

METANET4U 

**D2.3.eus.en
Language Report for
Basque
(Basque version)**

Version 1.1

2011-06-29



METANET4U

www.metanet4u.eu

The central objective of the Metanet4u project is to contribute to the establishment of a pan-European digital platform that makes available language resources and services, encompassing both datasets and software tools, for speech and language processing, and supports a new generation of exchange facilities for them.

This central objective is articulated in terms of the following main goals:

Assessment: to collect, organize and disseminate information that permits an updated insight into the current status and the potential of language related activities, for each of the national and/or language communities represented in the project. This includes organizing and providing a description of: language usage and its economic dimensions; language technologies and resources, products and services; main actors in different areas, including research, industry, government and society in general; public policies and programs; prevailing standards and practices; current level of development, main drivers and roadblocks; etc.

Collection: to assemble and prepare language resources for distribution. This includes collecting languages resources; documenting these language resources; upgrading them to agreed standards and guidelines; linking and cross-lingual aligning them where appropriate.

Distribution: to distribute the assembled language resources through exchange facilities that can be used by language researchers, developers and professionals. This includes collaboration with other projects and, where useful, with other relevant multi-national forums or activities. It also includes helping to build and operate broad inter-connected repositories and exchange facilities.

Dissemination: to mobilize national and regional actors, public bodies and funding agencies by raising awareness with respect to the activities and results of the project, in particular, and of the whole area of language resources and technology, in general.

METANET4U is a project in the META-NET Network of Excellence, a cluster of projects aiming at fostering the mission of META. META is the Multilingual Europe Technology Alliance, dedicated to building the technological foundations of a multilingual European information society.



Deliverable D2.3.eus: Language Report for Basque (Basque version)

METANET4U is co-funded by the participating institutions and the ICT Policy Support Programme of the European Commission



and by the participating institutions:



Faculty of Sciences, University of Lisbon



Instituto Superior Técnico



University of Manchester



University *Alexandru Ioan Cuza*



Research Institute for Artificial Intelligence,
Romanian Academy



University of Malta



Technical University of Catalonia



Universitat Pompeu Fabra

Revision History

Version	Date	Author	Organisation	Description
1.0	03-06-2011	I. Hernáez, E. Navas, I. Odriozola, K. Sarasola, A. Díaz de Ilarraza, I. Aizpurua, A. Díaz de Lezama, B. Oihartzabal, J. Salaberria	EHU	Daft version
1.1	29-06-2011	I. Hernáez, E. Navas, I. Odriozola, K. Sarasola, A. Díaz de Ilarraza, I. Aizpurua, A. Díaz de Lezama, B. Oihartzabal, J. Salaberria	EHU	FINAL version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.



METANET4U

D2.3.eus **Language Report for** **Basque**

(Basque version)

Document METANET4U-2011-D2.3.eus
EC CIP project #270893

Deliverable

Number: D2.3.eus

Completion: Final

Status: Submitted

Dissemination level: Public

Responsible: Asunción Moreno (WP2 coordinator)

Contributing Partners: Universitat Politècnica de Catalunya, Universitat
Pompeu I Fabra

Authors: I. Hernáez, E. Navas, I. Odriozola, K. Sarasola, A. Diaz de
Ilarraza, I. Aizpurua, A. Díaz de Lezama, B. Oihartzabal, J. Salaberria

© all rights reserved by FCUL on behalf of METANET4U

Aurkibidea

Laburpena	2
Arriskua, gure hizkuntzentzat, eta erronka, hizkuntza- teknologiarentzat	4
Hizkuntza-mugek oztopoak jartzen dizkiote Europako informazio- gizarteari	4
Gure hizkuntzak arriskuan	5
Hizkuntza-teknologia teknologia bideratzaile giltzarria da	6
Hizkuntza-teknologiaren aukerak	6
Hizkuntza-teknologiaren erronkak	7
Hizkuntzaren jabetzea	8
Euskara Europako informazio-gizartean	9
Datu orokorrak	9
Euskararen berezitasunak	9
Azken gertaerak	11
Hizkuntza-lanketa	12
Hizkuntza hezkuntzan	12
Nazioartean	13
Euskara Interneten	14
Irakurgai gehiago	14
Hizkuntza-teknologia euskararako	16
Hizkuntza-teknologiak	16
Hizkuntza-teknologia aplikatzeko arkitekturak	16
Aplikazio-eremu komunak	17
Hizkuntza-zuzentzailea	17
Sareko bilaketak	18
Ahozko interakzioa	19
Itzulpen automatikoa	21
Hizkuntza-teknologiaren atzeko lana	23
Hizkuntza-teknologia hezkuntzan	24
Hizkuntza-teknologiako programak	26
Euskararako tresnen eta baliabideen eskuragarritasuna	27
Tresna eta baliabideen taula	28
META-NET sarea	30
Lan-ildoak	30
Erakunde kideak	32
Erreferentziak	35

Laburpena

Europako hizkuntza asko aro digitalaren biktima izateko arriskuan daude, linean ordezkari eta baliabide gutxiegi baitituzte. Eskualdeko merkatu-aukera handiak ustiatu gabe gelditzen dira gaur egun, hizkuntzak ezarritako oztupoak direla eta. Orain zerbait egiten ez badugu, Europako herritar asko atzean geldituko dira gizarte eta ekonomia aldetik, beren ama-hizkuntza hitz egiten dutelako.

Hizkuntza-teknologia (HT) berritzaileari esker, berdintasuna, inklusibismoa eta arrakasta ekonomikoa oinarri dituen ezagutza-eta informazio-gizarte baten parte izango dira Europako herritarrak. Hizkuntza-teknologia eleaniztunak atea zabalduko du hizkuntza batetik bestera segituan, merke eta erraz komunikatzeko.

Gaur egun, nagusiki AEBko hornitzaile komertzialek eskaintzen dituzte hizkuntza-zerbitzuak. Google Translate doaneko zerbitzua adibide bat besterik ez da. Hizkuntza-teknologiaren ahalmen ikaragarria agerian gelditu da IBMren Watson sistema informatikoak izan berri duen arrakastarekin; bada, telebistako *Jeopardy* lehiaketara eraman zuten gizakien aurka lehiatzera, eta irabazi egin zuen. Europarrak garen aldetik, zenbait premiazko galdera egin behar dizkiogu geure buruari:

- ❑ Gure komunikazio- eta ezagutza-azpiegiturek monopolioen mende egon behar lukete?
- ❑ Benetan fida gaitezke beste batzuek berehala desaktiba ditzaketen hizkuntza-zerbitzuek?
- ❑ Modu aktiboan ari gara lehiatzen hizkuntza-teknologiaren ikerketa eta garapenaren merkatu globalean?
- ❑ Beste kontinenteetako hirugarren pertsona batzuek nahi dute gure itzulpen-arazoak eta Europako eleaniztasunarekin lotutako beste arazo batzuk konpontzea?
- ❑ Europaren aurrekari kulturek lagundu dezakete ezagutza-gizartea gorpuzten kalitatezko teknologia hobea, seguruagoa, zehatzagoa, berritzaileagoa eta sendoagoa eskainiz?

Euskarari buruzko liburu zuri honek erakusten du euskarak baduela industria bizi bat eta ikerketa-giro bat hizkuntza-teknologiaren inguruan. Euskara baturako badaude teknologia eta baliabide batzuk, baina oso gutxi, ingelesekoen aldean. Eta dauden teknologia eta baliabideak kalitate eskasagoak dira.

Txosten honetan egindako kalkuluen arabera, euskarak aurrerapausorik egingo badu, berehala jarri behar du lanean.



Arriskua, gure hizkuntzentzat, eta erronka, hizkuntza-teknologiarentzat

Komunikazioan eta gizartean izugarrizko eragina izaten ari den iraultza digital baten aurrean gaude. Komunikazio-teknologia digitalizatu eta sarekoetan izan berri diren aurrerapenak Gutenbergek inprenta asmatu zuenekoarekin alderatzen dira, batzuetan. Zer esaten digu analogia horrek Europako informazio-gizartearen eta geure hizkuntzen etorkizunari buruz?

Gutenbergen asmakizunaren ondoren, benetako aurrerapausoak egin ziren komunikazioan eta ezagutzen trukaketan, hainbat lan esker; esaterako, Lutherrek eginikoa, Biblia hizkuntza arrunt batera itzuli zuenean. Hurrengo mendeetan, teknika kulturalak garatu dira hizkuntza-prozesamendua eta ezagutza-trukaketa hobeto egiteko:

- ❑ Hizkuntza handien ortografia eta gramatika estandarizatzeak aukera eman zuen ideia zientifiko eta intelektual berriak azkar zabaltzeko.
- ❑ Hizkuntzen ofizialtasunak aukera eman zien herritarrei muga jakin batzuen barruan (sarritan, politikoak) komunikatzeko.
- ❑ Hizkuntzen irakaskuntzari eta itzulpenari esker, hizkuntzen arteko trukaketa etorri zen.
- ❑ Kazetaritzako eta bibliografiako jarraibideak sortzeak material argitaratuaren kalitatea eta eskuragarritasuna bermatu zuen.
- ❑ Hedabide berriek —egunkariak, irratia, telebista, liburuak eta beste batzuk— komunikazio-beharrei erantzun zieten.

Azken hogeitun urteotan, informazio-teknologiak lagundu egin du prozesu horietako asko automatizatzen eta errazten:

- ❑ Autoedizioko softwareak hartu du idazmakinarene eta monotipoaren tokia.
- ❑ Microsoft PowerPoint programak hartu du proiektagailuz erakutsitako gardenkien tokia.
- ❑ Mezu elektronikoen bidez faxez baino azkarrago bidali eta jasotzen dira dokumentuak.
- ❑ Skype erabiliz, interneteko telefono-deiak egin daitezke, eta elkargune birtualak sortu.
- ❑ Audio- eta bideo-fitxategien kodetze-formatuei esker, erraza da multimedia-fitxategiak trukatzeko.
- ❑ Bilatzaileetan gako-hitzak sartuz web-orrietara sar gaitezke.
- ❑ Lineako zerbitzuek itzulpen azkar eta gutxi gorabeherakoak sortzen dituzte; hor dugu, esaterako, Google Translate.
- ❑ Gizarte-hedabideen plataformek erraztu egiten dute elkarlana eta informazioa partekatzea.

Tresna eta aplikazio horiek lagungarriak badira ere, oraindik ez dute lortu informazio-gizarte europar eleaniztun eta jasangarri bat ezartzea; gizarte moderno eta inklusibo bat, non informazioa eta produktuak askatasunez ibiltzen diren alde batetik bestera.

Hizkuntza-mugek oztupoak jartzen dizkiote Europako informazio-gizarteari

Ezin dugu jakin etorkizuneko informazio-gizarte zehazki nolakoa izango den. Europaren energia-estrategiaz edo atzerriko politika

bateratuaz hitz egin behar denean, Europako atzerriko ministroak beren jatorrizko hizkuntzan mintzatzen entzun nahi izango ditugu, beharbada. Agian, plataforma bat izan nahi dugu, non hainbat hizkuntza hitz egiten dituzten eta era askotako hizkuntza-mailak dituzten pertsonak gai jakin bati buruz solasean arituko diren, teknologiak haien iritziak bildu eta laburpen txikiak egiten dituen bitartean. Baliteke, halaber, beste herrialde batean dagoen osasun-aseguruen bulego batekin hitz egin nahi izatea.

Argi dago gaur egungo komunikazioak, duela urte batzuetakoaren aldean, beste kalitate batekoa izan behar duela. Ekonomia eta informazio-eremu globalean, hizkuntza, hiztun eta eduki gehiagorekin egiten dugu topo, eta hedabide-mota berriekin berehala harremanetan jartzeko eskatzen digute. Gizarte-hedabideak (Wikipedia, Facebook, Twitter eta YouTube) izaten ari diren arrakasta icebergaren tontorra besterik ez da.

Gaur egun, hainbat gigabyteko testuak bidal ditzakegu mundu osora segundo gutxi batzuetan, ulertzen ez dugun hizkuntza batean dagoela ohartu baino lehen. Europako Batzordeak eskatuta duela gutxi egin den txosten baten arabera, Europako Internet-erabiltzaileen % 57k bere jatorrizko hizkuntzaz bestelako hizkuntzetan erosten ditu produktuak eta zerbitzuak (ingelesa da gehien erabiltzen den atzerriko hizkuntza, eta, haren ondoren, frantsesa, alemana eta gaztelania). Erabiltzaileen % 55ek irakurtzen ditu edukiak atzerriko hizkuntza batean, baina soilik % 35ek idazten ditu mezu elektronikoak edo sareko iruzkinak.ⁱ Duela urte gutxi batzuk, ingelesa izan zen sareko *lingua franca* — sareko edukiaren parte oso handi bat ingelesez zegoen—, baina egoera goitik behera aldatu da orain. Beste hizkuntza batzuetan idatzitako edukien kantitatea izugarri handitu da sarean (batez ere, asiar eta arabiar hizkuntzetan idatzitakoena).

Hizkuntza-mugek eragin duten nonahiko banaketa digitalak ez du toki handirik hartu diskurtso publikoan. Alabaina, galdera bat sortzen du behin eta berriz: “Europako zein hizkuntzak egingo du aurrera eta iraungo du sareko informazioaren eta ezagutzaren gizartean?”.

Gure hizkuntzak arriskuan

Inprentak informazio-trukaketa eskerga ekarri zuen Europara, baina bertako hizkuntza asko desagertzea ere eragin zuen. Eskualdeetako hizkuntzak eta hizkuntza txikiak apenas erabiltzen

ziren argitalpenetarako. «Horren ondorioz, hizkuntza asko ahozko transmisiora mugatu ziren —adibidez, kornubiera eta dalmaziera—, eta, beraz, mugatuta gelditu zen haien etengabeko ikaskuntza, zabalkundea eta erabilera.

Hizkuntza-aniztasuna da Europaren kultura-ondasun aberats eta garrantzitsuenetakoa (60 bat hizkuntza ditu). Europaren hizkuntza-aniztasuna haren arrakasta sozialaren ezinbesteko parte ere bada.ⁱⁱ Hiztun askoko hizkuntzek eutsiko diote suspertzen ari den gizartean eta merkatu digitalean duten tokiari, dudarik gabe, baina baliteke Europako hizkuntza asko komunikazio digitaletatik baztertuta geratzea eta garrantzia galtzea Interneteko gizartearen begietara. Hori ez litzateke batere ona izango. Alde batetik, aukera estrategiko bat galduko litzateke, eta horrek ahuldu egingo luke Europaren posizioa munduan. Bestetik, gertaera horiek ez datoz bat Europako herritar guztien (edozein hizkuntza izanda ere) berdintasunezko parte-hartzea bermatzeko helburuarekin. UNESCOk eleaniztasunaren inguruan eginiko txostenak dioenez,

hizkuntza funtsezko bitartekoa da oinarritzko eskubideez gozatzeko —adibidez, adierazpen politikoa, hezkuntza eta gizarteko parte-hartzea—.iii

Hizkuntza-teknologia teknologia bideratzaile giltzarria da

Lehen, hizkuntzaren irakaskuntzara eta itzulpenetara bideratzen ziren inbertsioak. Adibidez, kalkulu batzuen arabera, itzulpen, interpretazio, software-lokalizazio eta webgune-globalizazioaren merkatu europarra 8,4 mila milioi eurokoa zen 2008an, eta urtean % 10 haztea espero zen.^{iv} Alabaina, merkatu horren ahalmena ez da nahikoa oraingo eta geroko beharrak asetzeko.

