

Edizioa

Euskarak Mundu Digitalean Bizirauteko Kita — DLDPren gomendioak, euskararen bizitasun digitala hobetu dadin

Egileak:

Klara Ceberio Berger, Antton Gurrutxaga Hernaiz, Paola Baroni, Davyth Hicks, Eleonore Kruse, Valeria Quochi, Irene Russo, Tuomo Salonen, Anneli Sarhimaa, Claudia Soria

Lan hau Hizkuntza Dibertsitate Digitalaren Proiektuaren barruan egin da (www.dldp.eu/eu), Europar Batasunak finantzatua Erasmus+ Programaren barruan (2015-1-IT02-KA204- 015090 finantzazio-akordioa).

© 2018

Lan hau Nazioarteko Creative Commons Aintzatespen 4.0 lizentziapean dago.

Azalaren diseinua: Eleonore Kruse

Lege-oharra

Argitalpen honek egileen ikuspegia azaltzen du eta Erasmus+ Agentzia eta Batzorde Nazionala ez dira eduki horien erabileraren erantzule egiten.



www.dldp.eu



www.facebook.com/digitallanguagediversity



dldp@dldp.eu



www.twitter.com/dldproject

Gomendioen ikuspegi azkarra

Gaitasun digitala	
Adierazlea	Gomendioa
<u>Hizkuntza-baliabideen eskuragarritasuna</u>	<u>Erdi- eta goi-mailako hizkuntza-baliabideak garatzea</u>
	<u>Hiztegigintza: dibertsitatea, tamaina, espezializazioa eta dibulgazioa</u>
	<u>Corpusaren tamaina eta dibertsitatea handitzea</u>
	<u>Publikoki eskuragarri dauden datuak biltzea sare sozialetatik</u>
	<u>Corpusak aztertzeke tresnak erabiltzea, eta zure hiztegia elikatzea erabiltzen den hizkuntzari buruzko datuekin</u>
	<u>Hizketaren sintesi eta ezagutzarako lehenengo urratsak</u>

Presentzia eta erabilera digitala	
Adierazlea	Gomendioa
<u>Komunikazio elektronikorako erabilera</u>	<u>Aztertzea zenbat balio duen eskualde-hizkuntza edo hizkuntza gutxitua erabiltzeak pertsonen arteko komunikaziorako</u>
<u>Interneteko hedabideen eskuragarritasuna</u>	<u>Eduki-kantitatea handitzea eta Interneteko hedabide-motak dibertsifikatzea</u>
	<u>Testu-motako edukiaren kantitatea handitzea (webguneak, blogak, foroak)</u>
	<u>Dokumentuen eta erregistroen fitxategi bat (sarean oinarritua) sortzea edo elikatzea</u>
	<u>Saioak online ematea, software libreko tresnak erabiliz</u>
	<u>Istorio digitalak zeure hizkuntzan erregistratzea</u>
	<u>Azpitulazio-ekimenak sustatzea</u>
<u>Wikipedia</u>	<u>Igo ezazu zure Wikipedia maila bat gorago</u>
	<u>Wikipediaren tamaina eta kalitatea hobetzeko ekimenak</u>

Errendimendu	
Adierazlea	Gomendioa
<u>Interneteko zerbitzuen eskuragarritasuna</u>	<u>Interneteko zerbitzuak zure hizkuntzan erabiltzeko aukerak zabaltzea</u>
	<u>Zure eskualde-hizkuntza edo hizkuntza gutxitua erabiltzen dutenen komunitatetik informazioa eta esperientziak biltzea, zehazteko zer zerbitzu diren garrantzitsuenak eta erabilienak</u>
	<u>Aztertzea zenbat balio duen negoziotan erabiltzaileen hizkuntza erabiltzeak</u>
	<u>Smartphoneetarako app-ak garatzea</u>
<u>Sare sozial lokalizatuak</u>	<u>Sare sozialen erabiltzaile-interfazeak lokalizatzeko ekimenak</u>
<u>Software lokalizatua: sistema eragileak eta oinarrizko softwarea</u>	<u>Hizkuntza-komunitatean gehien erabiltzen den xede orokorreko software librea edo jabeduna lokalizatzeko ekimenak indartzea</u>
	<u>Bideo-jokoak aintzat hartzea biziberritze-aukera baliotsu gisa</u>
<u>Itzulpen automatikoko zerbitzuak</u>	<u>Hizkuntza-konbinazio kopurua handitzea; saiatzea ingelesa sartzen, jada sartuta ez badago</u>

1. Euskara hizkuntzaren profila

Euskara hizkuntza ez-indoeuropar bakarra da mendebaldeko Europan; beste inolako hizkuntzarekiko ahaidetasunik ezagutzen ez zaionez, hizkuntza isolatua dela esan ohi da. Gaur egun onartuena den sailkapenaren arabera, euskara euskalki hauetan banatzen da ekialdetik mendebaldera: gipuzkera, bizkaiera, lapurtera, goi-nafarrera, behe-nafarrera eta zuberera (Hualde eta Ortiz de Urbina, 2003)¹.

Euskara batua sortzeko lanak 1960ko hamarkadan hasi ziren eta Euskaltzaindiak 1968an argitaratu zituen lehen erabakiak (Arantzazuko Biltzarrean) orduan hasi zen, lehenengo aldiz eta ofizialki, euskara batuaren bidea. Hortik aurrera, estandarizazio-prozesua oso azkar joan zen.

Prozesuaren hasieran zailtasun batzuk izan ziren arren, estandarizazioak arrakastatsua izan da. Orokorrean, euskal gizartearen ondo ezarria eta onartua izan da (Hualde eta Zuazo, 2007)².

Euskarak 800.000 hiztun inguru ditu Pirinio mendebaldeko bi aldeetan: hegoaldeko lau probintzietan (Araba, Bizkaia, Gipuzkoa eta Nafarroa) Espainiako iparraldean; eta Ipar Euskal Herriko hiru lurralde historikoetan (Lapurdi, Nafarroa Beherea eta Zuberoa). Zazpi probintzia horiei Euskal Herria deritze.

Gaur egun koofiziala da Euskal Autonomia Erkidegoan (EAE), Araba, Bizkaia eta Gipuzkoan. Nafarroan aldiz, iparraldean bakarrik da ofiziala, eta Ipar Euskal Herrian ez da ofiziala.

NBERen arabera, euskara Euskal Autonomia Erkidegoan egoera ahulean dago, Nafarroan batez ere; eta Ipar Euskal Herrian (alde frantziarrean) desagertzeko arriskuan dago. EAEn egoera hobea da eta biziraupena bermatua duela dirudi.

Hezkuntzari dagokionez, egoera oso desberdina da eremuaren arabera (Gardner eta Zalbide, 2005)³. EAEko eta Nafarroako ikastetxe gehienetan euskaraz edo eredu elebidunean ikas daitekeen bitartean, Nafarroa hegoaldean eta iparraldean askoz gutxiago irakasten da euskaraz.

Unibertsitate mailan antzekoa gertatzen da. Egoera oso desberdina da eremuaren arabera, eta unibertsitate espezifikoaren arabera. Euskal Herriko Unibertsitatean (UPV-EHU) adibidez, ikasleek partzialki edo ia erabat ikas dezakete euskaraz, baina beste unibertsitateetan euskarazko ikasketak askoz urriagoak dira, gaztelaniaz edo frantsesez eskaintzen direnekin alderatuta, nahiz eta egoera hori aldatzeko ahalegin handia egiten ari diren.

Euskararen presentzia Euskal Herriko hedabideetan nahiko mugatua da oraindik. Euskarazko egunkari bakarra dago Euskal Herri osoan banatzen dena, Berria; badira egunero euskaraz argitaratzen diren beste tokiko egunkari batzuk ere. Astekari espezializatuak eta askotarikoak ere badaude eduki guztia euskaraz argitaratzen dituztenak. Nabarmentzekoa da azken urteotan sortu den egunkari digitalen kopurua.

Gaur egun, bi euskal irrati publiko dira soilik euskaraz aritzen direnak (Euskadi Irratia eta Euskadi Gaztea). Aldi berean, badira lurraldeko, eskualdeko eta herrietako irratia euskara hutsean aritzen direnak ere. Euskara ez da oso erabilia telebistan; hiru dira euskara hutsean ematen dutenak: ETB1, ETB3 eta Hamaika Telebista. Zenbait euskarazko tokiko zenbait telebista-kate ere.

Hizkuntza-teknologiei dagokienez, euskararako sortutako hainbat produktu teknologia eta baliabide daude. Badira hiztegi digitalak, corpusak, ahots-sintesisirako eta ahots ezagutzarako aplikazioak, ortografia-eta gramatika-zuzentzaileak. Itzulpen-automatikoko tresnak garatu dira gaztelaniarako eta euskararako (Hernández et al., 2012)⁴.

