia eta Sistema Informatikoak Saila (UPV-EHU)*. Donostia. 2010ko martxoaren 29a.

- Labaka, Gorka; España-Bonet, Cristina; Màrquez, Lluís & Sarasola, Kepa (2014). «A hybrid machine translation architecture guided by syntax», *Machine Translation Journal*, 28(2), 91-125.
- Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei-Jing (2002). «BLEU: a method for automatic evaluation of machine translation», *In Proceedings of the 40th annual meeting on association for computational linguistics*, 311-318, Philadelphia, Pennsylvania, July 6-12.
- Plitt, Mirko & Masselot, François (2010). «A productivity test of statistical machine translation post-editing in a typical localisation context», *In The Prague Bulletin of Mathematical Linguistics*, 7-16, Prague, Czech Republic: Universita Karlova.
- Schäfer, Falko (2003). «MT post-editing: How to shed light on the ‘unknown task’. Experiences made at SAP», *In 8th International workshop of the European Association for Machine Translation (EAMT 03)*, Dublin City University, Dublin, Ireland, May 15-17.
- Snover, Matthew; Door, Bonnie; Schwartz, Richard; Micciulla, Linnea & Makhoul, John (2006). «A study of translation edit rate with targeted human annotation», *In Proceedings of association for machine translation in the Americas* (Vol. 200, No. 6).

Resumen

La traducción automática es uno de los mayores retos a los que se enfrentan la inteligencia artificial y el procesamiento lingüístico desde sus inicios. El grupo de investigación IXA, en el marco del procesamiento automático del euskera, trabaja en el ámbito de la traducción automática mediante la aplicación y adecuación al euskera de las innovaciones que surgen en dicho campo. A lo largo de todos estos años el Grupo IXA se ha centrado en la creación de sistemas de traducción del castellano al euskera. No obstante, también se han desarrollado sistemas de traducción del euskera al castellano, así como sistemas que combinan pares de lenguas como el euskera y el inglés.

Las técnicas de traducción automática utilizadas actualmente distan mucho de las que se empleaban en los primeros años: empezando por la traducción automática basada en reglas hasta la traducción automática neuronal, pasando por la traducción automática estadística. El presente artículo describe el funcionamiento de cada una de esas técnicas de traducción, y señala sus ventajas e inconvenientes. Cada uno de los apartados se completa con una explicación del trabajo realizado y con los resultados obtenidos en euskera.

Por último, el artículo presenta diversos métodos para medir la empleabilidad de la traducción automática. Hace tiempo que las lenguas mayoritarias utilizan la postedición de la traducción automática como vía para acelerar la labor de los traductores humanos. Su uso no se ha extendido en el caso del euskera, en gran medida porque la calidad de las traducciones automáticas es más baja, si bien los estudios realizados por el Grupo IXA muestran que este tipo de traducciones puede resultar útil.

Résumé

La Traduction Automatique est l'un des plus grands défis auxquels est confronté, dès le début, l'Intelligence Artificielle et le Traitement Linguistique. Le groupe de recherche IXA s'intéresse également à la Traduction Automatique, dans le cadre du traitement automatique de l'euskara, en appliquant et en adaptant à cette langue les nouveautés qui surgissent dans ce domaine. Tout au long de ces dernières années, le Groupe IXA s'est consacré à la création de systèmes de traduction de l'espagnol à l'euskara. Néanmoins, il a également développé des systèmes de traduction de l'euskara vers l'espagnol, ainsi que des systèmes qui combinent des paires de langues telles que l'euskara et l'anglais.

Les techniques de Traduction Automatique employées actuellement sont bien différentes de celles utilisées au début. En effet, l'on a commencé par la Traduction Automatique basée sur des Règles pour aboutir à la Traduction Automatique Neuronale, en passant par la Traduction Automatique Statistique. L'article qui nous occupe décrit le fonctionnement de chacune de ces techniques de traduction, tout en soulignant leurs avantages et leurs inconvénients. Chacune des descriptions est complétée par l'explication du travail réalisé au niveau de l'euskara et des résultats obtenus.

Pour finir, cet article présente diverses méthodes pour mesurer le degré d'utilisation de la traduction automatique. Cela fait longtemps que les langues majori-

taires utilisent la post-édition de la traduction automatique comme moyen d'accélérer le travail des traducteurs humains. Dans le cas de l'euskara, son emploi ne s'est pas développé, cela étant dû, dans une grande mesure, au fait que la qualité des traductions n'est pas satisfaisante, mais les études réalisées par le Groupe IXA prouvent, toutefois, que ce genre de traduction peut être fort utile.

Abstract

Machine Translation is one of the greatest challenges ever faced by Artificial Intelligence and Linguistic Processing. In the research group IXA, as part of our work on the automatic processing of Basque, we also study Machine Translation through the application and adaptation to Basque of innovations that arise in the field. We have focused primarily on the creation of systems that translate from Spanish into Basque. Nevertheless, we have also developed systems of translation from Basque into Spanish, as well as systems that deal with pairs of languages like Basque and English.

The techniques of Machine Translation currently in use differ greatly from those used in the early years, and have evolved from Rule-Based Machine Translation to Neural Machine Translation, by way of Statistical Machine Translation. In this article, we describe the operation of these translation techniques, pointing out the advantages and disadvantages of each. We complete each section with an explanation of the work carried out and results obtained in Basque.

Finally, we present various methods for measuring the usability of Machine Translation. In the major languages, post-editing of Machine Translation has long been used as a way of speeding up the labor of human translators. Its use has not carried over into Basque however, largely because the quality of the translations is lower. Nevertheless, studies carried out by the IXA Group demonstrate that this type of translation can be useful.