Hizkuntza-teknologia teknologia bideratzaile giltzarria da, Europako hizkuntzak babestu eta bultzatu ditzakeena. Hizkuntza-teknologiak laguntza ematen dio jendeari elkarlanean aritzeko, negozioak egiteko, ezagutza besteekin banatzeko, eta eztabaida sozial eta politikoetan parte hartzeko, dena delako hizkuntza-mugak eta informatikako trebetasunak izanda ere. Hizkuntza-teknologiak laguntzen digu jada eguneroko lanetan, hala nola mezu elektronikoa idaztean, lineako ikerketa bat egitean edo hegaldi bat erreserbatzean. Eragiketa hauek egitean ere hizkuntza-teknologiaz baliatzen gara:

- ❑ Interneteko bilatzaile baten bidez informazioa aurkitzen dugunean.
- ❑ Testu-prozesadore batean ortografia eta gramatika egiaztatzen dugunean.
- ❑ Lineako denda batean produktu baten gaineko gomendioak begiratzen ditugunean.
- ❑ Nabigazio-sistema baten ahozko jarraibideak entzuten ditugunean.
- ❑ Lineako zerbitzu baten bidez web-orriak itzultzen ditugunean.

Lan honetan agertzen diren hizkuntza-teknologiak etorkizuneko aplikazio berritzaileen oinarritzko osagaia dira. Hizkuntza-teknologia, normalean, teknologia bideratzailea izaten da, eta aplikazio-plataforma handiago baten barruan joan ohi da, nabigazio-sistema edo bilatzaile baten barruan adibidez. Liburu zuri honetan aztertzen da teknologia komunak hizkuntza bakoitzerako zenbateraino dauden prestatuta.

Laster behar izango dugu Europako hizkuntza guztietarako hizkuntza-teknologia bat, eskuragarri dagoena, modu onean eros daitekeena eta software-esparru handiagoetan ondo integratuta dagoena. Erabiltzaileak ez du funtzio interaktiboak, multimediakoak eta eleaniztunak erabiltzerik hizkuntza-teknologiarik gabe.

Hizkuntza-teknologiaren aukerak

Hizkuntza-teknologiak aukera eman diezaieke Europako hizkuntza guztiei itzulpen automatikoak egiteko, edukiak sortzeko, informazioa prozesatzeko eta ezagutzak kudeatzeko. Hizkuntza-teknologiak balio dezake, halaber, etxetresna elektronikoa, aparailu, ibilgailu, ordenagailu eta robotentzako hizkuntzan oinarritutako interfaze intuitiboak garatzeko. Hainbat prototipo atera diren arren, aplikazio komertzial eta industrialak hasierako fasean daude oraindik. Ikerketan eta garapenean egin berri diren lorpenek benetako abagunea eman diote. Adibidez, itzulpen automatikoak (IA) zehaztasun handi samarra ematen du jada esparru jakin

batzuetan, eta aplikazio esperimentalek informazio eleaniztuna eta ezagutza-kudeaketa eskaintzen dute, baita Europako hizkuntza askotan edukiak sortzeko aukera ere.

Hizkuntza-aplikazioak, ahots bidezko erabiltzaile-interfazeak eta elkarrizketa-sistemak oso eremu espezializatuetan egon ohi dira eskuarki, eta errendimendu mugatua izan ohi dute maiz. Hizkuntza-teknologia hondamen-eremuetako erreskate-lanetarako erabiltzea ari dira ikertzen orain. Arrisku handiko inguru horietan, itzulpenaren zehaztasuna hil ala biziko kontua izan daiteke. Antzekoa gertatzen da osasungintzan ere. Hizkuntzarteko ahalmenak dituzten robot inteligenteek biziak salba ditzakete.

Hizkuntza-teknologiek sekulako merkatu-aukerak dituzte hezkuntzan eta entretenimenduen industrian; izan ere, jokoetan, joko hezigarrietan, simulazioetan eta prestakuntza-programetan integra daitezke. Hizkuntza-teknologiak zeregin garrantzitsua izan dezake beste hainbat tokitan ere; besteak beste, mugikorretako informazio-zerbitzuetan, ordenagailuz lagundutako hizkuntza-ikaskuntzarako softwarean, Internet bidezko ikaskuntzako inguruneetan, autoebaluazio-tresnetan eta plagioak aurkitzeko gailuetan. Gizarte-hedabideen aplikazioek (adib. Twitter eta Facebook) duten arrakastak iradokitzen du gero eta gehiago behar direla hizkuntza-teknologia sofistikutuak, gai direnak mezuak behatzeko, eztabaidak laburbiltzeko, iritzi-joerak iradokitzeko, erantzun emozionalak detektatzeko, copyrightaren arau-haustekak identifikatzeko eta erabilera desegokien jarraipena egiteko.

Hizkuntza-teknologiak aukera paregabea ematen dio Europar Batasunari, bai ekonomia aldetik, bai kultura aldetik. Eleaniztasuna arau bilakatu da Europan. Europako negoziokoak, erakundeak eta eskolak ere nazioartekoak eta askotarikoak dira. Herritarrek elkarrekin komunikatu nahi dute Europako Merkatu Batuan, oraindik ere hor dauden hizkuntza-mugez harago. Hizkuntza-teknologiak hor jarraitzen duten muga horiek gaitzen lagundu dezake, eta, era berean, hizkuntzaren erabilera askea eta irekia bultzatu. Gainera, Europako hizkuntzarako hizkuntza-teknologia eleaniztun eta berritzaileak munduko beste herrialdeekin eta haietako komunitate eleaniztunekin komunikatzen lagunduko liguke. Hizkuntza-teknologiek nazioarteko ekonomia-aukera ugari ematen dituzte.

Hizkuntza-teknologiaren erronkak

Azken urteotan hizkuntza-teknologiak aurrerapen handi samarra egin badu ere, aurrerapen teknologikoa eta produktuen berrikuntza erritmo motelean doaz gaur egun. Ezin dugu hamar edo hogeitaz urtez itxaron gure inguru eleaniztuneko komunikazioa eta produktibitatea areagotuko duten hobekuntza nabariak agertu arte.

Erabilera handiko hizkuntza-teknologiak —hala nola testu-prozesadoreetako ortografia- eta gramatika-zuzentzaileak— hizkuntza bakarrean izaten dira normalean, eta hizkuntza gutxi batzuetan bakarrik egon ohi dira eskuragarri. Komunikazio eleaniztunerako aplikazioek sofistikazio-maila bat eskatzen dute. Itzulpen automatikoa eta lineako zerbitzuak —esaterako, Google Translate edo Bing Translator— apartak dira dokumentu baten edukien gutxi gorabeherako itzulpenetarako. Baina, lineako zerbitzu eta IA aplikazio profesional horiek hainbat zailtasun izaten dituzte oso itzulpen zehatzak eta osatuak behar direnean. Okerreko itzulpen barregarri ezagunak asko dira (hor ditugu, esaterako, *Bush* edo *Kohl* izenen itzulpen literalak), eta agerian uzten dute zernolako erronkei egin behar dien aurre hizkuntza-teknologiak.

Hizkuntzaren jabetzea

Ordenagailuek hizkuntza nola tratatzen duten eta hizkuntzaren jabetzea horren zaila zergatik den azaltzeko, ikus dezagun gizakiok nola jabetzen garen lehen eta bigarren hizkuntzez, eta, gero, itzulpen automatikoko sistemen funtzionamenduaren eskema egingo dugu —zerbaitengatik du hizkuntza-teknologiaren alorrak horren lotura estua adimen artifizialaren arloarekin—.

Gizakiak bi modutan jabetzen dira hizkuntza-gaitasunez. Hasieran, hizkuntza bateko hitzunen arteko elkarreragina entzunez ikasten du umeak hizkuntza hori. Hizkuntzaren erabiltzaileek —gurasoek, anai-arrebek edo beste senide batzuek, esaterako— eraturako hizkuntza-adibide zehatzak entzuteak lehen hitzak eta esaldi laburrak esaten laguntzen die bi urte inguruko umei. Hizkuntzak ikasteko antolaketa genetiko bereziak eman digu gizakioi gaitasun hori.

Normalean, bigarren hizkuntza ikasteak ahalegin askoz handiagoa eskatzen du, umea ez baitago bertako hitzunen hizkuntza-komunitate baten barruan. Eskola garaian, atzerriko hizkuntzez jabetzeko, haien egitura gramatikala, hiztegia eta ortografia ikasten dira liburuetatik eta ikasmaterialetatik, eta, haietan, arau abstraktu, taula eta adibidezko testuen bidez azaltzen da hizkuntza. Atzerriko hizkuntza bat ikasteak denbora asko eta ahalegin handia eskatzen ditu, eta gero eta zailagoa da adinean aurrera egin ahala.

Hizkuntza-teknologiaren bi sistema-mota nagusiak gizakien antzera jabetzen dira hizkuntza-ahalmenez. Metodo estatistikoan, hizkuntza bakarrean idatzitako adibidezko testu zehatzen bildumetatik edo bi hizkuntza edo gehiagoko testu paralelo deritzenetatik lortzen da hizkuntza-ezagutza. Ikaskuntza automatikoko algoritmoen nolabaiteko hizkuntza-gaitasunak adieraz dezake hitzak, esaldi laburrak eta esaldi osoak zuzentasunez nola erabili hizkuntza batean edo nola itzuli hizkuntza batetik bestera. Metodo estatistikoetarako behar den esaldi-kopurua ikaragarria da. Lanaren kalitatea handiagotu egiten da zenbat eta testu gehiago aztertu. Milioika esaldiko testuen gainean probatzen dituzte maiz sistema horiek. Horregatik ibiltzen dira bilatzaileen hornitzaileak ahalik eta idatzizko material gehien bildu nahian. Testu-prozesadoreen ortografia-zuzentzaileek, linean eskuragarri dagoen informazioak eta Google Search eta Google Translate bezalako itzulpen-zerbitzuek metodo estatistikoa (datuek gidaturikoa) dute oinarrian.

Erregeletan oinarritutako sistemak dira bigarren hizkuntza-teknologia nagusia. Hizkuntzalaritzako, hizkuntzalaritza konputazionalako eta informatikako adituek azterketa gramatikalak (itzulpen-arauak) kodetu eta hiztegi-zerrendak (lexikoiak) osatzen dituzte. Erregeletan oinarritutako sistema ezartzeko, denbora eta esku-lan asko behar da. Era berean, oso aditu espezializatuak behar dira halako sistemak sortzeko. Erregeletan oinarritutako itzulpen automatikoko sistema garrantzitsuenetako batzuek etengabeko garapena izan dute azken hogeitau urteotan. Erregeletan oinarritutako sistemen alde ona da adituek kontrol zehatzagoa lor dezaketela hizkuntzaren prozesamenduaren gainean. Hortaz, softwareko akatsak sistematikoki zuzendu daitezke eta *feedback* zehatza eman dakioke erabiltzaileari, erregeletan oinarritutako sistemok hizkuntza ikasteko erabiltzen direnean batik bat. Finantza-sarrera mugatuak direla bide, erregeletan oinarritutako hizkuntza-teknologia hizkuntza handietan bakarrik erabil daiteke.

Euskara Europako informazio-gizartean

Datu orokorrak

Euskara, Nafarroako Erresumako hizkuntza nagusia zelako latinez “Lingua Navarrorum” esaten zitzaiona, mendebaldeko Europan bizirik dagoen hizkuntza preindoeuropar bakarra da. Hizkuntza bakartutzat jotzen da, ez baitaio loturarik aurkitu beste hizkuntzekin, antzinako akitanierarekin izan ezik. Euskararen jatorria nahiz beste hizkuntzekiko duen lotura gai gatazkatsuak eta interesgarriak dira oraindik ikerlarientzat.

Euskara, gaur egun, eskualde txiki batean hitz egiten da, Pirinioen mendebaldean, Espainiaren eta Frantziaren arteko mugaren bi aldeetan, euskaldunek *Euskal Herria* deritzen eskualdean. Hizkuntza lurra galduz joan da hainbat mendez, hegoaldean batez ere. Duela gutxi, Francoren diktadura garaian, euskara erabiltzea debekatu zutela eta, hizkuntzak berreskuratu ezinezko galera izan zuen. Ikaragarriko ahaleginak egin ziren hizkuntza biziberritzeko; batez ere, 1960ko hamarkadan, ikastolen sorrerari esker euskara hezkuntza-sisteman sartu zenean; klandestinitatean hasierako urteetan. Alabaina, euskararen berreskuratze-prozesua ez zen hasi 1980ko hamarkadara arte, autonomiak sortu eta Eusko Jaurlaritzari hizkuntzaren gaineko eskumen politikoak eman zitzaizkion arte.

Ahalegin ikaragarriak egin ziren arren, euskara hizkuntza “ahul” moduan agertu zen 2009an Unescok Munduko Arriskupeko Hizkuntzen Mapan^v atera zuenean. Gaur egun, kalkulatu da Euskal Herriko biztanleriaren^{vi} % 26 inguruk hitz egiten duela euskara, Espainiaren administraziopeko aldean nahiz Frantziaren administraziopeko aldean, baina bietan ez du estatus bera. Alde batetik, Euskal Herriaren Espainiako partea bi eskualde politikotan banatuta dago: Euskal Autonomia Erkidegoan, euskara koofiziala da gaztelaniarekin batera, baina gaztelaniaren alderako zenbait desberdintasunekin; Nafarroako Foru Erkidegoan, hiru eremu daude, euskararen legezko estatusaren arabera: euskalduna, erdalduna eta mistoa. Hizkuntzarekiko atxikimendua eta hizkuntza-eskubideak ezberdinak dira hiru eremu horietan. Bestetik, Frantziako aldean, Pirinio Atlantikoetako Departamentuaren mendebaldeko partean hitz egiten da euskara, baina inoiz ez du izan inolako legezko estatusik eta ez da ofiziala inongo erakundetan. Dena dela, duela urte batzuk (2004an), erakunde publiko bat sortu zen Iparraldean euskara bultzatzeko helburuarekin.