1 Hualde, J. I. & Ortiz de Urbina, J. (eds.) (2003). *A grammar of Basque*. (Mouton Grammar library, 9.) Berlin: Mouton de Gruyter.

2 Hualde, J. I. & Zuazo, K. (2007). The standardization of the Basque language. In *Language Problems and Language Planning*, 31(2):142-168.

3 Gardner, N. & Zalbide, M. (2005). Basque Acquisition Planning. In *International Journal of the Sociology of Language*, 174, 55-72.

4 Hernández, I., Navas, E., Odriozola, I., Sarasola, K., Díaz de Ilarraza, A., Leturia, I., Díaz de Lezana, A., Oihartzabal, B., Salaberria, J. (2012). Euskara Aro Digitalean – Basque in the Digital Age. META-NET White Paper Series: Europe's Languages in the Digital Age. Springer, Heidelberg. <http://www.meta-net.eu/whitepapers/volumes/basque>

2. Euskararen bizitasun digitala

2.1 Euskararen bizitasun digitalaren ebaluazioa

Euskararen bizitasun digitala ebaluatzeko, Hizkuntzaren Bizitasun Digitalaren Eskala aplikatu dugu, "How to use the Digital Language Vitality Scale" (Ceberio et al., 2018) dokumentuan deskribatutakoa. Eskala aplikatzeko behar zen informazioa bildu genuen adituen ezagutza eta hitzunen ezagutza uztartuz; lehenengoa euskal kideak ([Elhuyar Fundazioak](#)) eta DLDPre zenbait aholkularik eman ziguten, eta bigarrena onlineko [galdetegi](#) baten bidez bildu genuen.

Gure ikerketan oinarrituta, "Garabidean" mailan (4) koka daiteke euskara, eta 4,8ko puntuazioa lortu du. "Garabidean" maila hori honela deskribatzen da:

"Bizitasun digitalaren "Garabidean" mailan dagoen hizkuntza bat hein bateraino erabiltzen da komunikazioan eta sare sozialetan, baina litekeena da maiztasuna oraindik noizbehinkakoa izatea. Baliteke zenbait hedabide eta zerbitzu digital eskura egotea, bai eta tamaina ertaineko wikipedia bat ere; badaude oinarritzko hizkuntza-baliabide batzuk, eta litekeena da aurreratuagoren bat ere egotea. Hitzunen komunitateak erabiltzen dituen sare sozialetatik eta sistema eragileetatik bat, behintzat, lokalizatuta egon daiteke. Baliteke itzulpen automatikoko onlineko zerbitzu edo tresnaren bat eskura egotea, hizkuntza-konbinazio baterako, behintzat".

Beheko taulan hizkuntzaren bizitasun digitalerako erabilitako adierazleak eta euskarari dagokion puntuazioa jaso dira⁵:

Konexio ebidentzia	2
Alfabetatze digitala	4
Kodifikazioa	6
Digitalki gaitutako biztanleria-kopurua	6
Hizkuntza-baliabideen eskuragarritasuna	5
Komunikazio elektronikorako erabilera	5
Sare sozialetako erabilera	5
Interneteko hedabideen	5
Wikipedia	5
Interneteko zerbitzuen eskuragarritasuna	5
Sare sozial lokalizatuak	5
Software lokalizatua	5
Itzulpen automatikorako tresnak eta zerbitzuak	4
Domeinu espezifikoak	6

2.2 Erabilera digitala: nola, zer maiztasunekin eta zergatik erabiltzen duen jendeak euskara sarean

Inkestaren emaitzek erakusten dute euskara sasoi onean dagoela digitalki, eta aktiboki erabiltzen dela. Inkestatuek hizkuntza-gaitasun handia dute, eta ondo ezagutzen dituzte existitzen diren tresna eta baliabide digitalak. Hala ere, inkestaren hasieran aipatu bezala, inkestatuen profilak ez ditu erabat islatzen hitzun batzuen eta besteen profil soziolinguistikoak. Inkestatuen adinari dagokionez, gehienak 30 urtetik gorakoak dira. Jotzen da 15-34 urte bitartekoak direla aktiboena mundu digitalean⁶; beraz, ezin dugu baztertu talde hori azpiordezkatuta egotea inkesta honetan.

Inkestatu askok adierazi dute maiz erabiltzen dutela euskara Interneten (webguneetan, komunikazio elektronikotan, blogetan, Wikipedian...). Gauza bera sare sozialetan ere, bereziki Facebooken eta Twitterren, non euskara oso aktiboki erabiltzen baita. Halako sareak, gehienbat, erregistro informalekin lotuta daudenez, horrek agerian uzten du euskara eguneroko ahozko eta idatzizko onlineko komunikazioan erabiltzen dela eta erabilgarria dela horretarako, eta hori bizitasunaren seinale ukazina da. Hori faktore garrantzitsua da hizkuntza batek mundu digitalean biziraungo badu. Aldiz, euskara gutxiago erabiltzen da LinkedIn-en, negozioetara eta enplegura

5 Adierazleei eta puntuazioari buruzko argibideak dokumentu honetan daude: Ceberio et al. 2018. "How to use the Digital Language Vitality Scale". DLDP. <http://wp.dldp.eu/wp-content/uploads/2018/08/HowToUseTheDLVS.pdf>

6 <https://www.statista.com/statistics/272365/age-distribution-of-internet-users-worldwide/>

bideratutako sare sozialean. Hori euskararen egoeraren isla izan daiteke; izan ere, erabilerak gora egin du familia-ingurunean eta ingurune informalean, baina lan gehiago egin behar da testuinguru profesional eta formaletan.

Euskaraz lokalizatutako zerbitzu eta interfaze digitalei dagokienez, aipatzekoa da inkestatu gehienek badakitela existitzen direla aipatutako zerbitzuen bertsioak euskaraz. Hala ere, jakinaren gainean egon arren, inkestatu batzuek ez dute euskara erabiltzen beren gailu, aplikazio edo softwareetan. Horrek arrazoi asko izan ditzake, baina inkesta honek informazioa ematen digu euskal hiztunen ohitura digitalei buruz, kontuan izanik: inkestatuen herenak, gutxi gorabehera, uste duela errazago dela espainierazko tresnak erabiltzea; euskarazko softwarea aurkitzea eta instalatzea ez dela beste hizkuntzetan bezain erraza; eta, ondorioz, erabiltzaileek ahalegin gehigarria egin behar dutela.

Inkestatu batzuek eskatzen dute eskura dauden euskarazko baliabide guztiak zerrendatuta eta erabiltzaileentzat deskargatzeko prest dauden webgune bat. Horrez gain, aipatu da zabaldu beharra dagoela existitzen diren euskarazko zerbitzuei eta interfazei buruzko informazioa, bereziki gazteen artean.

Azkenik, adierazi nahi genuke eskari handiagoa dagoela euskarazko entretenimendu-produktuetarako, bereziki gazteentzat. Jende gehienak beste hizkuntza batzuetan kontsumitzen ditu ordenagailurako edo mugikorerrako jokoak, joko horiek euskaraz aurkitzea zailagoa egiten baitzaie.

3. Bizitasun digitala hobetzeko gomendioak

Hizkuntzak Mundu Digitalean Bizirauteko Kitaren xedea da eskualde-hizkuntza edo hizkuntza gutxituen hiztunei eta komunitateei aukera ematea, batetik, beren hizkuntzaren bizitasun-maila ebaluatzeko, eta, bestetik, ikustea zer nolako ekintza eta ekimen zehatzek hobetu dezaketen bizitasun-maila hori. Dokumentu honek bigarren helburu hori du hizpide, eta euskararen berriazko kasura dago egokituta.

Gomendio-sorta honetan, egin daitezkeen zenbait ekintza proposatzen ditugu (gehienbat herri-mailakoak), hizkuntza bati aurrera eginarazteko bizitasun digitalaren hurrengo urratserantz.

Gomendioak hiru ataletan daude banatuta, eta atal bakoitza bizitasun digitalaren adierazle-mota bati dago lotuta.