Ahozko euskarak oso dialekto-sakabanaketa handia du. Gaur egun, onartzen da sei euskalki daudela, elkarren artean alde nabariak dituztenak. Euskara batua ez zen ofizialki ezarri 1968. urtera arte, orduan egin baitzuen Euskaltzaindiak^{vii} lehen estandarizazio-proposamena. Euskalkiok ezberdintasun nabariak dituzte hainbat alderditan: hiztegia, fonetika, morfofonologia eta prosodia, azentua eta intonazioa. Euskalkiak ez dira entitate homogeenak; etengabe aldatzen dira batetik bestera, eta, batzuetan, ez dago hain argi bi edo hiruren arteko muga.

Euskararen berezitasunak

Euskara hizkuntza eranskaria eta flexio handikoa da, eta hizkuntza ergatibo-absolutiboa izatea da bereizgarri nagusia. Horrek esan nahi du aditz iragangaitz baten subjektua absolutibozko kasuan

(markatu gabekoan) joan ohi dela, eta kasu hori bera erabiltzen dela aditz iragankorren objektu zuzenarentzat; aditz iragankorren subjektua beste era batean markatzen da, ergatibozko kasuaren bidez: *-k* atzizkia.

Euskara postposiziozko hizkuntza da; beraz, kasuzko eta postposiziozko esaldiak sintagmaren amaieran atzizki bat edo gehiago gehituta eraten dira, eskema honen arabera:

erroa + (artikulua) + (numeroa) + [kasua(k)]

Adibidez, “mutilarenagana” honela dago osatuta: «mutil+a+Ø+r+en+gan+a», —“mutil” lema edo izen-erroa da; “a”, artikulua; “Ø”, singularreko marka; “r”, epentesi-partikula; “en”, edutezko genitiboa; “gan” izaki bizidunen marka; eta “a”, adlatiboa—.

Ezaugarri garrantzitsua da hori, hizkuntza naturalean eta hizketa-prozesamenduan kontuan hartu beharrekoa; bada, izen-sintagma bakoitzak 17 deklinabide-kasu izan ditzake, eta lau aldiz forma gehiago har ditzake zehaztasunaren eta numeroaren arabera. Hasierako 68 forma horiek are gehiago alda daitezke esaldiko beste zati batzuen arabera —hango izenen arabera ere deklinatzen direlako—. Kalkulatzen da bi mailatako errekurtsioarekin euskarazko izen batek 275 deklinabide-marka izan ditzakeela, eta hori oso ohiko fenomeno^{viii} da, gainera. Horrek aditzera ematen du beharrezkoa dela hainbat eratako bukaera horiek guztiak tratatzeko modu bat aurkitzea, oinarrizko hiztegi batetik abiatuta.

Aditzak dira euskara hizkuntza eranskaria dela erakusten duen beste adibide bat. Aditz laguntzailea aditz nagusi gehienekin batera joan ohi da, eta, subjektuarekin ez ezik, esaldian dagoen beste edozein objektu zuzen edo zeharkakorekin ere komunztatzen da. Europako hizkuntzen artean, pertsona askorekiko komunztadura hori euskarari, Kaukasoko hizkuntza batzuetan eta hungarieran baino ez da topatu (guztiak ez-indoeuroparrak). Euskarari, aditzen eskema honi jarraitzen zaio:

[aditz-erroa + aspektu-marka] [aditz laguntz.]

Adibidez, euskara batuan, «esaten zenizkidaten» honela dago osatuta: «esan» (aditz-erroa) + «ten» (maiztasun-aspektua) eta «zen+i+zki+da+Ø+te+n» aditz laguntzailea («zen» bigarren pertsonaren ergatibo-marka da; «i», aditz laguntzailearen erroa; «zki», hirugarren pertsona pluralaren absolutibo-marka; «da», lehen pertsona singularraren datibo-marka; «Ø», indikatibo-marka; «te», pluralaren ergatibo-marka; eta «n», iraganaldiko marka). Aditza horren konplexua izanda, aditz laguntzaile bakoitza morfemetan banatu beharrean osorik tratatu ohi da hizkuntza naturalaren prozesamendurako ikerketetan.

Esaldiko hitzen hurrenkerari dagokionez, oinarrizko eraikuntza sintaktikoa subjektua-objektuak-aditza da (gaztelanian, frantsesean eta ingelesean, aldiz, subjektua-aditza-objektuak da ohikoena). Esaldi barruko sintagmen hurrenkera alda daiteke mintzagaiaren arabera, baina sintagma barruko hitzen hurrenkera zurrunda da normalean. Bada, argitu behar da euskararen sintagma-hurrenkera mintzagaia-galdegia dela; hau da, esaldi neutroetan (norbaiti gertaera baten berri emateko esaldiak, kasurako), mintzagaia ematen da lehenik, eta galdegia ondoren. Halako esaldietan, aditz-sintagma amaieran joan ohi da. Laburbilduz,

galdegaia aditz-sintagmaren aurre-aurrean doa. Galderetan ere betetzen da arau hori; esaterako, «Zer da hau?» edo «Hau zer da?» esan daiteke, baina bietan «zer» galdetzailea «da» aditzaren aurre-aurrean doa. Arau hori horren garrantzitsua da euskaran, non beste hizkuntzetan ematen diren euskararen deskribapen gramatikaletan ere euskarazko hitza (*galdegai*) erabiltzen baita.

Euskararen ortografia fonemikoa da ia erabat: grafema bakoitza fonema bati dagokio, eta, beraz, hitz baten ahoskera erraz iragar daiteke idatzizko formatik. Dena dela, badaude salbuespen gutxi batzuk: <l> eta <n> letrek aurretik <i> letra eta atzetik bokal bat badute, bustidura egin ohi da; adib. mutila => <mutila>. Beste adibide bat "ez" ezezko partikularen amaierako kontsonante-fonemarena da, atzetik datorren fonema ahoskabe bihurtzen baitu; adib. ez dira => <eztira>.

Azken gertaerak

Euskaltzaindiak, euskal hizkuntzaz arduratzen den erakunde ofizialak, euskararen forma estandarizatua atera zuen 1960ko hamarkadan, euskara batua izenekoa. Euskara egoera formaletan (hezkuntzan, hedabideetan, literaturan...) eta euskal hiztun guztiek ulertzeko eran erabil zedin asmatu zen euskara batua, eta gaur egun ere horretarako erabiltzen da batik bat. Literatura klasikoko tradizioa dela eta, euskara batua erdialdeko euskalkian eta nafar-lapurteran oinarritzen da nagusiki. Ertzetako euskalkiak oso ezberdinak dira, nahiz eta mendebaldekoa hiztun gehien dituenetako bat izan erdialdekoarekin batera.

Euskara batuak oinarri sendoak ditu, eta aurrerapausoak egiten ari da sintaxiaren eta naturaltasunaren aldetik. Gaur egun, euskara ikasten duten ia guztiek euskara batua ikasten dute. Horren ondorioz, fenomeno berezi bat gertatu da Euskal Herri osoan: euskaldun zaharrak beren euskalkian mintzatzen dira herriko jendearekin eta euskara batuan euskaldun berriekin. Mendebaldean, hango euskalkiaren eta batuaren arteko desberdintasun nabariak direla eta, euskara-ikasleei iruditzen zaie ikasten ari diren hizkuntza oso urrun dagoela jendearen ahotan dabilen euskara horretatik. Bestalde, jada badaude euskara batua ama-hizkuntzatzat duten euskaldun batzuk, euskaldun berri askok beren umeei euskaraz egitea erabaki baitute, nahiz eta beren lehen hizkuntza gaztelania izan.

Hala eta guztiz ere, euskararen teoriariek^{ix} gero eta argiago ikusten dute, euskararen geroa berrmatuko bada, beharrezkoa dela euskara batua ez ezik egungo euskalkiak ere sustatzea. Hortaz, euskalkiek izango dute nolabaiteko garrantzia euskaraz eskainiko diren HTko aplikazioetan.

Euskal hizkuntza-teknologiaren komunitatea eta ikerlariak ohartu dira zer-nolako garrantzia duten teknologiek hiztun gutxiko hizkuntza batek XXI. mendean aurrera egin dezan, eta sekulako ahaleginak egin dituzte euskara gehien erabiltzen diren hizkuntzen maila berean jartzeko teknologia aldetik. Eskarmentu zientifiko sendoa du euskarak, bai eta aldameneko zenbait hizkuntzek ere, hala nola katalanak eta galizierak; hori ez da Europar beste inon gertatu, ez eta eskualdeko hizkuntza batzuek hizkuntzarteko produktu eta zerbitzuak garatzea ere.

Argi dago zeinen garrantzitsua den euskararako hizkuntza-teknologiaren industria garatzea, eta horrexegatik sortu da *Langune*^x elkartearen ere. Langune hizkuntzen industriaren alorreko Euskal Herriko enpresen elkarte da. Elkartearen sortu zen,

eta itzulpengintzaren, edukien, irakaskuntzaren eta hizkuntza-teknologiaren alorreko 30 enpresa baino gehiago biltzen ditu. Languneren helburu nagusia da hizkuntza-teknologiaren sektorea garatzea, eta erreferentzia-puntu bilakatzea Europako hizkuntzen industriarentzat, ahaleginak biderkatu gabe eta sinergiak lortuz. Langune hasi besterik ez da egin, baina urrats ikaragarriak ari da egiten.

Hizkuntza-lanketa

Euskararen ordezkari nagusia Euskaltzaindia da, euskal hizkuntzaren akademia ofiziala (1919). Hizkuntza ikertzen du, babesten saiatzten da eta erabilera-arauak ezartzen ditu. Onarpen ofizial osoa du Espainian (1976), eta onura publikorako kultura-elkartetzat onartzen da Frantzian (1995).

Euskal Autonomia Erkidegoan euskara hizkuntza ofizial deklaratu zenetik, Eusko Jaurlaritzak hainbat arau eta lege egin ditu euskararen erabilera babesteko eta bultzatzeko. Hainbat erakunde sortu dira harrezkero: Euskararen Aholku Batzordea (1982), EiTb (Euskal Irrati Telebista, 1982), HABE (Helduen Alfabetatze eta Berreuskalduntzerako Erakundea, 1983) eta beste hainbat.

Euskara Biziberritzeko Plan Nagusia (EBPN) 1998an jarri zen abian, tresna estrategiko moduan eta hiru helburu nagusirekin: adostasun batera iritsi erakundeek xede eta ekintzen artean, eratze-programetarako lehentasunak ezarri eta euskararen alde lan egiten duten erakundeek, enpresen eta elkarteek jarduerak koordinatu. Plan estrategiko horren barruan, aldi behin egiten diren inkesta soziolinguistikoak baliagarriak dira beste helburu batzuk eta zuzenketa-ildoak ezartzeko. Eusko Jaurlaritzak badu euskarari buruzko web-atari bat (www.euskara.euskadi.net), eta, han, hizkuntzari, haren historiari eta gaur egungo egoerari buruzko informazioa ez ezik, hizkuntzarekin lotutako era guztietako zerbitzu, produktu eta aplikazioetarako estekak ere badaude —eratzeko-programa publikoetarakoak barne—. Frantziako aldean, “Euskararen Erakunde Publikoa”^{xi} 2004an sortu zen, interes publikoko elkarte moduan, lau erakunde publiko —herri- edo eskualde-erakundeak— eta estatua elkartuta, eta eskualdean hizkuntza-politika bateratua sortu eta ezartzeko asmoz.

Hizkuntza hezkuntzan

Euskal Autonomia Erkidegoan, 1983an sartu zen euskara hezkuntza-sistema publikoan, Lehen eta Bigarren Hezkuntzan euskararen eta gaztelaniaren erabilera arautzen duen legearekin. Lehen eta Bigarren Hezkuntzarako, hiru eredu sortu ziren, eta ikastetxe bakoitzari aukeran eman zitzaion zein eredu eskaini. A ereduan, komunikazio-hizkuntza gaztelania da, eta euskara “Euskal hizkuntza eta literatura” irakasgai ematen da. D ereduan —euskaraz, *c* letra ez da erabiltzen, normalean—, euskara da komunikazio-hizkuntza, eta ikasgai bat ematen da gaztelaniaz, “Gaztelania eta literatura”. B ereduan edo tarteko ereduan, ikasgai batzuk gaztelaniaz ematen dira (batez ere, irakurketa eta idazketa eta matematika) eta beste batzuk euskaraz (nagusiki, zientziak eta plastika). Alabaina, A eredua gero eta ikasle gehiago galduz joan zen, eta B eredua gero eta ikasle gehiago hartuz, Haur Hezkuntzan eta Lehen Hezkuntzan batik bat —ikasleen erdiak baino gehiagok D ereduan ikasten du aldi horietan—. Dena den, 15 urteko ikasleen % 85ek gaztelaniaz egin zituen PISA programaren azterketak^{xii}, eta % 15ek bakarrik egin zituen euskaraz; horrek argi uzten du gaztelania dela hizkuntza nagusia hezkuntzan.

Nafarroako Erkidegoan, euskarak hainbat mailatako estatus ofiziala duen tokian, laugarren eredu bat ere jarri zuten, euskara derrigorrezko ikasgai gisa ere eskaintzen ez zuena. Iparraldeko probintziei dagokionez, euskarazko Lehen Hezkuntza eskola-sare pribatu batek ematen du, Seaskak, eta, gaur egun, 2.700 bat ikasle ditu 29 ikastetxetan —Bigarren Hezkuntzako ikastetxe bat eta lizeo bat bakarrik daude—.

Azkenaldian, eredu berriak ari dira proposatzen eta probatzen, ingelesaren ikasketa goiztiarrari garrantzia ematen dioten ereduak. Eusko Jaurlaritzak eredu hirueleduna jarri du martxan duela gutxi, eta, Nafarroan, berriz, gaztelaniaz eta ingelesez ematen den hezkuntza elebiduna jarri dute, nahiz eta euskara aukeran eskaintzen duten.

Hezkuntza-maila handiagoetan, gaztelania da nagusi, ezbairik gabe. Hiru unibertsitate daude, eta publikoa, bakarra: Euskal Herriko Unibertsitatea (EHU). EHUK euskaraz ikasteko aukera ematen du, eta, euskaraz eta gaztelaniaz eskaintza bera egiteko sekulako ahaleginak egin diren arren, oso gradu gutxi ikas daitezke euskara hutsean. Nabarmentzekoa da Hizkuntzaren Azterketa eta Prozesamendua^{xiii} master eta doktorego-programa euskara hutsean ematen dela 2001. urtetik. Mondragon unibertsitate pribatuak euskaraz eskaintzen ditu gradu gehienak, eta master batzuk ere bai. Hirugarren unibertsitateak, Deustuko Unibertsitateak, ikasketa batzuk baino ez ditu eskaintzen euskaraz.