3.1 Bizitasun digitalaren hiru motatako adierazleak

3.1.1 Gaitasun digitala

Guretzat, *gaitasun digitalak* esan nahi du zenbaterainoko azpiegitura-euskarria eta euskarri teknologikoa duen hizkuntza batek mundu digitalean funtzionatu ahal izateko. Aurrebaldintza bat da hizkuntzak idazketa-sistema bat eduki behar duela, gutxienez; halakorik ezean, ezinezkoa baita mundu digitalean funtzionatzea. Oinarrizko baldintza batzuk bete behar dira, hala nola Internetarako konexioa edukitzea eta alfabetatze digitala, komunitate batek hizkuntza bat mundu digitalean erabiliko badu. Era berean, hizkuntza-baliabideak eta tresnak existitzeak eta eskura egoteak zerikusi handia du hizkuntza batek testuinguru digitaletan funtzionatzearekin. Adibidez, smartphoneetan zuzentzaile ortografikoak edukitzeak susta dezake, printzipioz, hizkuntza hori erabiltzea, aukera ematen baitu errazago eta azkarrago idazteko. Hizkuntza baten gaitasun digitalak hura mundu digitalean erabiltzeko potentzialari bakarrik egiten dio erreferentzia, baina ez du inola ere bermatzen komunitate batek erabili egingo duela. Halaxe gertatzen da, esaterako, Europako eskualde-hizkuntza edo hizkuntza gutxitu askorekin: nahiz eta hizkuntza horietako gehienek betetzen dituzten gaitasun digitalerako eskakizun guztiak, askotan, gutxi erabiltzen dira hizkuntza horiek hitz egiten diren herrialdeetako estatu-hizkuntza ofizialaren aldean. Erabilera digital eskasa beste faktore batzuen ondoriozkoa ere izan daiteke; adibidez, sumisio psikologikoa, idatzizko hizkuntzan gaitasunik ez izatea, edo espazio digitalik ez edukitzea (foroak, blogak) hizkuntza erabiltzeko.

3.1.2 Presentzia eta erabilera digitala

Gaitasun digitalaren azpiegitura-maila bermatuta, hainbat hedabidetan erabil daiteke hizkuntza, eta xede ugariarako. Adierazleen bigarren multzoak (6.etik 9.era) adierazten du nola eta zenbat erabiltzen den hizkuntza bat mundu digitalean: ea erabiltzen den (eta zenbateraino) komunikatzeko, sormenezko edukia sortzeko edo hezkuntza- edo entretenimendu-xedeetarako, besteak beste. Adierazle-multzo horren izendatzaile komuna da hizkuntza horretan eduki digitala sortzeari lotuta daudela, komunikaziorako edo beste xede batzuetarako erabiltzen dela ere. Esan bezala, nolabaiteko goranzko progresio bat iradokitzeko moduan daude ordenatuta adierazleak: testu-mezuak, mezularitza eta posta elektronikoa, adibidez, hizkuntza horretan Wikipediako artikulua idaztea edo liburu elektronikoak edo bideo-jokoak garatzea baino funtzio oinarrizkoagozat jotzen dira. Hala ere, horrek ez du esan nahi ordena hori mailaz maila igotzeko eskaileratzat hartu behar denik. Beraz, ez da nahitaezkoa Wikipedia edukitzea bideo-jokoak sortzen edo lokalizatzen hasteko. Komunikazio-funtzioa beste funtzio batzuk baino oinarrizkoagozat jotzen da, Gibson-en arabera [8]. Hizkuntzaren erabilera digital horiek progresio bat dakarte, halaber, hizkuntzaren erabilera pribatuagoetatik erabilera publikoago eta, askotan, ofizialetara. Mota horretarako, lau adierazle nabarmendu ditugu: *komunikazio elektronikorako erabilera, sare sozialetako erabilera, Interneteko hedabideen eskuragarritasuna eta Wikipedia.*

3.1.3 Errendimendu digitala

Hizkuntza batekin mundu digitalean zer egin daitekeen adierazten duten adierazleak biltzen ditu errendimendu digitalak. Hala, beste modu batera ikusten da zenbateraino erabiltzen den hizkuntza bat zenbait hedabidetan. Ikuspegi horrek gehiago aztertzen du zer xedetarako erabiltzen den hizkuntza, zenbat hedabidetan erabiltzen den baino. Talde horretarako, bost adierazle identifikatu ditugu. *Interneteko zerbitzuen eskuragarritasuna, sare sozial lokalizatuak, software lokalizatuak, itzulpen automatikorako tresnak eta zerbitzuak.*

3.2 Gomendioen egitura

Gomendio bakoitza honela dago egituratuta:

- » Azalpen-testu bat, gomendioa arrazoitzeko eta deskribatzeko.
- » Gomendioak xede dituen **hartzailleak**: batetik, partikularrak, erabiltzaileen taldeak eta elkarteak, eta, bestetik, erakundeak, ikerketa-taldeak, software-garatzailleak eta enpresak.
- » **Adibideak**: gomendioan proposatu diren ekimenak gauzatu diren kasu arrakastatsuak edo interesgarriak, edo gomendioa nola inplementa daitekeen erakuts dezaketenak.
- » **Gehiago irakurtzeko**: gomendioari buruzko informazio gehiago ematen duten artikulua, blog-sarrerak edo txosten akademikoak.
- » **Trebakuntza Programako (TP) horri buruzko modulua**, gomendioari dagokion informazioa ematen duena.

4. Gaitasun digitala

Ekintza-ildo nagusia: Zure hizkuntza ingurune digitalerako prestatzea

- » Oinarrizko gaitasun gisa, sustatu eskualde-hizkuntzako edo hizkuntza gutxituko alfabetatzea.
- » Bermatu konexio ona eta eguneratua eta Interneterako sarbide orokortua.
- » Sustatu eskualde-hizkuntzaren edo hizkuntza gutxituaren hiztunen (erabiltzaile digital potentzialak) gaitasun digitala (ertain-altua).
- » Garatu hizkuntza-baliabideak eta tresnak, eta harrarazi parte horretan zenbait eragileri (erabiltzaileen komunitateak, ikerketa-taldeak, enpresak, eragile politikoak)

Hizkuntza-baliabideen eskuragarritasuna

Hizkuntza-baliabideak ezinbesteko baldintza dira hizkuntzan oinarritutako ordenagailurako aplikazio aurreratuagoetarako.

Nahiz eta ohikoa izan hizkuntza-baliabide terminoa erabiltzea, kontuan izan behar da tresnak ere sartzen direla izendapen horretan. Baliabideen artean daude hiztegiak, testu-bildumak edo corpusak, gramatikak, datu-base lexikalak eta terminologikoak, eta ezagutza-baseak eta ontologiak. Tresnen artean, berriz, honako hauek daude: zuzentzaile ortografikoak eta estilistikoak, aztertzaile morfosintaktikoak, kategoria gramatikalen etiketatzaileak (POS taggers) eta aztertzaile sintaktikoak, terminologia-erazleak eta hitz anitzeko esapideak (Multi Word Expressions, MWE) erazleak, itzultzaile automatikoak eta hizketa-sintesisirako eta -ezagutzerako tresnak.

Hizkuntza batek hizkuntza-baliabideei dagokienez duen bizitasun-maila ebaluatzeko, "How to use the Digital Language Vitality Scale" (Ceberio et al., 2018) dokumentuan, hiru mailatan sailkatu ditugu baliabideak eta tresnak: oinarrizkoak, erdi-mailakoak eta aurreratuak.

- » Oinarrizkoak: hiztegi elektroniko elebakarrak eta elebidunak, corpus digitala (<100 milioi hitz), zuzentzaile ortografikoa.
- » Erdi-mailakoak: corpusean oinarritutako hiztegi elebakarra, corpus digitala (>100 milioi hitz), corpus paraleloak, web-corpusak, terminoak erazteko softwarea, kategoria gramatikalen etiketatzea, oinarrizko itzulpen automatikoa (arauetan oinarritua), hizketa-sintesia.
- » Aurreratuak: corpus handiak (>1.000 milioi), corpus eleaniztuna, aztertzaile sintaktikoak, WordNet, prozesatze semantikoa, itzulpen automatiko aurreratua (estatistikoa, hibridoa, neurala), hizketa-ezagutzea.

Jakina, sailkapen hori orientaziozkoa besterik ez da, eta ez da zorrotzegi hartu behar; baina erabilgarria izan daiteke adierazle honen balioak egituratzeko eta beheko gomendioak antolatzeko.

R3 Oinarrizko hizkuntza-baliabideak garatzea

R3.4 Konkordantziak eta antzeko tresnak erabiltzea corpora kontsultatzeko

Esan bezala, oso garrantzitsua da hizkuntza batek baliabideak edukitzea, adibidez, testu-corpusak.

Badira tresna erraz bezain eraginkorrak, konkordantzia-programak, oso lagungarriak izan daitezkeenak hitzei eta haien erabilerari buruzko informazioa lortzeko, eta prozesatu gabeko testuekin erabil daitezkeenak, hau da, hitzei buruzko informazio linguistikoa eransteko. Hauei buruzko informazioa lor daiteke: hitzen maiztasuna, konbinazio ohikoenak edo kolokazioak; eta emaitzak erabil daitezke hiztegi-gintzarako eta hizkuntza prozesatzeko beste zeregin batzuetarako.

Hala ere, halako baliabideak ahalik eta modu eraginkorrenean erabiltzeko, egoerarik onena da testu horiek linguistikoki prozesatuta egotea. Hitz edo lema batek zenbait forma edo token

(osagai lexikoak) har ditzake, hala nola zenbait genero-forma, plural-forma, aditz irregularrak edo, hizkuntza batzuetan, inflexio-formak. Hala, kalitateko datuak lortzeko, gomendagarria da testua aztertuta edukitzea, oinarrizko mailan behintzat; adibidez, jakitea zer lema dagokion testuko hitz bakoitzari, eta zein den haren kategoria gramatikala. Urrats bat aurrerago joanda, batzuetan, analisi morfologikoa ezinbestekoa izaten da, normalean eskualde-hizkuntzak edo hizkuntza gutxituak aberatsak izaten baitira morfologikoki, eta horrek, askotan, datu-eskasia eragiten du.

Kategoria gramatikalen etiketatzaileak (POS tagger) automatikoki egiten du lan hori, eta mugari garrantzitsua da hizkuntza baten garapen teknologikoan. Zure hizkuntzan halako tresnarik badago, gomendagarria da erabiltzea, eta corpusa arakatzea, etiketatu denean.

Hartzaileak: partikularrak, kolektiboak, ikerketa-taldeak

Adibideak:

- » [AntConc 3.4.0 Tutorial 1: Getting Started – YouTube](#)
- » [TextSTAT – Tutorial - YouTube](#)

Gehiago irakurtzeko:

- » [AntConc: A freeware corpus analysis toolkit for concordancing and text analysis](#)
- » [TextSTAT – Simple Text Analysis Tool](#)
- » [What are the most useful programs for forming text corpus or dictionary?](#)

Horri buruzko modulua TPn: 6

R4 Erdi- eta goi-mailako hizkuntza-baliabideak garatzea

Maila honetan, litekeena da erabiltzaileek jada eskura edukitzea oinarrizko baliabideak, eta erdi-mailako zenbait baliabide jada garatuta egotea. Puntu honetan, garaia da erdi-mailako baliabideen sorta osatzeko, eta baliabide eta tresna aurreratuak garatzea planteatzeko.

R4.1 Hiztegi gintza: dibertsitatea, tamaina, espezializazioa eta dibulgazioa

“Garabidean” mailan, litekeena da hiztegi elektronikoko bat baino gehiago egotea —gehienak elebidunak izango dira—, eta hiztegi horiek, askotan, sarean egon ohi dira.

Hizkuntzaren berriazko egoeraren arabera, zenbait ekintza egin daitezke, dela hiztegi berriak sortzeko, dela lehendik daudenak aberasteko.

Lau hitz gako hauen inguruan eman daitezke gomendioak: dibertsitatea, tamaina, espezializazioa eta dibulgazioa. Bestela esanda, “Garabidean” mailan dagoen hizkuntza batek hiztegi elebidunak eta elebakarrak behar ditu, tamaina eta estaldura desberdinetakoak, eta zenbait erabiltzaile-mota eta xedetarako zuzenduak (ikasleak, itzultzaileak edo arloko espezialistak). Horrez gain, funtsezkoa da hiztegi horiek Interneten eskuragarri egotea, eta, bereziki, kontsultarako app-ak edukitzea.

Hiztegi bat egitea adituentzako zeregina da, baina hizkuntza baten hiztun gai batek har dezake parte hiztegi bat egiteko proiektu bateko zenbait zereginetan behintzat. Gaur egungo teknologiari esker, duela, jo, hogeitau urte baino askoz errazago sor, edita eta argitaratu daitezke hiztegiak.

Hala ere, zenbait motatako proiektuek giza, teknologia- eta finantza-azpiegitura behar dute, eta hori zaila izan daiteke hizkuntza gutxituentzat. Hori dela eta, proiektuak, tradizionalki, erakundeen parte-hartzearekin edo bultzadarekin egin dira, edo argitaratzaile pribatuenarekin, zeinek papereko bertsioak saltzean oinarritutako negozio-eredua baitute.

Hori guztia oso azkar ari da aldatzen, gero eta papereko hiztegi gutxiago saltzen baitira urtetik urtera; hori hain da horrela, non ez baita errealizatzen jotzen merkatuak, gaur egun, hizkuntza gutxitu baten hiztegi baten argitalpena finantzatzea, are gutxiago hiztegi elebakar batena.

Beraz, babes publikoan, mikromezenasgoan eta elkarlanean oinarritzen da hiztegegintza, batez ere hizkuntza gutxituetan.

Horiek horrela, lehenengo gomendioa da ekimenak abiaraztea erakundeei eta hiztun-taldeei halako proiektuetan parte harrarazteko.

Behean, hiztegiak prestatzeko eta argitaratzeko lana erraztu dezaketen tresnen zenbait erreferentzia ematen dira.

Hartzaileak: kolektiboak, enpresak, erakundeak.

Adibideak:

Hiztegi lexikografikoak (elebakarrak edo elebidunak):

- » [FieldWorks \(FLEx Quick Tour on Vimeo\)](#)
- » [Lexonomy \(Gentle introduction to Lexonomy\)](#)
- » [Tshwanalex \(TLex Suite: Dictionary Compilation Software\)](#)¹

Hiztegi tekniko edo espezializatuak, proiektu terminologikoak:

- » [TermWiki In My Language | TermWiki.com](#)
- » [TermKate — An integral web platform for the creation and publishing of terminological dictionaries](#)

Hiztegi-aplikazioak:

- » [SIL Dictionary app builder](#)

Further reading:

- » Kroskryty, P.V. (2015). [Designing a dictionary for an endangered language community: Lexicographical deliberations, language ideological clarifications](#). University of Hawaii Press.
- » Garret, A. (2018). [Online dictionaries for language revitalization](#), to appear in [The Routledge handbook of language revitalization](#), edited by Leanne Hinton, Leena Huss, and Gerald Roche (Routledge, 2018).
- » Kotorova, E. (2016). [Dictionary for a Minority Language: the Case of Ket](#). In Proceedings of the XVII EURALEX International Congress.

Horri buruzko modulua TPn: 6

R4.2 Corpusaren tamaina eta dibertsitatea handitzea

Bizitasun digitalaren “Garabidean” mailan dagoen hizkuntza batean, corpus-baliabide bat edo gehiago egongo dira, ziurrenik. Hurrengo urratsa da corpus handi samar bat izatea helburu, bai eta corpus elebidunak edo paraleloak eta corpus espezializatuak prestatzen hastea ere.

Nahiz eta, zentzu zabal batean, testu-bilduma oro jo daitekeen, printzipioz, corpustzat, askori iruditzen zaie zenbait baldintza daudela horretarako. Lehenik eta behin, erabilgarria izatea hizkuntza orokorraren erabilerari buruzko datuak lortzeko (erreferentzia-corpusa), edo arlo, genero, erregistro edo garai jakin bateko erabilerari buruzko datuak lortzeko (corpus espezializatu). Helburu hori da gai eztabaidatuenetako bat corpusak diseinatzean, eta zerikusia du laginaren tamainarekin eta ordezkagarritasunarekin, edo, gutxienez, sartutako testu-mota batzuen eta besteen arteko orekarekin. Bigarrenik, gaur egun ia ezinezkoa da corpus bat ez egotea formatu digitalen. Azkenik, oinarritzko mailan, corpus batek prozesatu gabeko testua dauka, baina, batez ere morfologia aberatsa duten hizkuntzen kasuan, ia ezinbestekotzat jo ohi da corpusa linguistikoki komentatuta egotea, non, adibidez, lemari, kategoria gramatikalari eta token bakoitzaren kasuari buruzko informazioa jasotzen baita. Corpusak diseinatzeko eta eraikitzeko gomendioak hemen aurkitu daitezke: [Corpus design criteria](#) (Atkins et al., 1992) eta [Corpus Linguistics: corpus](#)

¹ NEz da doakoa, baina badago hizkuntza akademikorako edo desagertzeko arriskuan dauden hizkuntzetarako lizentzia berezi bat

[building principles](#) (Markus Dickinson, 2015); kodetzeari buruz, ikus [15 Language Corpora - The TEI Guidelines - Text Encoding Initiative](#).