Nazioartean

2009ko urtarriletik, Etxepare Euskal Institutua da euskal hizkuntza eta kultura mundu osoan zabaltzeaz arduratzen den euskal erakunde publikoa. Institutu horren asmoa da euskararen irakaskuntza, ikaskuntza eta erabilera bultzatzea mundu osoan, eta euskara hizkuntza komuntzat daukaten komunitate guztien ekarpenak biltzen ditu. Institutuaren xedea da, halaber, euskal kultura nazioarteko komunitatean barreiatzea, euskara hitz egiten duten talde horiei erreferentzia berezia eginez, euskal diaspora barne. Historian, euskaldun askok alde egin dute Euskal Herritik munduko beste txoko batzuetara, arrazoi ekonomiko eta politikoak direla eta; euskal diaspora izena jarri zitzaion aberritik kanpo bizi den euskal jatorriko jendeari. Gaur egun, euskal jatorriko jende dezente bizi da Txilen, Argentinan, Bolivian, Ekuadorren, Kolonbian, Kuban, Mexikon, Venezuelan, Kanadan eta Estatu Batuetan. Haietan guztietan, euskal kulturaguneak daude, Euskal Etxeak, helburu bera lortzeko sortutakoak: euskal kultura eta nortasunari eustea. 24 herrialdetako hiri handi gehientsuenetan daude Euskal Etxeak^{xiv}.

Euskararen jatorriak eta egitura bereziak euskal hizkuntza eta kultura ikasteko interesa piztu dute. Gaur egun, Amerikako eta Europako 13 herrialdetako 29 unibertsitatetan ikas daiteke.

Nazioarteko erakundeetan duen tokiari dagokionez, Espainiako gobernuak saiatu da Europako erakundeen hizkuntza ofizialetan euskara sartzen, katalanarekin eta galegoarekin batera. Baina, gaur egun, ez dira hizkuntza ofizialtzat jotzen; erdiofizialak dira, eskoziera, gaelikoa eta galesarekin batera. Euskara oso leku gutxitan erabil daiteke: Eskualdeetako Lantaldearen eta Kontseiluaren saioetan hitz egin daiteke, baina Europar Legebiltzarraren osoko bilkuretan, ez. Herritarrek eskubidea dute, halaber, Europako erakundeei euskaraz idazteko eta erantzuna hizkuntza berean jasotzeko, baina Espainiako Gobernuaren bitartez egin behar dute beti, eta hark ordaindu behar ditu gastuak.

Euskara sartuta dago Europar Batasuneko Eskualdeetako Hizkuntzen eta Hizkuntza Txikien Zerrendan^{xv}, eta, beraz, jasotzen du laguntza Europar Legebiltzarrak eskualde-hizkuntzetako eta hizkuntza txikietako ekintzak sustatzeko egindako ebazpenetatik.

Hizkuntza-teknologiak beste ikuspegi batetik egin diezaioke aurre erronka horri, atzerriko hizkuntzako testurako itzulpen automatikoa edo hizkuntzarteko informazio-berreskurapena bezalako zerbitzuak eskainiz, eta, hala, lagundu egin dezake berezko ingeles-hiztunak ez direnek dituzten desabantaila pertsonal eta ekonomikoak murrizten.

Euskara Interneten

2010. urtearen lehen hiruhilabetekoan, Euskal Herriko etxeen % 61,4tan (513.000) ordenagailua zegoen. 460.000 familia baino gehixeago zeuden, eta horietatik % 54,9k Interneterako sarbidea zuten beren etxeetan. Horrenbestez, 15 urte edo gehiagoko milioi bat lagun baino gehiago Internet-erabiltzaileak ziren. Gehienek esan zuten egunero konektatzen zirela. % 22,9k bakarrik erabiltzen zuten euskara Interneten.^{xvi} Hala eta guztiz ere, euskaldunen artean Internet-erabiltzaileen talde sendo eta gogotsu bat dago. Euskarazko blogek, euskarazko Wikipediak nahiz lineako zerbitzuek eta doako softwarean oinarritutako sistema eragile eta tresnen kokapenak euskara eta euskal kultura Interneten nahiz IKTetan egotearen aldeko apustua egin dute, eta, hala, euskara zabaldu dute. Esaterako, Euskal Wikipediak 97.000 artikulua baino gehiago ditu; wikipedia guztien artean artikulua gehien dituen 39.a da. Eta ahalegin handiak egin dira software-programa^{xvii xviii} eta baliabide arruntak^{xix xx xxi} euskaraz eskaintzeko.

Lehen mailako domeinu berri bat erregistratu da, .eus, eta 2012. urtean jarriko da abian. Aurretiko izen-emateak 193 dira jada. Proposaturiko .eus domeinua euskal hizkuntza eta kulturaren komunitatea Interneten ordezkatzeko duen izena izango da. Ikur hori euskal kultura eta euskara sustatzeko tresna bihurtuko da, eta, alde horretatik, .eus domeinua mekanismo eraginkorra izango da euskara mundu osoan estandarizatzeko. .eus domeinuak, Interneteko toki birtualean, euskara modu eraginkorrean sustatzen dela ziurtatuko du, eta, era berean, nazioartean onespina bermatuko du. Era berean, .eus domeinuak Interneten kulturantzatasuna indartu eta zabalduko du, hizkuntza- eta kultura-komunitateei beren domeinua izaten uzteak Interneten bihotz-bihotzean jartzen baitu kulturantzatasuna. Hizkuntzarekin eta kulturekin zerikusia duten domeinuak indargarri eta onuragarri dira hizkuntza- eta kultura-komunitate horientzat, baina Internetentzat berarentzat ere bai.^{xxii}

Hizkuntza-teknologiarentzat garrantzitsua da Internet gero eta indartsuagoa izatea, bi arrazoirengatik. Batetik, eskuragarri dauden hizkuntzari buruzko datu digitalizatuak iturri aberatsa dira hizkuntza naturalaren erabilera aztertzeko, informazio estatistikoa bilduz nagusiki. Bestetik, hizkuntza-teknologia erabiltzen duten era askotako aplikazio-eremuak eskaintzen ditu Internetek.

Irakurgai gehiago

HUALDE, J.L.; LAKARRA, J.; TRASK, R. L.: *Towards a History of the Basque Language*, John Benjamins Publishing Company, 1995.

HUALDE, J.L.; ORTIZ DE URBINA, J.: *A Grammar of Basque*, Mouton de Gruyter, Berlin, 2003.

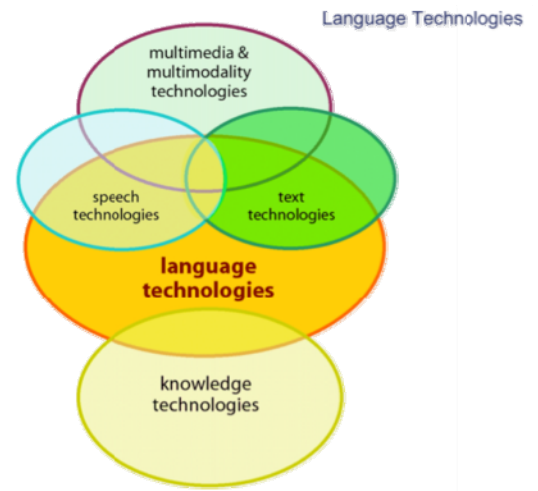
ZUAZO, K.: *El euskera y sus dialectos*, Alberdania, 2010.



Hizkuntza-teknologia euskararako

Hizkuntza-teknologiak

Hizkuntza-teknologiak giza hizkuntzarekin lan egiteko espezializatutako informazio-teknologiak dira. Horregatik, giza hizkuntzaren teknologia izenpean ere ezagutzen dira maiz teknologia hauek. Giza hizkuntza ahoz nahiz idatziz agertzen da. Hizkuntza-komunikazioko modurik zaharrena eta naturalena hizketa bada ere, informazio konplexua eta giza ezagutzaren zati handiena testu idatzien bidez gorde eta transmititzen da. Hizketa- eta testu-teknologiek bi modu horietan prozesatzen edo sortzen dute hizkuntza. Baina, hizkuntzak baditu hizketan nahiz testuetan agertzen diren alderdiak ere, hala nola hiztegiak, gramatikaren zati handi bat eta esaldien esanahia. Hortaz, hizkuntza-teknologiako atal asko ezin dira bietako batean sartu, hizketa-teknologian ala testu-teknologian. Horien artean daude hizkuntza ezagutzarekin lotzen duten teknologiak. Eskuineko irudiak hizkuntza-teknologiaren egoera irudikatzen du. Elkarrekin komunikatzen garenean, beste komunikazio-modu batzuk eta beste informazio-bide batzuk erabiltzen ditugu hizkuntzarekin batera. Hizketarekin batera, imintzioak eta aurpegierak erabiltzen ditugu. Testu digitalak irudiekin eta hotsekin lagunduta joan ohi dira. Filmetan, hizkuntza ahoz eta idatziz ager daiteke. Beraz, hizketa- eta testu-teknologiak gainjarri egiten dira eta interakzioan daude askotariko komunikazioa eta multimedia-dokumentuak errazago prozesatzeko aukera ematen duten beste teknologia askorekin.



Hizkuntza-teknologia aplikatzeko arkitekturak

Hizkuntza prozesatzeko software-aplikazio ohikoenetan, hizkuntzaren hainbat alderdi eta haien zereginak kopiatzen dituzten zenbait osagai egon ohi dira. Eskuineko irudian, testuak prozesatzeko sistema batean topa dezakegun arkitektura bat ageri da, asko sinplifikatuta. Lehenengo hiru moduluek sarrerako testuaren egitura eta esanahia hartzen dute kontuan:

- ❑ Aurretratamendua: datuak garbitu, formatua kendu, sarrerako hizkuntza detektatu eta abar.
- ❑ Anlisi gramatikala: aditza eta haren objektuak, modifikatzaileak eta abar aurkitu; esaldiaren egitura detektatu.
- ❑ Anlisi semantikoa: desanbiguazioa ("hori" hitzaren zein adiera da egokia testuinguru jakin batean?), anaforak eta erreferentziako adierazpenak (adib. "bera" edo "autoa") ulertzea; esaldi baten esanahia ordenagailuak irakurtzeko moduan eman.

Zeregin espezifikoak moduluek era askotako eragiketarako egiten dituzte, hala nola sarrerako testu baten laburpen automatikoa, datu-baseko bilaketak eta beste hainbat. Hemen behean, aplikazio-eremu komunak erakutsiko ditugu, eta eremu horietako modulu nagusiak nabarmendu. Hor ere, aplikazioen arkitekturak oso sinplifikatuta eta idealizatuta ageri dira, hizkuntza-teknologiako (HT) aplikazioak edonork ulertzeko moduan azaltzarren. Tresna eta baliabide garrantzitsuenak azpimarratuta daude testuan, eta kapitulu amaierako taulan ere ageri dira. Aplikazio-eremu komunei buruzko ataletan, dagokion euskarazko esparruan lanean ari diren industrien ikuspegi orokor bat ere ematen da.

Aplikazio-eremu komunak aurkeztu ondoren, HTen ikerketa eta hezkuntzako egoeraren azalpen labur bat emango dugu, eta, bukatzeko, egin diren eta egiten ari diren ikerketa-programen berri

emango dugu. Atal honen amaieran, adituaren ikuspuntutik HTen tresna eta baliabide komunen egoera nolako den azalduko dugu, hainbat alderditatik (eskuragarritasuna, heldutasuna eta kalitatea). Taula honek ikuspegi orokor on bat ematen du euskararako HTen egoeraren gainean.

Aplikazio-eremu komunak

Hizkuntza-zuzentzailea

Testu-prozesadore bat erabili duen edonork egin du topo **ortografia-zuzentzaile** batekin —ortografia-akatsak markatu eta zuzenketak proposatzen dituen osagaia da—. Ralph Gorin-ek ortografia zuzentzeko lehen programa asmatu zuenetik berrogei urte igaro ondoren, gaur egungo hizkuntza-zuzentzaileek ez dute, besterik gabe, erazutako hitzen zerrenda bat ortografia zuzeneko hitzen hiztegi batekin alderatzen; gero eta sofistikatuagoak dira. Morfologiarako hizkuntzaren menpeko algoritmoez gainera (adibidez, plurala egiteko), batzuk gai dira orain sintaxiarekin lotutako akatsak identifikatzeko, hala nola aditz baten falta edo pertsona edo numeroan subjektuarekin komunztatzen ez duen aditza (“Haiek gutuna idazten ari **da*”). Alabaina, eskura dauden ortografia-zuzentzaile gehienek (Microsoft Word-ekoa barne) ez dute akatsik aurkituko Jerrold H. Zar-en poema bateko (1992) lehen bertso honetan:

*Eye have a spelling chequer,
It came with my Pea Sea.
It plane lee marks four my revue
Miss Steaks I can knot sea.*

Mota horretako akatsak aurkitzeko, testuingurua aztertu beharra dago maiz, hala nola euskaraz ergatiboa erabili behar den ala ez erabakitzeke orduan:

*Liburua neskak dauka

Irakurlea neska da.*

Horretarako, bietakoren bat egin behar da: hizkuntza espezifikoko gramatika-arauak formulatu —trebetasun eta eskulan handia eskatzen du horrek— edo hizkuntza-eredu estatistiko delakoa erabili. Ereduok hitz jakin bat inguru zehatz batean (hots, aurretik eta atzetik dituen hitzak) ateratzeko zer probabilitate dagoen kalkulatzeko dute. Adibidez, *neskak dauka* hitz-sekuentzia agertzeko probabilitatea *neska dauka* agertzekoa baino askoz handiagoa da. Hizkuntza-eredu estatistikoa automatikoki atera daiteke hizkuntza-datu zuzenen kantitate handietatik (hots, corpusetatik). Orain arte, ingelesezko hizkuntza-datuetan oinarrituta garatu eta ebaluatu dira metodo horiek. Horrek ez du esan nahi, ordea, euskarara zuzenean transferitu daitezkeenik, euskarak inflexio handiagoa eta morfologia eranskaria du eta. Egia esanda, zailtasun izugarriak daude euskararako hizkuntza-ereduak sortzeko, ezinezkoa baita balizko hitz-forma guztiak biltzea.

Hizkuntza-zuzentzailea ez da testu-prozesamenduko tresnetan bakarrik erabiltzen; testuak sortzen laguntzeko sistemetan ere erabiltzen da. Produktu teknikoek gora egin duten bezalaxe, dokumentazio teknikoak ere asko ugartu da azken hamarkadetan. Enpresak, bezeroen aldetik erabilera txarragatiko kekek edo

matxuragatiko erreklamazioak (gaizki idatziriko edo ulerturiko jarraibideengatik) jasotzeko beldurrez, dokumentazio teknikoan arreta gehiago jartzen hasi dira, eta nazioarteko merkatuan sartzen ere bai. Hizkuntza naturalaren prozesamenduan egin diren aurrerapenek testuak sortzen laguntzeko softwarea ekarri dute; programa horrek dokumentazio teknikoak jartzen du idazlearen eskura, arau jakin batzuk eta (enpresaren) terminologia-murrizketa batzuk dituzten hitzak eta esaldi-egiturak erabil ditzan.

Euskararako gehien erabiltzen den zuzentzaile ortografikoa Xuxen da^{xxiii}, IXA unibertsitateko ikerkuntza-taldeak (ixa.si.ehu.es) garatu eta Eleka Ingeniaritza Linguistikoa enpresak eskaintzen duena. Zuzentzaile ortografiko hori ez da hiztegi bat erabiltzera mugatzen, ingelesean edo inflexio txikiagoko beste hizkuntza batzuetan egin ohi den moduan. Horren ordez, analisi morfologikoa egiten du. Zuzentzaile ortografiko honen bertsio berrienak gramatika eta estiloa ere zuzentzen ditu. Bertsio horretan, Hizkia^{xxiv} enpresak eta UZEI^{xxv} erakundeak garatutako kodea ere badago.

Zuzentzaile ortografikoetan eta editatzen laguntzeko programetan ez ezik, ordenagailuz lagundutako hizkuntza-ikaskuntzaren esparruan ere garrantzitsua da hizkuntza-zuzentzailea, eta sareko bilatzaileetara bidalitako dudak automatikoki zuzentzeko ere erabiltzen da; adib. Google-ren "esan nahi zenuen" iradokizunak.

Sareko bilaketak

Sareko, intranetetako edo liburutegi digitaletako, bilaketak dira gaur egun gehien erabiltzen den hizkuntza-teknologia, eta, hala ere, gutxi garatuta dago. Google bilatzailea 1998an sortu zen, eta mundu osoko bilaketa guztietatik % 80tan erabiltzen da gaur egun^{xxvi}.

Lehenengo bertsio hartatik ez da aldaketa esanguratsurik egon, ez bilaketaren interfazeaz, ez berreskuratutako emaitzen aurkezpenean. Oraingo bertsioan, Google-k gaizki idatzitako hitzak ortografikoki zuzentzen ditu, eta, 2009an, oinarritzko bilaketa semantikorako ahalmenak gehitu zizkieten bere algoritmo-nahasketari^{xxvii}; hala, bilaketa zehatzagoa egin daiteke, sartutako terminoen esanahia testuinguruan aztertzen da eta. Google-ren arrakastak erakusten du, eskura datu asko eta horiek indexatzeko teknika eraginkorrak izanda, nagusiki estatistiketan oinarritzen den metodo batek emaitza onak eman ditzakeela.

Dena dela, informazio-eskaera sofistikuagoetarako, hizkuntza-egaztuta sakonagoa integratu beharra dago. Ikerketa-laborategietan, makinak irakurtzeko moduko thesaurusak eta hizkuntza-baliabide ontologikoak —adibidez, Wordnet— erabili dituzten esperimenduek hobekuntzak izan dituzte orriak bilaketa-terminoen sinonimoen bidez bilatzeko aukerari esker. Aurrerapen horietarako ere hizkuntzaren baliabide espezifikoak behar dira. Euskal Herriko Unibertsitateko IXA ikerkuntza-taldeak EuskalWordnet (BasWN) asmatu du, eta ELRAren bitartez erosi daiteke.

Bilatzaileen hurrengo belaunaldiak hizkuntza-teknologia sofistikuagoa izan beharko du. Bilaketa-kontsulta galdera bat edo beste mota bateko esaldi bat bada —gako-hitz batzuk izan beharrez—, kontsulta horri erantzun egokiak emateko, esaldi hori sintaktikoki eta semantikoki aztertu behar da, eta indize batek egon behar du dagozkion dokumentuak azkar berreskuratuzeko. Adibidez, demagun erabiltzaile batek kontsulta hau egiten duela: “Azken bost urteotan beste enpresa batek xurgatu dituen enpresa

guztien zerrenda bat eman iezadazu". Erantzun egokia lortzeko, analisi sintaktikoa egin beharra dago, esaldiaren egitura gramatikala aztertzeko eta jakiteko erabiltzaileak bilatzen duena xurgatuak izan diren enpresak direla, eta ez beste enpresa batzuk xurgatu dituzten enpresak. *Azken bost urteotan* esamoldea ere prozesatu egin behar da, jakiteko zein urteri egiten dien erreferentzia.

Azkenik, kontsulta prozesatua egituratu gabeko datu-kantitate handi batekin lotu beharra dago, erabiltzaileak lortu nahi duen informazioa aurkitze aldera. Informazio-berreskurapena esaten zaio horri, eta dokumentu egokiak bilatu eta mailakatzen ditu. Gainera, enpresa-zerrenda bat sortzeko, dokumentu bateko hitz-kate bat enpresa-izen bat dela adierazten digun informazioa erauzi behar dugu. Era horretako informazioa entitateen izenen ezagutzaile izeneko bitartez dago eskuragarri.

Are zailagoa da bilaketa bat beste hizkuntza batean idatzitako dokumentuekin lotzea. Hizkuntzarteko informazio-berreskuratzeko, balizko iturburu-hizkuntza guztietara itzuli behar dugu kontsulta automatikoki, eta berreskuratutako informazioa helburuko hizkuntzara eraman. Testu-formatuez bestelakoetan ematen diren datuen proportzioa handitu den heinean, gero eta gehiago eskatzen dira multimedia-informazioa berreskuratzeko zerbitzuak, hots, irudi-, audio- eta bideo-datuen bilaketak. Audio- eta bideo-fitxategietarako, hizketa-ezagutze modulua izan behar da, hizketaren edukia testu edo irudikapen fonetiko bihurtzeko, eta, hala, erabiltzailearen kontsultak haiekin lotzeko.

Enpresa horien garapena interes bereziko atariei gehigarriak eta bilatzaile aurreratuak eskaintzean datza, gaiari dagokion semantika erabiliz. Oraindik ere prozesatze-indar handia eskatzen duela eta, bilatzaile horiek testu-corpus nahiko txikietan bakarrik erabil daitezke. Prozesatze-denbora bilatzaile estatistiko arrunt batena baino mila aldiz handiagoa da, gutxienez —adibidez, Google-k eskaintzen duen bilatzailearekin alderatuta—. Bilatzaile horiek eskari handia dute gai espezifikoko domeinuen modelazioan, eta ezinezkoa da mekanismo horiek web-mailan erabiltzea.

Euskal Autonomia Erkidegoan, Eleka Ingeniaritza Linguistikoa enpresa txikia buru-belarri aritu da lanean, euskararako aplikazioak eta sarean oinarritutako zerbitzuak garatzen. HTren ikerketa-emaitzak eta baliabideak integratzen dituzte normalean, hala nola IXA taldearen eta Elhuyar Fundazioaren lematizatzaileak eta datu-base lexikoak. Elebila bilatzaile eleaniztunak kontuan hartzen ditu euskararen berezitasunak, eta hainbat hizkuntza-tresna eta -baliabide integratzen ditu, kalitatezko euskarazko emaitzak lortzeko. Beste adibide bat Miatu izeneko tresna da; liburutegi bat da, eta helburu bereziarekin indexatutako datu-basean ganean lematizatzaileak eta analisi morfologikorako beste tresna batzuk erabiliz bilaketak egiteko funtzioak eskaintzen ditu. www.zientzia.net zientziarekin lotutako ataria eta www.ikasbil.net eduki pedagogikoko ataria sortzeko erabili da.

Ahozko interakzioa

Erabiltzaile bati grafikoak, teklatura eta saguaren partez ahozko hizkera erabiliz makinekin interakzioan jartzeko aukera ematen dioten interfazeen oinarrian dagoen teknologia da Ahozko Interakzioa. Gaur egun, enpresek beren bezeroei, langileei edo lankideei telefonoz eskaintzen dizkieten automatizazio partzial edo osoko zerbitzuetan erabiltzen dira, normalean, ahotsaren bidezko

erabiltzaile-interfazeak (AEI). Lan-arlo hauetan asko erabiltzen dira ahotsaren bidezko erabiltzaile-interfazeak: banketxeak, logistika, garraio publikoa eta telekomunikazioak. Ahozko interakzioaren teknologia gailu jakin batzuen interfazeetan ere egon ohi da —esaterako, autoan txertatuta doazen nabigazio-sistemetan—, eta erabiltzaile-interfaze grafikoen sarrera-irteera modalitateen ordez ahozko hizkera ezartzeko ere erabiltzen dira, Smartphone gailuetan esaterako.

Ahozko interakzioaren oinarrian, lau teknologia hauek daude:

- Hizketaren ezagutza automatikoak zer hitz esan diren ateratzen du, erabiltzaileak egindako hots-sekuentzia batetik.
- Analisi sintaktikoak eta interpretazio semantikoak erabiltzaile batek esandakoaren egitura sintaktikoa aztertzen du eta interpretatu egiten du sistemaren helburuaren arabera.
- Elkarrizketa-kudeaketa beharrezkoa da erabiltzailearen interakzioko sistema-zatian zer egin zehazteko, behin erabiltzailearen inputa emandakoan eta sistemaren funtzioak kontuan hartuta.
- Hizketaren sintesirako teknologia (TTS edo *Text-to-Speech*) esandako hitz horiek hots bihurtzeko eta erabiltzailearentzako irteera gisa emateko erabiltzen da.

Erronkarik handienetarikoa da hizketaren ezagutza automatikoko sistema batek zehaztasun osoz ezagutzea erabiltzaile batek esandako hitzak. Horretarako, bietakoren bat egin behar da: erabiltzailearen balizko esaldiak gako-hitzen sorta mugatu batera murriztu, edo hizkuntza-ereduak sortu eskuz, hizkuntza naturalaren erabiltzailearen esaldi-sorta handia hartzen dutenak. Lehenengoak ahotsaren bidezko erabiltzaile-interfazeen erabilera zorrotz eta zurruna dakar eta erabiltzailearen onarpen eskasa izango du, ziur asko; bestalde, hizkuntza-ereduak sortu, doitu eta zaintzeak asko igo ditzake kostuak. Haatik, hizkuntza-ereduak erabiltzen dituzten eta hasieran erabiltzaile bati zer nahi duen adierazteko malgutasuna ematen dioten —adibidez, “Lagundu zaitzakat?” edo antzeko galderen bidez— ahots bidezko erabiltzaile-interfazeek automatizazio-abiadura hobea dute eta erabiltzaileen onarpen zabalagoa; horregatik, abantailak ikusten zaizkio elkarrizketa-metodo zurrunagoen aldean.

Ahots bidezko erabiltzaile-interfazeen emaitzetarako, enpresek erabiltzen dituzten esaldiak alde zuzenetik grabatutakoak izan ohi dira eta hiztun profesionalek esandakoak —ahal izanez gero, enpresakoak bertakoak—. Esaldi estatikoak diren kasuan, hitzen formulazioa erabilera-testuinguru jakin baten edo erabiltzaile horren datu pertsonalen menpekoa ez denean, erabiltzailearen esperientzia ona izango da. Aldiz, esaldi batek zenbat eta eduki dinamikoagoa hartu behar duen kontuan, are txarragoa izango da erabiltzailearen esperientzia, audio-fitxategi solteak elkarrekin lotzeagatik sortutako prosodia kaskarragoa izango baita. Gaur egungo TTS sistemak, ostera, hobeak dira esaldi dinamikoaren naturaltasun prosodikoiari dagokionez, nahiz eta oraindik ere hobetu daitezkeen.

Ahozko interakzioaren teknologiaren merkatuan, azken hamarkadan, teknologia-osagaien arteko interfazeak asko estandarizatu ziren, eta aplikazio jakin baterako software-tresna jakin batzuk sortzeko irizpideak ere agertu ziren. Era berean, merkatua asko indartu da azken hamar urteotan, batez ere hizketaren ezagutza automatikoaren eta TTSen esparruetan. Esparru horietan, G20 herrialdeetako —populazio dezentea eta

indar ekonomiko handia duten herrialdeak— merkatu nazionalak mundu osoko bost enpresa baino gutxiagoren eskuetan daude; Europan, Nuance eta Loquendo dira garrantzitsuenak. 2007. urtetik, Eusko Jaurlaritzak emandako babesari esker, Nuanceren produktu-katalogoan sartuta dago euskara. Alabaina, hizketaren ezagutza automatikoaren eskaintza tamaina txiki eta ertaineko hiztegi-aplikazioetara mugatzen da, eta ez da eskaintzen diktaketa-produkturik. TTSrako, emakumezko ahots bakarra eskaintzen da. Espainiako merkatuan, Verbio Speech Technologies^{xxxiii} enpresa kataluniarrak bietarako eskaintzen du euskara, hizketaren ezagutza automatikorako nahiz TTSrako. Euskararako diktaketa-sistema komertzialik ez dago, ordea.

Elkarrizketa-kudeaketako teknologia eta ezagutzei dagokienez, enpresa nazionalak dira nagusi merkatuetan, ETEak normalean. TTSen Espainiako merkatuan, enpresa gehienak aplikazio-sortzaileak dira. Espainiako merkatuko enpresa nagusiak hauek dira: Indsys^{xxxix} (Intelligent Dialogue Systems), Fonetic^{xxx}, Ydilo^{xxxii} eta NaturalVox^{xxxii}. Horietako zenbaitek badute eskaintza mugatu bat euskararentzat. Euskararako doaneko TTS softwarea ere badago, Euskal Herriko Unibertsitateko (EHU) Aholab^{xxxiii} ikerkuntza-taldeak eskainia.

Gaur egungo teknologiatik harago begiratuta, aldaketa esanguratsuak egongo dira, Smartphone gailuak hedatu egingo baitira bezeroekiko harremanak kudeatzeko plataforma berri moduan —telefono, Internet eta posta elektronikorekin batera—. Joera horrek eragina izango du ahozko interakziorako teknologiaren erabileran ere. Alde batetik, epe luzera, behera egingo du telefonian oinarritutako ahotsaren bidezko erabiltzaile-interfazeen eskariak. Bestetik, gero eta gehiago erabiliko da ahozko hizkera Smartphonetarako sarrera-modalitate erabilerraz moduan. Joera hori erakusten dute hiztuna kontuan hartu gabeko hizketaren ezagutzaren zehaztasunean egin diren hobekuntza nabariak —Smartphone-erabiltzaileei zerbitzu zentralizatu moduan jada eskaintzen ari zaizkien ahozko diktaketa-zerbitzuetarako egindakoak—. Ezagutzaren eginkizuna aplikazioen azpiegiturara bideratzeko joera hori ikusita, uste da hizkuntza-teknologia komunaren aplikazio espezifikoko erabilerak garrantzia hartuko duela.