Halako ezaugarriak dituen corpus bat oso informazio-iturri aberatsa da, zenbait xedetarako erabil eta kontsultatu daitekeena: hizkuntzaren erabileraren analisi linguistikorako, beste baliabide eta tresna batzuk garatzeko, hipotesiak frogatzeko eta hizkuntza-prozesamendu automatikorako tresnak trebatzeko. Adibidez, hiztegi-gintzari mesede handia egin dio corpus-analisiak: hitz berriei buruzko informazioa lor dezakegu, eta jakin dezakegu zein diren hitz erabilienak edo hitz baten aldaerak, edo zer adjektibo erabiltzen diren gehien izen jakin batekin.

Hala ere, baldintza horiek betetzea, batez ere lehenengoa, erronka bat izan daiteke hizkuntza nagusientzat ere — horiek, normalean, ez dituzte baliabide gutxi izaten —. Halako corpus bat eraikitzea lan handia da; ahalegin kolektiboa eskatzen du, bai eta dirua eta funtsak inbertitzea ere, eta hori eskualde-hizkuntza edo hizkuntza gutxituen komunitate batzuen aukeretatik kanpo egon daiteke.

Arazo nagusietako bat da nola lortu formatu digitaleko testu-kantitate handi samar bat. Argitaratzaileekin eta egileekin negoziatzea lan nekeza izan daiteke, eta lanak ez dira beti eskura egoten corpusean kostu txikian sar daitekeen formatu batean. Oztopo hori gainditzen lagundu diezagukeen aukera bat da sarean oinarritzea, edo sarea hartzea corpus gisa. Interneten, testuak digitalizatuta daude jada, baina estrategia horren bideragarritasuna ez dago bakarrik eskualde-hizkuntza edo hizkuntza gutxituan dagoen web-eduki kantitatearen mende, baita eduki hori aurkitzeko eta biltzeko dugun gaitasunaren mende ere. Esaterako, gure xede-hizkuntzan idatzita dauden orrialdeak identifikatu behar ditugu. Ez baduzu modurik horretarako tresna bat garatzeko, badaude tresna batzuk zeregin hori errazten dutenak, hala nola [TextSTAT](#) eta [BootCaT](#). Lehenengoak miaketa-sistema bat erabiltzen du, eta zehaztu behar duzu zer URL harrapatu nahi duzun. Identifikatzeko zer webgunek duen edukia lanean ari zaren hizkuntzan, lagungarria izan daiteke hizkuntza-komunitateari eskatzea eman diezazkizutela hizkuntza horretan dauden orrialde ezagunen URLak. BootCat-ek beste funtzionalitate bat gehiago du: jatorri-hitzen multzo bat espezifika dezakezu, eta, Bing erabiliz bilaketa-tresna gisa, tresnak hitz horiek konbinatzen ditu, zenbait bilaketa egiteko eta eskuratutako dokumentuak biltzeko. Zure hizkuntzan bakarrik existitzen diren hitzak erabil ditzakezu, edo arlo jakin batekoak (arloko corpus espezifikoak eraikitzeko).

Edukiak biltzeko beste modu bat testu-biltegiak dira. Metodo horrek lotura estua du gomendio honekin: Dokumentuen eta erregistroen fitxategi bat (sarean oinarritua) sortzea (atal honetan: Presentzia eta erabilera digitala).

Eskualde-hizkuntza edo hizkuntza gutxitu baten corpus elebakar eta orokor baten tamainari buruz ari garenean, erronka polita izan daiteke 100 milioi hitzetara iristea. Nahiz eta erabili izan diren corpus txikiagoak eskualde-hizkuntza eta hizkuntza gutxituei buruzko ikerketa eta proiektuetan, eta nahiz eta nekez jar dezakegun zalantzan haien erabilgarritasuna, arrazoi sendoa dago corpusak handia izan behar duela esateko. Hizkuntza naturalak prozesatzeko tresna batzuk, datuetan oinarritutakoak, datu linguistikoetan trebatzen dira (askotan, prozesatu gabeko edota komentatutako testu eran), eta haien kalitatea corpusaren tamainaren oso mende dago. Corpusak erabilgarriak dira, halaber, erreferentzia-bertsioetako hizkuntza-tresnak ebaluatzeko, zeinak gizakiek berrikusten baitituzte eta oso baliotsuak baitira.

Corpus paraleloak bata bestearen itzulpena diren testuen bildumak dira (alegia, sorburu-testu bat eta haren itzulpena). Corpus paraleloak oso baliabide baliotsua dira itzultzaileentzat, baina baita hiztegi elebidunak egiteko ere, jakin baitezakegu nola itzuli den hitz, esaera edo termino jakin bat. Corpus paralelo mota oinarritzako dokumentu elebidunen bilduma da. Baldin badakigu zer dokumentu diren bata bestearen itzulpena, esan dezakegu corpusa dokumentu-mailan lerrokatuta dagoela. Baina askoz erabilgarriagoa da esaldi-mailan lerrokatutako corpus paraleloak lortzea. Haien eraikitze prozedura ideal bat da itzulpen-memoriak (IM) hartzea abiapuntutzat. IMak giza itzulpenaren emaitza dira, [Ordenagailuz Lagundutako Itzulpen \(OLI\)](#) tresna bat erabili denean. Bestela, automatikoki lerrokatu ditzakegu dokumentu elebidunak, tresna hauek erabiliz, besteak beste: [Wordfast Aligner](#) edo [hunalign – sentence aligner](#). Beste aukera bat da Interneten testu elebidunak bilatzea. Lan hori ez da erraza, baina inspiragarriak izan daitezke eskualde-hizkuntza edo hizkuntza gutxitu batzuetan egin diren zenbait esperientzia; esaterako, Bitextor, ILSP

Focused Crawler eta PaCo2. Corpus paraleloen beste iturri bat izan daitezke zenbait hizkuntzetako wikipediatatik hartutako artikuluen atalak, artikuluko horiek bata bestearen itzulpena badira. Wikipediak ez dira berdin-berdinak. Zorrotz hitz eginda, corpus alderagarritzat jotzen dira. Baina badakigu eduki batzuk berdinak direla, eta zenbait teknika garatu dira eduki horiek aurkitzeko eta lerrokatzeko.

Azkenik, corpus espezializatuak datu-iturri fidagarriak dira hiztegi espezializatuak edo terminologikoak egiteko. Corpus orokorren kasuan bezala, corpus espezializatuak eraikitzeke modu tradizionala da arlo jakin bateko testuak eta dokumentuak biltzea, baina hauta dezakegu sarea corpustzat hartzen duen metodoa ere, zeina ikusi baita datu-iturri nahiko egokia dela (ikus, adibidez, Gurrutxaga et al., 2010). Arlo jakin batean espezializatutako corpus bat eraikitzeke saretik, oso erabilgarria izan daiteke jatorri-hitzen metodoa. BootCaT-ek aukera hori ematen du.

Hartzaileak: kolektiboak, ikerketa-taldeak, erakundeak.

Adibideak:

- » [BootCaT – Simple Utilities to Bootstrap Corpora and Terms from the Web](#) || [BootCaT front-end tutorial](#)
- » [TextSTAT – Simple Text Analysis Tool](#)
- » [Building your own corpus – TextSTAT and AntConc](#)
- » [ILSP Focused Crawler - ILSP NLP](#)
- » [Bitextor: the automatic bitext generator](#)
- » [hrWaC and slWac: Compiling Web Corpora for Croatian and Slovene](#)
- » [W2c- large multilingual corpus | Martin Majlis - Academia.edu](#)
- » [TenTen Corpus Family – Wikipedia](#)

Gehiago irakurtzeko:

- » Atkins, S., Clear, J., & Ostler, N. (1992). [Corpus design criteria](#). *Literary and linguistic computing*, 7(1), 1-16.
- » Gurrutxaga, A., Leturia, I., Pociello, E., San Vicente, I. & Saralegi, X., (2010). [Exploiting the Internet to build language resources for less resourced languages](#). In *7th SaLTmIL Workshop on Creation and use of basic lexical resources for less-resourced languages LREC 2010*, Valetta, Malta.
- » Lynn, T., Foster, J., Dras, M. & Dhonnchadha, E.U. (2012). [Active Learning and the Irish Treebank](#), In *Proceedings of the Australasian Language Technology Association Workshop 2012* (pp. 23-32), Dunedin, NZ.
- » San Vicente, I. & Manterola, I. (2012). [PaCo2: A Fully Automated tool for gathering Parallel Corpora from the Web](#). In LREC (pp. 1-6).
- » Smith, J.R., Quirk, C. & Toutanova, K. (2010). [Extracting parallel sentences from comparable corpora using document level alignment](#). In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 403-411). Association for Computational Linguistics.