Itzulpen automatikoa

Hizkuntza naturala itzultzeko ordenagailu digitalak erabiltzearen ideia A. D. Booth-ek izan zuen 1946an, eta esparru hori ikertzeko finantzaketa handia egin zen 1950eko hamarkadan, eta, berriro, 1980eko hamarkadan. Hala ere, itzulpen automatikoak (IA) ez dio oraindik behar bezala erantzun hasierako urteetan sortu zuen itxaropenari.

IAk, bere oinarritzko mailan, hizkuntza natural bateko hitzak kendu eta beste batekoak jarri besterik ez du egiten. Hori baliagarria izan daiteke esapide gutxiko hizkera oso mugatua darabilten esparruetan, hala nola eguraldi-iragarpenetan. Estandarizazio gutxiagoko testuak ondo itzultzeko, ordea, testu-unitate handiagoak (esapideak, esaldiak eta pasarte osoak ere bai) xede-hizkuntzako baliokide aproposenekin lotu behar dira. Horko zailtasunik handiena giza hizkuntzaren anbiguotasuna da, erronkak ezartzen baititu hainbat mailatan; esaterako, adiera-desanbiguazioa lexiko-mailan (“Jaguar” hitzak autoari edo animalari egin diezaioke erreferentzia) edo beste maila batzuetan, adibidez:

Egon garenetan ez dugu topatu

[Egon garen aldietan ez dugu topatu] edo

[Egon garen tokietan ez dugu topatu]

Aitak semeari bere bizikleta eman dio

[Aitarena ala semearena?]

Halakoak konpontzeko modu bat hizkuntza-arauetan oinarritzen da. Familia bereko hizkuntzekin ari bagara lanean, beharbada zuzeneko itzulpena egin daiteke bigarren adibidearen gisako perpausetan. Baina, sarritan, arauan oinarritutako sistemek (edo ezagutzak gidaturikoek) sarrera-testua aztertu eta tarteko adierazpide sinboliko bat sortzen dute, eta hortik sortzen da xede-hizkuntzako testua. Metodo horiek arrakasta izan dezaten, hiztegi handiak izan behar dituzte, informazio morfologiko, sintaktiko eta semantikoa biltzen dutenak, eta gramatika-arauen bilduma handia, hizkuntzalari aditu batek tentuz diseinaturikoa.

1980ko hamarkadaren amaieratik hasita, ordenagailua garrantzia hartuz eta merkatuz joan zen heinean, gero eta jakin-min handiagoa pizten zuten IArako eredu estatistikoek. Eredu estatistiko horien parametroak testu elebidunen corpusen analisitik atera dira; hor dugu, esaterako, Europarl corpus paraleloa, Europako Legebiltzarraren aktak 11 hizkuntza europarretan ematen dituena. Datu nahikoa izanez gero, IA estatistikoa nahiko baliagarria da atzerriko hizkuntzan idatzitako testu baten gutxi gorabeherako esanahia ateratzeko. Alabaina, ezagutzak gidaturiko sistemekin alderatuta IA estatistikoak (edo datuek gidaturikoak) duen desabantaila da emaitza agramatikalak sortzen dituela maiz. Bestalde, datuek gidaturiko IAK, gramatika idazteko giza ahalegin txikiagoa behar izateaz gainera, badu beste abantaila bat: ezagutzak gidaturiko sistemei ihes egiten dizkieten berezitasunak ondo trata ditzake, esapideak kasurako.

Ezagutzak gidaturiko IAren eta datuek gidaturiko IAren indarguneak eta ahulguneak elkarrekiko osagarriak direnez, bietako metodologiak nahasten dituzten metodo hibridoetara jotzen dute ikerlari guztiek gaur egun. Hori egiteko era bat baino gehiago daude. Batean, bi sistemak erabiltzen dira —ezagutzak gidaturikoa eta datuek gidaturikoa—, eta hautapen-modulu batek erabakitzen du zein den esaldi bakoitzaren irteerako esaldi onena. Esaldi luzeetarako, ordea, ez du topatzen emaitza egokirik. Konponbide hobea da irteera askotako esaldien zatirik onenak elkartzea; hori nahiko zaila izan daiteke, askotariko aukerei dagozkien atalak ez baitira beti agerikoak eta lerrokatu egin behar izaten baitira.

Euskararentzat, IA bereziki zaila da. Euskara hizkuntza eranskaria, morfologia aberatsekoa eta flexio-maila handikoa izanda, hiztegiaren analisia eta hiztegi-estaldura zailagoa da. Gainera, esaldiko osagaien hurrenkera dela bide, lan nekeza da corpus paraleloak kudeatzea.

Matxin transferentzian oinarritutako IA sistema bat da, gaztelaniatik euskararako, IXA taldeak Euskal Herriko Unibertsitatean garatutakoa. Irekia da, berrerabilgarria, eta euskarri elkarreragingarria eskaintzen du beste hizkuntza-bikoteentzat ere (matxin.sourceforge.org). Kode irekiko

beste programa batzuk erabiltzen ditu, hala nola Freeling, eta euskararen morfologia berrerabiltzen du morfologia-sorkuntzarako. IXA taldeak Itzulpen Automatiko Estatistikoko sistema bat ere sortu du euskara eta gaztelaniarako, hitzen segmentazioa eta berrordenaketa egiten duena (EUSMT, <http://ixa2.si.ehu.es/openmt-demo/>). IA sistema horiek garatzeko, lankidetzan estua dago unibertsitateko ikerkuntza-taldearen, Eleka Ingeniaritza Linguistikoa enpresa txikiaren eta Elhuyar Fundazioaren artean —azken horrek hizkuntza-baliabide asko jartzen ditu—. Eleka enpresak Batua-Bizkaiera bihurtzailea ere ateratu du. Alacanteko Unibertsitateko Transducens taldeak ere garatu du euskaratik gaztelaniara itzultzeko hasierako sistema bat, Apertium plataforma erabiltzen duena. Google Itzultzaileak alfa bertsio bat eskaintzen du euskararentzat.

Lucy Software enpresak —IAko aplikazioen sortzaile garrantzitsuenetakoa da nazioartean— filial bat dauka Espainian, Lucy Iberica^{xxxiv}, lehen Translendum zena. Eusko Jaurlaritzak enpresa hori hautatu zuen 2008an, gaztelaniatik euskararako itzulpen-sistema bat sortzeko, eta 2011an berriro hautatu zuen lan horrekin jarrai zezan.

Erabiltzailearekiko espezifikoa den terminologia eta lan-prozesuen integrazioa behar bezala egokituz gero, oro har, uste da Iaren erabilerak produktibitatea asko hobetu dezakeela. Halaber, IA sistemen kalitatea oraindik asko hobetu daitekeela uste da. Hainbat erronka daude oraindik; besteak beste, hizkuntza-baliabideak esparru edo erabiltzaile-talde jakin batera egokitzea eta lehendik dauden prozesuetan integratzea, terminoen datu-baseekin eta itzulpen-memoriekin batera. Gainera, hizkuntza-bikote asko falta dira oraindik.

Hizkuntza-teknologiaren atzeko lana

Hizkuntza-teknologiako aplikazioak sortzeko, era askotako lanak egin behar dira, eta, batzuetan, lan horiek ez dira agertzen erabiltzailearekiko elkarreaginean, baina zerbitzu-funtzio garrantzitsuak betetzen dituzte sistemaren “erraietan”. Horregatik, ikergai garrantzitsuak dira, eta esparru akademikoko hizkuntzalaritza konputazionalaren barruko banakako diziplinak dira orain.

Galderei erantzutea ikerketa-arlo bizia da orain; corpus glosatuak egin dira horretarako eta lehiaketa zientifikoak antolatu dira. Ideia da gako-hitzetan oinarritutako bilaketa atzean uztea (bilatzaileak garrantzitsuak izan daitezkeen dokumentuen sorta handi bat ematen du), eta bilaketa zehatza egin ahal izatea (erabiltzaileak galdera zehatz bat egiten du, eta sistemak erantzun bakar bat ematen dio): “Zer adin zuen Neil Amstrongek ilargira iritsi zenean?” - “38”. Argi dago hori lehen aipatu dugun web-bilaketaren esparru komunarekin dagoela lotuta, baina, galderei erantzutea, gaur egun termino zabala da eta hainbat zalantza sortzen ditu, hala nola “Zer galdera-*mota* ezberdindu behar lirake eta nola tratatu behar lirake”, “Nola azter eta erka daitezke erantzuna izan dezaketen dokumentuak” (bat ez datozen erantzunak eman ditzakete?) eta “Nola erauz daitezke informazio espezifikoa (erantzuna) testu batetik testuingurua gehiegi baztertu gabe?”.

Horrek lotura du informazio-erauzketa (IE) delako lanarekin ere; garrantzi eta eragin handiko alorra izan zen hori 1990eko hamarkadaren hasieran, hizkuntzalaritza konputazionalaren barruan “aldaketa estatistikoa” gertatu zen garaian. IEren xedea da

dokumentu-mota espezifikoetan informazio espezifikoa aurkitzea; adibidez, egunkarietako pasarteetan kontatzen diren enpresa-xurgatzeetan partaide nagusiak zein diren aurkitzea. Landutako beste alor bat terroristen erasoei buruzko albisteena da; horietan, erasoaren egilea, helburua, ordua, tokia eta ondorioak jasotzen dituen txantilo batera pasatzea da arazoa. Esparru espezifikoko txantiloak betetzeko gaitasuna IEren bereizgarri nagusia da, eta, horrexegatik, “atzeko” teknologiaren beste adibide bat da; ondo zedarritutako ikerketa-eremuak ditu, baina, erabilgarria izan dadin, aplikazio-inguru egoki batean txertatu behar da.

Bi “mugako” eremu daude: testu-laburpenak eta testu-sorkuntza; batzuetan, aplikazio autonomoak dira, eta, beste batzuetan, azpiko aplikazio laguntzaileak. Testu-laburpenak, agerikoa den moduan, testu luze bat laburtzeko lanari egiten dio erreferentzia, eta MS Word-en barruko funtzio moduan eskaintzen da, esaterako. Batez ere estatistiketan oinarrituta egiten du lan; lehenik, testu bateko hitz “garrantzitsuak” aurkitzen ditu (adibidez, testuan oso maiz eta hizkuntzaren erabilera orokorrean askoz gutxiagotan agertzen diren hitzak), eta, gero, hitz garrantzitsu asko dituzten esaldiak detektatzen ditu. Esaldi horiek dokumentuan markatu edo handik erauzi egiten ditu, eta laburpena egiteko erabiltzen ditu. Zeregin horretan —eta horixe da bere zeregin nagusia—, testu-laburpena eta esaldi-erauzketa gauza bera dira: testua txikiagotu egiten da bere esaldien azpimultzo batera. Testu-laburpenerako tresna komertzial guztiek ideia horixe erabiltzen dute. Zertxobait ikertu den beste metodo bat laburpenean esaldi *berriak* sartzea da; hots, jatorrizko testuan forma horretan agertu beharrik ez duten esaldiez osaturiko laburpena egitea. Horretarako, testua sakonagotik ulertu behar da, eta, beraz, ez da horren metodo sendoa. Azken batean, testu-sortzaile bat normalean ez da izaten aplikazio autonomo bat, software handiagoko batean txertatutako aplikazioa baizik; adibidez, informazio klinikoaren sistemetan, pazienteen datuak bildu, gorde eta prozesatu egiten dira, eta txostena sortzea da sisteman txertatutako testu-sortzailearen funtzioetako bat.

Ikerkuntza-eremu horiek guztiak ez daude ingeleserako bezainbeste garatuta euskararako. Bada, ingelesean, hainbat eta hainbat lehiaketa ireki egin dira galderei erantzutearen, informazio-erauzketaren eta testu-laburpenen esparruetan, Estatu Batuetako DARPA eta NISTek antolatuta. Horrek aurrerapen handiak ekarri ditu esparruan, baina jomuga ingelesa izan da beti; lehiaketa batzuetan sartu izan dira hainbat hizkuntza, baina euskara ez da inoiz horietako bat izan. Hori dela eta, corpus glosatu edo baliabide gutxi daude eskuragarri zeregin horietarako. Testu-laburpenetarako sistemak, metodo estatistikoak bakarrik erabiltzen dituztenean, hizkuntzaren mendekoak dira, neurri handi batean, eta, beraz, zenbait ikerketa-prototipo daude eskuragarri. Testu-sorkuntzarako, osagai berrerabilgarriak gainazala egiteko modulueta (“sorkuntza-gramatiketara”) mugatuta egon izan dira; hor ere, eskura dauden software gehienak ingeleserako dira.

Hizkuntza-teknologia hezkuntzan

Hizkuntza-teknologia diziplinarteko alorra da, eta hainbat adituren lana hartzen du; besteak beste, hizkuntzalariak, informatikariak, matematikariak, filosofoak, psikolinguistak eta neurozientzialariak. Hortaz, gaur egun, Espainian, hizkuntzalari konputazional izateko oinarritzko prestakuntza filologia edo hizkuntzalaritzako gradu baten barruan emango da, beharbada —irakasgai komunaren artean hizkuntzalaritza konputazionala ematen bada—, edo informatika-fakultatean, bestela. Lehen aukera eskaintzen duten

unibertsitateak hauek dira: Universitat de Barcelona, Universitat Pompeu Fabra, Universitat Oberta de Catalunya eta Universidade de Vigo. Hizkuntzalaritza konputazionala irakasgaitzat ematen duten informatika-fakultate nagusiak, berriz, beste hauek dira: Universidad Politécnica de Madrid, Universidad Carlos III, Universidad Autónoma de Madrid, Universitat d'Alacant, Universidad Nacional de Educación a Distancia eta Euskal Herriko Unibertsitatea. Bi aukerak eskaintzen dituenik ere bada: Universidad Complutense.

Gratuondoko ikastaroen helburu zehatzagoa duen prestakuntza profesionala eskaintzen dute. Doktore-ikastaro batzuetan ematen dira hizkuntzaren eta hizketaren prozesamenduarekin lotutako masterrak eta irakasgaiak. Euskal Herriko Unibertsitateak hizkuntzaren prozesamenduari buruzko doktoretza-ikastaro oso bat eskaintzen du, euskara hutsean ere ikas daitekeena. Beste master edo doktoretza-ikastaro batzuetako ikasleei ere eskaintzen zaizkie hizkuntza-teknologiako moduluak, hizketaren prozesamendua batik bat (adib. EHUren Sare Mugikorretako Informazio- eta Komunikazio-teknologiak masterreko ikasleei).