Horri buruzko modulua TPn: 6

R4.3 Publikoki eskuragarri dauden datuak biltzea sare sozialetatik

DLDPk gomendatzen du ahalik eta gehien erabiltzea publikoki eskuragarri dauden sare sozialetako mezuak eta argitalpenak, eskualde-hizkuntza eta hizkuntza gutxituetan idatzitakoak. Batez ere Facebooken, badira orrialde publiko espezifikoak hizkuntza gutxituei lotutako kultur eta hizkuntza-gaiari inguruko eztabaidak sustatzen dituztenak. Hor, jendeak irudi bateko objektuak izendatzen ditu, alderatzeko ea denek erabiltzen duten hitz bera; beste batzuek, berriz, olerkiak eta kantuak partekatzen dituzte.

Zenbait bideri jarraitu dakieke sare sozialetako datuak aintzat hartzean:

- » Sarean miatutako testuekin osatutako corpus bat sare sozialetako datuekin integra daiteke;
- » Sare sozialetako testu-datuen datu-multzo bat erabil daiteke aztertzeke zenbateraino diren uniformeak eta koherenteak hizkuntza beraren aldaera batzuk eta besteak;

- » Garatzaileek datu-multzo bera erabil dezakete kategoría gramatikalen mailan lematizatzen edo komentatzen duen softwarea eraikitzeke.

Erronka horri ekin aurretik, nahitaezkoa da datuen lizentzia arretaz egiaztatzea.

Halako testu-mota zaratsutsuzat jotzen da askotan, erabiltzaileek sortutako eduki editatu gabea baita, eta ortografia- edo gramatika-akatsak eduki baititzake. Halako testuek izaera linguistikoko horretan oinarritutako corpus-mota espezifiko bat osatuko lukete, baina beste xede batzuetarako da erabilgarria (hizkuntzaren erabilerak denboran izan duen bilakaera aztertzeke, hizkuntzaren erabileraren denbora errealeko argazki bat egiteke, kode-alternantzia aztertzeke eta beste hizkuntza-fenomeno interesgarri batzuk aztertzeke).

Hartzaileak: ikerketa-taldeak, software-garatzzaileak.

Adibideak:

- » [The Corpus of Facebook Welsh Language Texts](#)
- » indigenoustweets.com

Gehiago irakurtzeko:

- » Honeycutt, C. & Cunliffe, D. (2010). [The use of the Welsh language on Facebook: An initial investigation](#). In *Information, Communication & Society*, 13(2), pp.226-248.
- » Lynn, T., Scannell, K. & Maguire, E. (2015). [Minority language Twitter: Part-of-speech tagging and analysis of Irish tweets](#). In ACL-IJCNLP 2015, p.1.
- » Lackaff, D. & Moner, W. J. (2016). [Local languages, global networks: Mobile design for minority language users](#). In *Proceedings of the 34th Annual International Conference on the Design of Communication (SIGDOC '16)*.

Horri buruzko modulua TPn: 6

R4.5 Corpusak aztertzeke tresnak erabiltzea, eta zure hiztegia elikatzea erabiltzen den hizkuntzari buruzko datuekin

Badira tresna erraz bezain eraginkorrak, [konkordantzia-programak](#), oso lagungarriak izan daitezkeenak hitzei eta haien erabilerari buruzko informazioa lortzeke, eta prozesatu gabeko testuekin erabil daitezkeenak, hau da, hitzei buruzko hizkuntza-informazioa eransteke aztertu ez diren testuekin. Hauei buruzko informazioa lor daiteke: hitzen maiztasuna eta konbinazio ohikoenak edo kolokazioak; eta emaitzak erabil daitezke hiztegi-intzarako eta hizkuntza prozesatzeko beste zeregin batzuetarako. Tresna-mota horiei buruzko informazio xeheagoa eta erreferentziak jada eman dira gomendio honetan: R3.1 Konkordantziak eta antzeke tresnak erabiltzea corpora kontsultatzeko.

Baina zure hizkuntzan kategoría gramatikalen etiketatzaile bat baldin badago, askoz aukera handiagoa dago hizkuntzaren erabilerari buruzko datuak lortzeke. Jakin daiteke zer maiztasunekin erabiltzen diren ez bakarrik testuko hitz-formak (tokenak), baita hiztegi-sarrerak ere (lemak); hiztegiak ez dauden hitz berriak aurkitu daitezke; eta ikertu daiteke nola konbinatzen diren hitzak esaerak, kolokazioak eta termino konplexuak osatzeko. Horretarako, corpora kontsultatzeko sistema bat behar da, edo corpora aztertzeke tresna bat (corpus-analysis.com webgunean, halako zenbait tresna aurkituko dituzu).

Hartzaileak: ikerketa-taldeak, software-garatzzaileak.

Adibideak:

- » [Corpkit: a tool for investigating text](#)
- » [The IMS Open Corpus Workbench \(CWB\)](#)
- » [NoSketch Engine](#)
- » [BlackLab - An open source corpus search engine](#)
- » [Unitex - GramLab](#)

Gehiago irakurtzeko:

- » [The IMS Open Corpus Workbench \(CWB\) Corpus Encoding Tutorial](#)
- » [5 Reasons: Anke Luedeling on "Corpus Linguistics"](#)

Horri buruzko modulua TPn: 6

R4.6 Hizketaren sintesi eta ezagutzarako lehenengo urratsak

Hizketaren sintesia giza hizketaren simulazioa da, ordenagailuz sortua. Hizketa-teknologiaren bi zeregin nagusietako bat da (bestea hizketa-ezagutzea da). Hizkuntza bat teknologia horien munduan barneratzen hasten denean, ohikoena sintesiarekin hastea da, zeregin hori hizketa-ezagutzea baino errazagoa baita.

Hizketa-teknologiak garatzeak aukera ugari ematen ditu, bereziki, desgaitasuna dutenentzako sistemak sortzeko (pantaila-irakurleak ikusmen mugatua dutenentzat, hizketa-sintetizatzaileak hizketa-urritasuna dutenentzat, edo ezagutze-sistemak gailu informatiko konbentzionalak erabiltzea eragozten dieten desgaitasunak dituztenentzat). Bestalde, ikusi da hizketa-sintetizatzaileak oso baliotsuak direla online ikasteko, ikasleek hiztun natiboekin harremanetan jarri ezin dutenean. Adibidez, hori bereziki eraginkorra izan da irlandera ikasten duten atzerritarrentzat, zaila egiten baitzaie irlanderaren ortografia ahoskatzea. Horrez gain, erabiltzailea hizkuntza naturala erabiliz ordenagailuarekin edo gailu mugikorrekin komunika daitekeen sistema bat osatzeko, beharrezkoak dira sintesia eta ezagutzea.

Hizketa-teknologiak garatzeko, maila tekniko aurreratua behar da. Horregatik, ikerketa-taldeei eta software-garatzzaileei zuzentzen zaie gomendio hau. Hala ere, sintesi-teknologia jakin batzuek, hala nola kateamendu-sintesiak, benetako hizketaren laginak erabiltzen dituzte, eta, puntu honetan, garrantzitsua da hiztunen parte-hartzea, lokuzioak eman baititzakete hizketaren datu-base bat sortzeko, eta beren ahotsarekin ere egin baitezakete ekarpena.

Hartzaileak: ikerketa-taldeak, software-garatzzaileak

Adibideak:

- » [The MARY Text-to-Speech System \(MaryTTS\)](#)
- » [Festcat - Catalan Speech Synthesis](#)
- » [eSpeakNG](#)
- » [abair.ie – The Irish Language Synthesiser](#)
- » [Aholab TTS](#)
- » [The Welsh National Language Technologies Portal - Speech](#)
- » [Mozilla Common Voice](#)

Gehiago irakurtzeko:

- » [Speech Synthesis for Minority Languages: A Case Study on Scottish Gaelic](#)
- » [Issues in Porting TTS to Minority Languages](#)
- » [Frisian TTS, an example of bootstrapping TTS for minority languages](#)
- » Chasaide, A. N., Chiaráin, N. N., Wendler, C., Berthelsen, H., Murphy, A., & Gobl, C. (2017). [The abair initiative: Bringing spoken irish into the digital space](#). In *Proceedings of Interspeech 2017*.

5. Presentzia eta erabilera digitala

Ekintza-ildo nagusia: Erabilera sustatzea, eta edukia sor eta partekatu dadin sustatzea

- » Aurkitu eta probatu moduak jendea bultzatzeko komunikazio elektronikoko pribatuan eta sare sozialetan beren eskualde-hizkuntza edo hizkuntza gutxitua erabiltzera;
- » Sustatu honelako eduki-motak sortzea: webguneak, blogak, foroak, bai eta Interneteko irratia eta telebista ere;
- » Eskualde-hizkuntza edo hizkuntza gutxituetako hedabideak igotzeko eta partekatzeke ekimenak;
- » Egin elkarlan irekia azpitolaketarako;
- » Wikipedia: sortu, editatu, zuzendu, eguneratu.

5.1 Komunikazio elektronikorako erabilera

R5 Aztertzea zenbat balio duen eskualde-hizkuntza edo hizkuntza gutxitua erabiltzeak pertsonen arteko komunikaziorako

Pertsonen arteko komunikazio digitalari dagokionez, komunikazio elektronikoa zenbait motatakoa izan daiteke. Bat-bateko mezularitzak, kasu, ahozkotasanaren ezaugarri asko ditu. Hala, ingurune digitalean, komunikazio elektronikoa erabiltzen ez den hizkuntza bat hitz egiten ez den hizkuntza bat da. Komunikazio elektronikoa beste alderdi garrantzitsu bat da erregistro formalagoetan erabiltzen dela, hala nola testuinguru profesionaleko mezua elektronikoa.

Hartzaileak: partikularrak, kolektiboak, enpresak, erakundeak.

Adibideak:

- » [Tribo, a loita normalizadora e cotiá entre os máis pequenos](#)
- » [Campaign to use Welsh in the European Football Cup](#)
- » [Mintzanet: an Option for Practicing the Basque Language Anywhere in the World](#)

5.2 Interneteko hedabideen eskuragarritasuna

R7 Eduki-kantitatea handitzea eta Interneteko hedabide-motak dibertsifikatzea

Interneten hizkuntza jakin batean edukia egotea hizkuntza horren bizitasunaren adierazle nabarmena da. Sareko edukia edonoren esku dago, eta webgune, blog, foro, audio edo bideo eran eduki gutxi duen hizkuntza batek ez du ikusgaitasunik eta presentziarik ingurune digitalean.

Zenbait neurri eraginkor har daitezke hizkuntza horretan lehendik dauden hedabide-kopurua handitzeko, eta mota gehiago sartzeko.

R7.1 Testu-motako edukiaren kantitatea handitzea (webguneak, blogak, foroak)

Hizkuntza orentzat oinarrizkoa eta funtsezkoa da sarean horrelako edukia egotea. Interneten presentzia handi samarra lortu duten eskualde-hizkuntza edo hizkuntza gutxitu gehienetan, web-eduki gehiena kategoria horietan dago.

Beraz, horrelako edukia sortzea eta argitaratzea sustatu eta errazten duen edozerk mesede egingo die hizkuntzari eta haren erabiltzaileei.

Gaur egun, hainbat aplikaziok eta zerbitzuek errazten dute lan hori. WordPress, Blogger eta Wix dira ezagunenetako batzuk, baina askoz aukera gehiago aurkitu ditzakezu “Gehiago irakurtzeko” ataleko erreferentzietan.

Hartzaileak: partikularrak, kolektiboak, erakundeak.

Adibideak:

- » [The Maya Tz’utujil initiative](#): the Maya Tz’utujil initiative is a space on Facebook, Twitter and WordPress to teach, learn, and broadcast the Tz’utujil language, which primarily covers the geographic national area of Guatemala, and includes multiple collaborative partners who have become fully linked to the digital initiative.
- » [Basque Language on the Web: Making an Impact](#)

Gehiago irakurtzeko:

- » [6 Differences Between Blogging in a Minority Language versus English](#)
- » [How to Choose the Best Blogging Platform in 2018 \(Compared\)](#)
- » [The 16 best free blogging platforms](#)
- » [Best 10 Free Website Builders | 2018’s Best Website Builders](#)

Horri buruzko modulua TPn: 5

R7.2 Dokumentuen eta erregistroen fitxategi bat (sarean oinarritua) sortzea edo elikatzea

Bizitasun digital oso txikia duen hizkuntza batentzat, dokumentuen eta erregistroen fitxategi bat (sarean oinarritua) sor daiteke. Hizkuntza hitz egiten ez dutenek ere egin dezakete ekarpena. Baliabidea erabilgarria izan daiteke hizkuntza bat berrezartzeko, edo interesa duten baina seguru ez dauden hiztunak hizkuntza erabiltzen hastera bultzatzeko.

- » Sortu eta elikatu zure hizkuntzako testu-biltegiak
- » Dibertsifikatu zure hizkuntzan dauden hedabide-motak, audio- eta bideo-edukia barne
- » Egin ekarpena lehendik dauden biltegiei, hala nola Wikitongues-i

Hartzaileak: kolektiboak, erakundeak, instituzioak, partikularrak.

Adibideak:

- » [Indigitization – Toolkit for the Digitization of First Nations Knowledge](#)
- » [Endangered Languages Documentation Programme \(ELDP\)](#)
- » [Greenstone Digital Library Software](#)
- » [Wikitongues](#)
- » [BasaBALIWiki](#)
- » [Yadiko Ukruri initiative ‘jitomagaro uai’](#)
- » [Cultural Codex](#)
- » [AIKUMA Project - preserving endangered languages](#)

Gehiago irakurtzeko:

- » Nichols, D.M., Witten, I.H., Keegan, T.T., Bainbridge, D. & Dewsnip, M. (2005). [Digital libraries and minority languages](#). In *New Review of Hypermedia and Multimedia*, 11(2), 139-155
- » [Nenek – A cloud-based collaboration platform for the management of Amerindian resource languages](#)

Horri buruzko modulua TPn: 5

R7.3 Saioak online ematea, software libreko tresnak erabiliz

Onlineko irratietxe bat modu eraginkorra izan daiteke hizkuntza bat zabaltzeko eta ikusgai egiteko, bereziki komunitate barreiatuen kasuan. Software librea egonez gero, nahiko erraza izan daiteke. Behean azaltzen diren istorioetan inspiroz zaitezke.

Hartzaileak: partikularrak, kolektiboak, erakundeak.

Adibideak:

- » [Brazil's First Indigenous Online Radio Station Uses Digital Media to Promote Native Languages and Communities](#)
- » [Meet the Young Ecuadorians Behind the First Kichwa-Language Radio Show in the US](#)

Gehiago irakurtzeko:

- » [Icecast](#), a streaming media (audio/video) server that can be used to create an Internet radio station.
- » [Top 5 Free Tools to Live Stream Your Event Online – Capterra Blog](#)

Horri buruzko modulua TPn: 5

R7.4 Istorio digitalak zeure hizkuntzan erregistratzea

Honela definitzen da istorio-kontaketa digitala (Barret, 2005):

“Istorio-kontaketa digitala istorioak kontatzeko antzinako artearen adierazpen modernoa da. Istorio digitalek irudiak, musika, narratiba eta ahotsa nahasiz hartzen dute indarra, eta, hala, dimentsio sakona eta kolore bizia ematen dizkiete pertsonaiei, egoerei, esperientziei eta pertzepzioei”.

Istorio-kontaketa digitala hezkuntzan erabiltzen da, besteak beste (Hoven, 2009):

“Hizkuntzak ikasteko eta hizkuntza-irakasleak trebatzeko munduko hezkuntza-programa askotan, gero eta gehiago erabiltzen da istorio-kontaketa digitala hausnarketa bidezko ikasketa edo ebaluazioa bultzatzeko. Istorio-kontaketa digitalak baliabide ugari eman diezazkieke hizkuntza-irakasleei eta ikasleei, ikasleei laguntzeko beren burua atzerriko hizkuntzekin eta kulturekin edo galtzen ari diren lehenengo hizkuntzekin (H1) eta kulturekin (K1) identifikatzen, eta erosoago sentitzen hizkuntza horiek benetako xedeetarako erabiltzeko”.

Ikuspegi horretatik, baliagarria izan daiteke hizkuntza biziberritzeko.

Hartzaileak: Partikularrak.

Adibideak:

- » [Digital Storytelling: What it is... And... What it is NOT](#)
- » [How Storytelling Can Do Wonders in Blogging](#)

Gehiago irakurtzeko:

- » [Frequently-Asked Questions about Digital Storytelling](#)
- » Hoven, D. (2009). [Digital Storytelling in Indigenous Education: Internet Technologies to \(Re-\) Establish L1 and C1 Literacy and Fluency](#). In *Internet-Based Language Learning: Pedagogies and Technologies*, p.47.