Euskal Autonomia Erkidegoko hiru unibertsitateetan banatuta dauden zenbait ikerketa-talde ari dira gai hauek lantzen: hizketaren prozesamendua, hizketaren sintesia eta bihurketa, hizketaren eta hitzunen ezagutza, hizkuntzaren ezagutza, hizkuntza naturalaren prozesamendua, testutik testurako itzulpena eta hizketatik hizketarako itzulpena. Guztiak dira Hizkuntza Naturalaren Prozesamendurako Espainiar Elkartearen kide (SEPLN, Sociedad Española para el Procesamiento del Lenguaje Natural). SEPLN irabazi asmorik gabeko erakundea da, esparru akademikoko nahiz industria-arloko 300 kide baino gehiago dituena, eta 1984an sortu zen, helburu honekin: irakaskuntza, ikerkuntza eta LNPren garpenarekin lotutako jarduerak sustatu eta zabaltzea, Espainian bertan nahiz nazioartean. SEPLN elkarteak mintegiak, sinposioak eta hitzaldiak antolatzen ditu, eta Espainiako nahiz nazioarteko erakundeekiko lankidetzak sustatzen ditu.

SEPLNk urteroko hitzaldi bat antolatzen du, eta, urtetik urtera, LNP lantzen duten ikerlari gehiago erakartzen ditu, Espainiatik nahiz kanpotik datozenak. Elkarteak aldizkari bat ere kaleratzen du, eta web-zerbitzari bat du, hizkuntza naturalaren prozesamenduarekin lotutako gaiei buruzko informazioa eta kideentzako foro ireki bat eskaintzen dituena.

Espainiako Hizketa Teknologietako Sarea (RTTH)^{xxxv} foro komun bat da, eta han elkartzen diren hizketa-teknologiaren ikerlariak (250 baino gehiago dira gaur egun) zeregin osagarriak egin eta esperientziak elkarbanatzen dituzte, helburu hauekin:

- Hizketa-teknologiaren ikerkuntza sustatzea, alor horretara ikerlari gazte gehiago erakartzeko prestakuntza, ikasle-trukeak, bekak eta sariak eskainiz.
- Negozio-ikerkuntzarako inbertsioak erakartzea, beste negozio-aukera batzuk eskaintzen dituzten aplikazio berriak aurkituz.
- Aurrerapenak egitea elkarteak sortzen, eta sareko kideak integratzea, Espainiak gaztelaniaren ikerkuntzan duen nagusitasunari eusteko eta hizkuntza koofizialei ere bultzada emateko (katalana, euskara eta galegoa).

RTTH elkarteak “Hizketaren teknologiari buruzko jardunaldiak” antolatzen ditu urtero, 2000. urtetik. Ikastaro horren xedea da

elkargune bat izatea, iberiar penintsulako hizkuntzetan dauden hizketa- eta hizkuntza-teknologiaren gaineko ikerkuntzaren emaitzen berri emateko eta horiei buruzko eztabaida sustatzeko. Enpresen eta unibertsitatearen arteko elkarlana ere sustatzen du. Era askotako jarduerak antolatzen ditu: txosten teknikoaren aurkezpenak, hitzaldi magistralak, proiektu-txostenen eta laborategiko jardueren aurkezpenak, erakustaldiak eta azken doktore-tesien aurkezpenak.

Hizkuntza-teknologiako programak

Euskararako teknologia-programak Eusko Jaurlaritzak eta Espainiako Gobernuak bultzatu dituzte, batez ere.

Espainiako Hezkuntza Ministerioak eta Zientzia eta Berrikuntzako Ministerioak ikerketa-programa nazionalen bidez bultzatu dute informazio-teknologiaren alorreko ikerketa. Programa horietarako, ikerketa-proiektu asko egin behar izan dira, eta elkarlana egin nazioarteko ikerketa-zentro eta -enpresekin. Euskararen prozesamendu automatikorako aplikazio komertzialen eta aurrerapen teknologikoaren oinarria proiektu horien ondorioz sortu da, hein batean.

2000. urtetik hona, Espainiako Gobernuak, Ikerketarako eta Teknologia Berrikuntzarako Plan Nazionalaren barruan, proiektu batzuk bultzatu ditu hizketa-teknologia eleaniztunen alorrean: TEHAM, AVIVAVOZ eta BUCEADOR. Haien helburu nagusia zen hizketaren ezagutza, hizketaren itzulpena eta testutik hizketarako sintesia hobetzea Espainiako hizkuntza ofizial guztietan: euskara, galegoa, katalana eta gaztelania.

Industria Teknologia Garatzeko Zentroa (CDTI) Espainiako erakunde publiko bat da, Zientzia eta Berrikuntza Ministerioaren mendekoa, eta Espainiako enpresen maila teknologikoa areagotzen laguntzea du helburu. CDTI-ik I+G proiektuak ebaluatu eta finantzatzeko, CENIT (2010ean bukatutakoa) eta AVANZA bezalako programen bitartez.

Eusko Jaurlaritzak ikerketa eta berrikuntzaren alde egiten du “Zientzia, teknologia eta berrikuntzarako plana”ren bitartez (ZTBP). Plan horren barruan, erakunde eta ikerketa- eta berrikuntza-agentzia batzuk eratu dira azken urteotan: Zientzia, Teknologia eta Berrikuntzaren Euskal Kontseilua (ikerketa eta berrikuntza bultzatzeko eta garatzeko ekintzak egiten dituen erakunde politiko ahaltsuena), InnoBasque (berrikuntzaren euskal agentzia) eta IkerBasque (zientziaren euskal fundazioa, talentudun ikerlariak euskal zientzia- eta teknologia-sistemara erakartzeaz arduratzen dena). ZTBP planaren tresna garrantzitsuenak ikerketa- eta berrikuntza-proiektuetarako deialdiak dira: ETORTEK programa —Zientzia, Teknologia eta Berrikuntzarako Euskal Sareko erakundeei zuzendutakoa— eta ETORGAI programa —enpresa pribatuei zuzendutakoa—.

Azken ZTBP planean (2010ekoan) identifikatutako alor estrategiko bat hizkuntza-teknologiarena izan da, aurrekoetan bezalaxe. Hala, azken hamar urteotan, HIZKING21, ANHITZ eta BERBATEK^{xxxvi} proiektuak gauzatu dira ETORTEK programaren barruan. Euskararako gaur egun dauden baliabide eta tresna gehienak proiektu horien bidez eskuratutakoak dira.

Euskararako tresnen eta baliabideen eskuragarritasuna

Taula honetan, euskararako hizkuntza-teknologiako zerbitzuen gaur egungo egoeraren berri ematen da. Punta-puntako aditu batzuek gaur egun dauden tresnak eta baliabideak puntuatu dituzte, irizpide hauen arabera kalkulu informatuetan oinarrituta:

- 1 **Kantitatea:** Hizkuntza honetarako tresnarik/baliabiderik badago? Zenbat eta tresna/baliabide gehiago egon, orduan eta puntuazioa altuagoa.
 - 0: inolako tresnarik/ baliabiderik ez
 - 6: tresna/baliabide asko, hainbat motatakoak
- 2 **Eskuragarritasuna:** Tresnak/baliabideak eskuragarri daude? Hau da, kode irekikoak dira, edozein plataformatan erabiltzeko modukoak, edo prezio altua eta oso kondizio mugatuak dituzte?
 - 0: tresna/baliabide guztiak edo ia guztiak asko ordaindu behar dira
 - 6: tresna/baliabide gehienak doakoak eta irekiak dira — kode irekikoak edo Creative Commons en zentzuzko lizentzia batzuekin erabiltzekoak—, eta berrerabiltzeko eta beste xede batzuk ezartzeko aukera ematen dutenak
- 3 **Kalitatea:** Eskura dauden tresna, aplikazio eta baliabideek zenbateraino betetzen dituzte dagozkien tresnen funtzionamendu-irizpideak eta baliabideen kalitate-adierazleak? Tresna/baliabide horiek funtzionamenduan daude eta ondo mantenduta?
 - 0: baliabide/tresna mugatua
 - 6: kalitatezko tresna, giza kalitateko oharrak daude baliabide batean
- 4 **Estaldura:** Tresna onenek zenbateraino betetzen dituzte dagozkien estaldura-irizpideak (estiloak, generoak, testu-motak, hizkuntza-fenomenoak, sarrera/irteera motak, IA sistema duten hizkuntzen kopurua)? Baliabideek zein neurritan ordezkatzan dute helburuko hizkuntza edo mendeko hizkuntzak?
 - 0: xede bereziko baliabidea edo tresna, kasu espezifikotarako, estaldura oso eskasa, erabilera oso espezifikotan bakarrik erabiltzekoa.
 - 6: oso estaldura zabaleko baliabidea, oso tresna gogorra, erabilera askokoa, hizkuntza asko har ditzakeena.
- 5 **Heldutasuna:** Esan daiteke tresna/baliabidea heldua dela, egonkorra, eta merkaturako prest dagoela? Eskura dauden tresna/baliabide onenak bere horretan erabil daitezke edo egokitu egin behar dira? Teknologia horren funtzionamendua egokia da eta produzitzeko prest dago, edo produkzio-sistematan erabil ezin daitekeen prototipo bat besterik ez da? Jendeak baliabide/tresna horiek onartu dituen eta HT sistematan arazorik gabe erabiltzen diren begiratu daiteke hori jakiteko.
 - 0: hasierako prototipoa, sistema mugatua, kontzeptu-proba, adibiderako baliabide-proba
 - 6: jada integratu edo aplikatzeko prest dagoen osagaia

6 Iraunkortasuna: Tresna/baliabidea ondo manten/integra daiteke gaur egungo IT sistemetan? Tresnak/baliabideak har ditzake, neurriren batean, dokumentazioa/eskuliburuak, erabilera-kasuen azalpenak, aurrealdeak, GUIak eta abar? Programazio-plataforma estandar/egokia erabiltzen du (adib. Java EE)? Badago industriako/ikerketako parametrorik edo parametro gisakorik, eta, halakorik balego, tresnak/baliabideak betetzen ditu haiek (datu-formatuak eta abar)?

- 0: erabat pribatua, datu-formatu bereziak eta APIak
- 6: parametro guztiak betetzen ditu, erabat dokumentatuta

7 Moldagarritasuna: Eskura dauden tresna edo baliabide onenak zenbateraino molda/heda daitezke beste zeregin/esparru/genero/testu-mota/erabilera-kasu batzuetara?

- 0: ia ezinezkoa da tresna/baliabide bat beste zeregin baterako moldatzea, baita eskura baliabide edo langile mordoa izanda ere.
- 6: oso moldagarritasun-maila altua; oso modu errazean eta eraginkorrean egin daiteke moldaketa

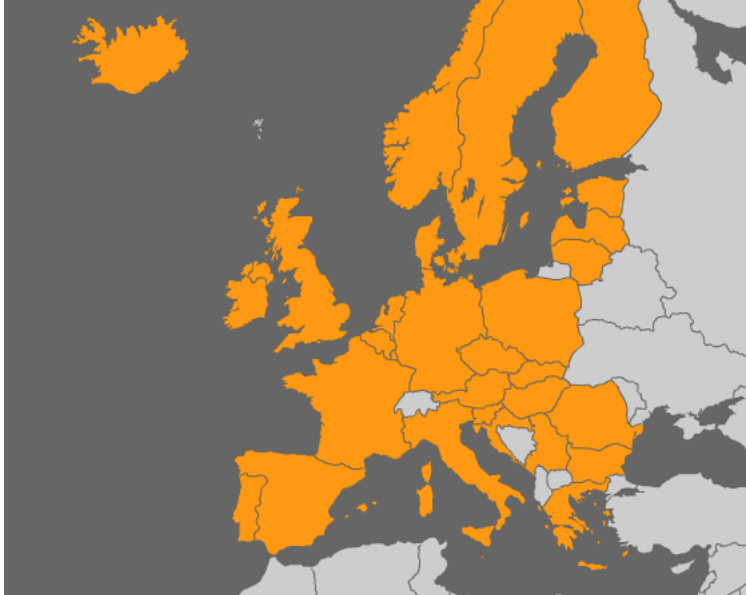
Tresna eta baliabideen taula

	Kantitatea	Eskuragarritasuna	Kalitatea	Estaldura	Heldutasuna	Iraunkortasuna	Moldagarritasuna
Hizkuntza-teknologia (tresnak, teknologiak, aplikazioak)							
Tokenizazioa, morfologia (tokenizazioa, kategoria-ezartzea, analisi/sorkuntza morfologikoa)	4	3	5	5	5	4	4
Analisi sintaktikoa (azalekoa edo sakonekoa)	3	2	4	4	4	4	3
Perpauen semantika (adiera-desanbiguazioa, argumentu-egitura, rol semantikoak)	3	3	4	3	3	3	4
Testu-semantika (korreferentziaren konponbidea, testuingurua, pragmatika, inferentzia)	1	1	1	1	1	1	1
Diskurtsoaren prozesamendu aurreratua (testu-egitura, koherentzia, kohesioa, egitura erretorikoaren teoria, argumentuaren araberako zonifikazioa, argumentazioa, testu-ereduak, testu-motak eta abar)	1	1	1	1	1	1	1
Informazio-berreskurapena (testu-indexatzea, multimediatiko IB, hizkuntzen arteko IB)	3	2	4	4	4	3	4
Informazio-erazketa (entitateen izenen ezagutza, gertaeren/erlazioen erazketa,	3	3	4	4	3	3	3

	Kantitatea	Eskuragarritasuna	Kalitatea	Estaldura	Heldutasuna	Iraunkortasuna	Moldagarritasuna
iritzien/sentimenduen identifikazioa, testu-meatzaritza/analisiak)							
Hizkuntza-sorkuntza (perpau-sorkuntza, txosten-sorkuntza, testu-sorkuntza)	1	0	0	0	0	0	0
Testu-laburpenak, galderak erantzutea, informazioa eskuratzeko teknologia aurreratuak	1	1	2	2	1	1	1
Itzulpen automatikoa	3	4	2	3	3	3	3
Hizketaren ezagutza	2	1	1	1	4	4	2
Hizketaren sintesia	2	3	5	2	5	4	3
Elkarrizketa-kudeaketa (elkarrizketa-ahalmenak eta erabiltzaile-modelizazioa)	0	0	0	0	0	0	0
Hizkuntza-baliabideak (baliabide-, datu- eta ezagutza-baseak)							
Erreferentziako corpusak	3	4	5	2	5	4	4
Corpus sintaktikoak (zuhaitz-bankuak, mendekotasun-bankuak)	1	4	4	2	3	4	4
Corpus semantikoak	1	1	1	1	1	1	1
Diskurtso-corpusak	1	0	0	0	0	0	0
Corpus paraleloak, itzulpen-memoriak	2	4	4	5	4	4	5
Hizketa-corpusak (hizketa-datu gordinak, etiketadun/ohardun hizketa-datuak, ahozko elkarrizketako datuak)	3	2	4	2	3	4	4
Askotariko datuak eta multimedia-datuak (testu-datuak, audio/bideoekin konbinatuta)	2	3	5	1	2	2	2
Hizkuntza-ereduak	1	3	2	2	2	3	4
Lexikoiak, termino-bankuak	5	4	5	6	6	6	6
Gramatikak	3	4	3	4	4	4	4
Thesaurusak, sare semantikoak	3	5	4	4	5	5	5
Ingurunearen ezaguerarako baliabide ontologikoak (adib. goiko ereduak, lotutako datuak)	2	2	2	2	2	2	2

META-NET sarea

Europako Batzordeak sortutako bikaintasunezko sarea da META-NET. Sareak Europako 31 herrialdetako 47 kide ditu, gaur egun. META-NETek Europa Eleaniztunaren Teknologia Aliantza (META) babesten du, hizkuntza-teknologiako aditu eta erakundeen talde europar gero eta handiagoa.



1. irudia: META-NETen ordezkariak duten herrialdeak

META-NET beste ekimen batzuekin elkarlanean ere aritzen da; horietako bat da CLARIN (Common Language Resources and Technology Infrastructure), Europan giza zientzien ikerkuntza digitala ezartzeko proiektua. META-NETek oinarri teknologikoak eman nahi ditu informazio-gizarte zinez eleaniztuna sortzeko Europan, eta hari eusteko. Horrelako gizarte bat lortu nahi da:

- ❑ Hizkuntzen arteko komunikazio eta lankidetzarako aukera ematen duena.
- ❑ Hizkuntza guztietan aukera berdinak ematen dituen informazioa eta ezagutzak eskuratzeko.
- ❑ Europarrei informazio-teknologia aurreratua eskaintzen diena sarean eta modu onean.

META-NETek Europako hizkuntza guztietarako bultzatzen eta sustatzen ditu teknologia eleaniztunak. Teknologiai esker, era askotako aplikazio eta esparruetan erabil daitezke itzulpen automatikoa, eduki-sorkuntza, informazio-prozesamendua eta ezagutza-kudeaketa. Sareak hobetu egin nahi ditu egungo metodoak, eta, beraz, baliteke hizkuntzen arteko komunikazio eta lankidetzak ere hobetzea. Europar guztiek informazioa eta ezagutzak lortzeko eskubide bera dute, edozein hizkuntza izanda ere.

Lan-ildoak

META-NET 2010eko otsailaren 1ean jarri zen abian, hizkuntza-teknologiaren (HT) ikerkuntzan aurrera egiteko asmoarekin. Sareak merkatu digital eta informaziogune bakar batean elkartutako Europa baten alde egiten du. META-NETek bere helburuetatik harago doazen jarduerak ere egin ditu. META-



*The Multilingual Europe
Technology Alliance (META)*

VISION, META-SHARE eta META-RESEARCH sarearen hiru lan-ildoak dira.



2. irudia: META-NETen hiru lan-ildoak

META-VISIONen helburua da akziodunen komunitate bizia eta eragin handikoa sortzea ikuspegi partekatu baten eta ikerketa estrategiko komun baten inguruan. Proiektu horren lan-ildo nagusia da Europar HTren komunitate koherente eta kohesiboa eratzea, akziodunen talde zatitu eta anitzetako ordezkariak elkartuz. META-NETen lehen urtean, jendearenganako irisgarritasuna izan zen gai nagusia jardunaldi hauetan eginiko aurkezpenetan: FLReNet Forum (Espainian), Language Technology Days (Luxenburgon), JIAMCATT 2010 (Luxenburgon), LREC 2010 (Maltan), EAMT 2010 (Frantzia) eta ICT 2010 (Belgikan). Hasierako kalkuluen arabera, META-NET dagoeneko hizkuntza-teknologiako 2.500 aditu baino gehiagorekin jarri da harremanetan, haiekin batera garatzeko bere helburuak eta ikuspegiak. Bruselako META-FORUM 2010 bileran, META-NETek 250 parte-hartzaile baino gehiagori eman zion bere ikuspegia eraikitzeko prozesuaren berri. Hainbat saio interaktibotan, sarearen ikuspegiaren gaineko feedbacka eman zuten parte-hartzaileek.

META-SHARE proiektuak baliabideak trukatzeko eta partekatzeko bitarteko ireki eta partekatua eskaintzen du. Biltegiekin osatutako parekoen sareak hizkuntza-datuak, tresnak eta web-zerbitzuak izango ditu, kalitatezko metadatuekin dokumentatuak eta kategoria estandarizatuetan antolatuak. Baliabideak erraz eskura daitezke, eta bilaketa uniforme da. Baliabideok kode irekikoak izan daitezke —doakoak, beraz— edo ordainduta eskuratu beharreko salgai mugatuak. META-SHAREren jomuga lehendik dauden hizkuntza-datuak, tresnak eta sistemak dira, bai eta teknologia, produktu eta zerbitzu berriak sortu eta ebaluatzeko beharrezkoak diren produktu berriak ere. Hizkuntza-datuak eta tresnak berrerabili, konbinatu, beste xede batzuetarako prestatu eta berriz diseinatzeko lanek berebiziko garrantzia dute. Lehenago edo geroago, META-SHARE hizkuntza-teknologiaren merkatuko baliabide garrantzitsu bat izango da enpresa txiki, ertain eta handietako garapen-adituentzat, lokalizazio-adituentzat, ikerlariarentzat, itzultzaileentzat eta hizkuntza-adituentzat. META-SHAREren asmoa da hizkuntza-teknologiaren ziklo osoa garatzea: ikerketatik hasi, eta produktu eta zerbitzu berritzaileetaraino. Lan horren giltzarria da META-SHARE Europako eta munduko azpiegituraren zati garrantzitsu eta baliotsua izatea hizkuntza-teknologiaren komunitatearentzat.

META-RESEARCH proiektuak zubiak eraikitzen ditu hurbileko teknologia-esparruetara iristeko. Xedea beste esparruetan aurrerapenak eragitea da, eta hizkuntza-teknologiaren onerako izan daitekeen ikerketa berritzailea aprobeztatzea. Proiektuaren

asmoa da itzulpen automatikoari semantika gehitzea, itzulpen automatiko hibridoko lan-banaketa optimizatzea, itzulpen automatikoetan testuinguruaz baliatzea eta IArako oinarri enpirikoa prestatzea. META-RESEARCH beste alor eta diziplina batzuetan ere ari da lanean, hala nola ikaskuntza automatikoan eta web semantikoaren komunitatean. META-RESEARCHek zeregin nagusi hauek ditu: datuak bildu, datu-sortak prestatu eta hizkuntza-baliabideak ebaluaziorako prestatzea; tresnen eta metodoen inbentarioak biltzea; eta ikastaroak eta prestakuntza-saioak antolatzea komunitateko kideentzat. Dagoeneko agerian utzi du semantikak eragin handia izan dezakeela itzulpen automatikoaren alderdi batzuetako jardunbide egokietan. Gainera, gomendio batzuk ematen ditu informazio semantikoa itzulpen automatikoan integratzeko zailtasunak ekiditeko. META-RESEARCH itzulpen automatikorako hizkuntza-baliabide berri bat ere ari da burutzen, *Annotated Hybrid Sample MT Corpus* izenekoa, ingelesa-alemana, ingelesa-gaztelania eta ingelesa-txekiera hizkuntza-bikoteentzako balio duena. META-RESEARCHek sarean ezkutuan dauden corpus eleaniztunak biltzen dituen software bat ere sortu du.

Erakunde kideak

Taula honetan, META-NETen parte diren erakundeak eta haien ordezkariak ageri dira.

Herraldea	Erakundea	Ordezkaria(k)
Austria	Vienako Unibertsitatea	Gerhard Budin
Belgika	Anberesgo Unibertsitatea	Walter Daelemans
	Lovainako Unibertsitatea	Dirk van Compernelle
Bulgaria	Bulgariako Zientzia Akademia	Svetla Koeva
Kroazia	Zagrebeko Unibertsitatea	Marko Tadić
Zipre	Zipreko Unibertsitatea	Jack Burston
Txekiar Errepublika	Pragako Karlova Unibertsitatea	Jan Hajic
Danimarka	Kopenhageko Unibertsitatea	Bolette Sandford Pedersen eta Bente Maegaard
Estonia	Tartuko Unibertsitatea	Tiit Roosmaa
Finlandia	Aalto Unibertsitatea	Timo Honkela
	Helsinki Unibertsitatea	Kimmo Koskenniemi eta Krister Linden
Frantzia	CNRS/LIMSI	Joseph Mariani
	Evaluations and Language Resources Distribution Agency	Khalid Choukri
Alemania	DFKI	Hans Uszkoreit eta Georg Rehm
	RWTH Aachen University	Hermann Ney

Herrialdea	Erakundea	Ordezkarria(k)
	Sarreko Unibertsitatea	Manfred Pinkal
Grezia	Institute for Language and Speech Processing, "Athena" R.C.	Stelios Piperidis
Hungaria	Hungariako Zientzia Akademia	Tamás Váradi
	Budapesteko Teknologia eta Ekonomia Unibertsitatea	Géza Németh eta Gábor Olaszy
Islandia	Islandiako Unibertsitatea	Eiríkur Rögnvaldsson
Irlanda	Dublin City University	Josef van Genabith
Italia	Consiglio Nazionale Ricerche, Istituto di Linguistica Computazionale "Antonio Zampolli"	Nicoletta Calzolari
	Fondazione Bruno Kessler	Bernardo Magnini
Letonia	Tilde	Andrejs Vasiljevs
	Matematika eta Informatika Institutua, Letoniako Unibertsitatea	Inguna Skadina
Lituania	Lituanieraren Institutua	Jolanta Zabarskaitė
Luxenburgo	Arax Ltd.	Vartkes Goetcherian
Malta	Maltako Unibertsitatea	Mike Rosner
Herbehereak	Utrechteko Unibertsitatea	Jan Odijk
	Groningengo Unibertsitatea	Gertjan van Noord
Norvegia	Bergengo Unibertsitatea	Koenraad De Smedt
Polonia	Poloniako Zientzia Akademia	Adam Przepiórkowski eta Maciej Ogrodniczuk
	Łódzko Unibertsitatea	Barbara Lewandowska-Tomaszczyk eta Piotr Pęzik
Portugal	Lisboako Unibertsitatea	Antonio Branco
	Sistema Ingeniaritzaren eta Ordenagailuen Institutua	Isabel Trancoso
Errumania	Errumaniako Zientzia Akademia	Dan Tufis
	Alexandru Ioan Cuza unibertsitatea	Dan Cristea
Serbia	Belgradeko Unibertsitatea	Dusko Vitas, Cvetana Krstev eta Ivan Obradovic
	Mihailo Pupin institutua	Sanja Vranes
Eslovakia	Eslovakiako Zientzia Akademia	Radovan Garabik
Eslovenia	Jozef Stefan institutua	Marko Grobelnik

Herraldea	Erakundea	Ordezkaría(k)
Espainia	Barcelona Media	Toni Badia
	Kataluniako Unibertsitate Teknikoa	Asunción Moreno
	Pompeu Fabra unibertsitatea	Núria Bel
Suedia	Göteborgeko Unibertsitatea	Lars Borin
Erresuma Batua	Manchesterreko Unibertsitatea	Sophia Ananiadou
	Edinburgoko Unibertsitatea	Steve Renals

Erreferentziak

- ⁱ European Commission Directorate-General Information Society and Media, *User language preferences online*, Flash Eurobarometer #313, 2011 (http://ec.europa.eu/public_opinion/flash/fl_313_en.pdf).
- ⁱⁱ European Commission, *Multilingualism: an asset for Europe and a shared commitment*, Brussels, 2008 (http://ec.europa.eu/education/languages/pdf/com/2008_0566_en.pdf).
- ⁱⁱⁱ UNESCO Director-General, *Intersectoral mid-term strategy on languages and multilingualism*, Paris, 2007 (<http://unesdoc.unesco.org/images/0015/001503/150335e.pdf>).
- ^{iv} European Commission Directorate-General for Translation, *Size of the language industry in the EU*, Kingston Upon Thames, 2009 (<http://ec.europa.eu/dgs/translation/publications/studies>).
- ^v <http://www.unesco.org/en/languages-and-multilingualism/>
- ^{vi} <http://en.eustat.es>
- ^{vii} http://www.euskaltzaindia.net/index.php?option=com_content&Itemid=1&id=18&lang=en&layout=blog&view=section
- ^{viii} IXA group: *Automatic morphological analysis of Basque*, Literary & Linguistic Computing, vol. 11, No. 4, pp. 193-203, 1996.
- ^{ix} Koldo Zuazo: *Euskararen sendabelarrak*, Alberdania, 2000.
- ^x http://www.langune.com/home?set_language=en
- ^{xi} www.mintzaira.fr
- ^{xii} <http://www.isei-ivei.net/cast/pub/pisa2009/PISA2009-EUSKADI-1INFORME.pdf>
- ^{xiii} <https://ixa.si.ehu.es/master/en>
- ^{xiv} <http://www.euskaletxeak.net/i>
- ^{xv} http://ec.europa.eu/education/languages/languages-of-europe/doc139_en.htm
- ^{xvi} http://en.eustat.es/estadisticas/opt_o/id_118/ti_Information_Society/subarbol.html#axzz1LTNljBpS
- ^{xvii} http://www.euskara.euskadi.net/r59-20660/eu/contenidos/informacion/euskarazko_softwarea/eu_9567/aurkib.html
- ^{xviii} <http://softkat.ueu.org/>
- ^{xix} http://aclweb.org/aclwiki/index.php?title=Resources_for_Basque
- ^{xx} <http://www.hiztegia.net/>
- ^{xxi} <http://www.nolaerran.org>
- ^{xxii} <http://www.puntueus.org/en/>
- ^{xxiii} <http://www.xuxen.com>
- ^{xxiv} <http://hizkia.pagesperso-orange.fr>
- ^{xxv} <http://www.uzei.com>
- ^{xxvi} <http://www.spiegel.de/netzwelt/web/0,1518,619398,00.html>

xxvii

http://www.pcworld.com/businesscenter/article/161869/google_rolls_out_semantic_search_capabilities.html

xxviii <http://www.verbio.com>

xxix <http://www.indisys.es/default.aspx>

xxx <http://www.fonetic.es/>

xxxi <http://www.ydilo.com/esp/index.php>

xxxii <http://www.naturalvox.com/>

xxxiii <http://aholab.ehu.es/tts>

xxxiv <http://www.lucysoftware.com/>

xxxv <http://www.rthabla.es>