Horri buruzko modulua TPn: 5

R7.5 Azpitolazio-ekimenak sustatzea

Ingurune digitalean, elkarlanean lan egiteko aukerek mesede egin diote filmak eta bideoak azpitolatzeari. Hainbat proiektutan, boluntarioek film baten azpitoluak itzultzen dituzte, gero eta hizkuntza gehiagotara itzuli ere. Hori aukera bat da eskualde-hizkuntza edo hizkuntza gutxituentzat. Lehen zaila eta garestia bazen ere, orain egia bihur daiteke eskualde-hizkuntza edo hizkuntza gutxituen hiztunen elkarlanari esker, dela jatorrizko bertsioko filmak ikusteko beren eskualde-hizkuntza edo hizkuntza gutxituko azpitoluekin, dela proiektio handiagoa emateko eskualde-hizkuntza edo hizkuntza gutxituan ekoiztako jatorrizko lanei, haiek hizkuntza handiagoetan azpitolatuz.

Hartzaileak: partikularrak, kolektiboak.

Adibideak:

- » [Amara: Caption, Subtitle and Translate Video](#)
- » [PerMondo – Introduction to subtitling](#)
- » [TED translations](#)
- » [Contribute translated content - YouTube Help](#)

Gehiago irakurtzeko:

- » [Crowdsourcing Subtitles for Endangered Languages](#)
- » [Review: Amara is a Web-based service that lets anyone transcribe and translate online video](#)
- » [How Crowdsourced Video Translation Works: Webinar Q&A with Amara](#)
- » [Is crowdsourcing translation a threat or an opportunity for the audiovisual market?](#)
- » [Azpitoluak, a Project for Basque Subtitles](#)
- » [Dowling, M., Lynn, T. & Way, A. \(2017\). A Crowd-sourcing Approach for Translations of Minority Language User-Generated Content. In Proceedings of 1st Workshop on Social MT, Prague, Czech Republic.](#)

Horri buruzko modulua TPn: 5

5.3 Wikipedia

Oso litekeena da wikipedia bat egotea zure hizkuntzan, baina baliteke txikia izatea, eta hiztunek ez erabiltzea edota ez izatea haren berri. Kontuan izanik Wikipedia hizkuntza baten bizitasunaren adierazle nabarmena dela, garrantzitsua da hura haztea eta hobetzea. Wikipediaren zerrenda xehean, ikus dezakezu ea badagoen wikipedia bat zure hizkuntzan, eta, hala bada, zein diren haren zifra garrantzitsuenak, hala nola artikuluko kopurua edo erabiltzaile aktibo kopurua.

Ez badago wikipediarik zure hizkuntzan, gomendatzen da proiektu bat hastea edo Wikipediaren erabiltzaile-interfazea itzultzea.

R9 Igo ezazu zure Wikipedia maila bat gorago

R9.2 Wikipediaren tamaina eta kalitatea hobetzeko ekimenak

Jada badago wikipedia bat hizkuntza horretan, ziurrenik tamaina ertainekoa (10.000 eta 100.000 artikuluko bitarte), baina, kontuan izanik Wikipedia hizkuntza baten bizitasunaren adierazle nabarmena dela, garrantzitsua da hango artikuluko kopurua eta haien luzera handitzea, eta haien kalitatea hobetzea.

Wikipediak mota askotako edukiak ditu, eta, horrenbestez, aukera ematen du zenbait arlo eta alderditan lan egiteko, hura hobetze aldera. Hemen, artikuluei dagozkien ezaugarri hauek hartuko ditugu aintzat:

- » **Kopurua**
Kopurua handitzea beti da positiboa, baina, "Garabidean" dagoen hizkuntza batentzat erronka

bat ezartzeko, 100.000 sarreratarata iristea edo hurbiltzea lorpen handia litzateke. Esan gabe doa mugarrari hori gainditzea berebiziko arrakastatzat jo behar litzatekeela.

» **Mota**

Bultzatu erabiltzaileak eskualde-hizkuntza edo hizkuntza gutxituko wikipedian tokiko edo eskualdeko edukia sartzera. Erabiltzaile gehiago erakartzeko, aukera bat da zure eskualde-hizkuntza edo hizkuntza gutxituko wikipedia besteetatik bereiztea. Horretarako, beste wikipedietan ez dagoen informazioa sartu beharko litzateke, edo wikipedia horietako informazioa baino informazio xeheago eta landuagoa. Baina, bereziki maila honetan, horrek ez du esan nahi artikulua orokorragoak sartu behar ez direnik, hala nola zientzietakoak, humanitateetakoak, historiakoak, arteari buruzkoak edo erabiltzaileentzat eta haien egunerokorako garrantzitsua izan daitekeen beste edozein eduki-mota (pertsona ospetsuei, markei, kirolei eta abarri buruzko orrialde ezagunak), batez ere hiztunen hurrengo belaunaldiak baliabide digital horiek erabiltzera bultzatu nahi baditugu. Halako informaziorik egon ezean, hizkuntzak arriskua du Wikipedian erreproduzitzeko gizartean pairatzen duen diglosia-egoera, eta horri buelta eman ezin izateko. Estrategia mistoa izan behar du. Metodo horri nola ekin erabakitzeke, lagungarria izango da bilatzea zein artikulua diren irakurrienak eta zein artikuluri egun zaizkien ekarpen gehien.

» **Luzera**

Bizitasun digitalaren "Garabidean" mailan dagoen hizkuntza batentzat, merezi du artikuluen luzera aintzat hartzen hasia, garrantzi handiko parametroa baita. Wikipediara sarreraren definizio hutsaz harago informazio entziklopedikoaren bila sartzen diren erabiltzaileak gure hizkuntzako wikipediara sartzera nahi badugu informazio hori bilatzera, funtsezkoa da behar hori oinarritzeko mailan aseko duen luzera ematea artikuluei, nahiz eta jakitun izan beti aurkituko dituztela datu gehiago hizkuntza handietako wikipedietan.

» **Calidad lingüística**

Hizkuntza gutxituetako wikipedia batzuetan, kezka handia sortzen du artikuluen hizkuntza-kalitateak. Oso garrantzitsua da oreka bilatzea: kalitatea kaskarra bada, baliteke hizkuntza horretako wikipediaren izena zikintzea, baina, arreta handiegia jartzen bada kalitate-maila altua eskatzen, litekeena da erabiltzaile batzuk gogogabetzea eta haiek parte ez hartzea. Adibidez, sardinierazko wikipediako artikulua idazten dituzten egileek hiru aldaeraren artean hauta dezakete (Limba Sarda Comuna, Logudoresu eta Campidanesu), eta aldaera horietarako kanpoko zuzentzaile ortografiko baterako esteka ematen da, eta iradokitzen da egiaztatzea testuaren koherentzia, hura Wikipedian argitaratu aurretik.

Erabiltzaileei Wikipedia editatzen parte harrarazteko moduak:

- » Editatoki bat antolatzea tokiko jaialdietan, eta itzulpen-proiektuak sustatzea hezkuntzan, baita alfabetatze digitala hobetzeko modu gisa ere.
- » Bigarren Hezkuntzako eskolak bultzatzea ikasleei Wikipedia editatzen irakastera (ikasgela batek orrialde bat sor dezake zure herriari edo kirol-klubari buruz, adibidez).
- » Mintegi bat antolatzea trebatzaileak trebatzeko (eskualde-hizkuntza edo hizkuntza gutxituaren bidez), eta baldintza zuzenak ezartzea. Editatzaile berriei beren eskualde-hizkuntza edo hizkuntza gutxituan ematen bazaie mintegia, erakargarriagoa egingo zaie haren ordezko hizkuntza handian ematen bazaie baino.
- » Wikipediako orrialde bisitatuenak identifikatzea (oro har), eta orrialde horiek lehentasunezko artikulua gisa nabarmentzea, editatzaileek itzul ditzaten.
- » Saiatzea tokiko agintariak esku har dezaten Wikipedia sustatzen, hura erabiltzea bultzatzen duten kanpainen bidez; laguntza ematea Wikipediari ekarpena egiten dioten komunitateei; eta edizioarekin eta zuzenketekin laguntzea.

Hartzaileak: partikularrak, kolektiboak.

Adibideak:

- » [Celtic Knot - Wikipedia Language Conference](#)
- » [Basque Wikimedians User Group](#)
- » [Collaboration with Wikipedia in 2016 and 2017](#)
- » [Catalan Wikipedia - Creation](#)

Gehiago irakurtzeko:

- » [Galipedia, the Wikipedia in Galician, is now 15 years old](#) (English version by Google Translate from the [original](#) in Spanish)
- » [The Basque Wikipedia, Local Knowledge Gone Global \(and back\)](#)
- » [Wikipedia in Catalan, leader in the 1,000 most important articles](#) (English version by Google Translate from the original in Catalan)

Horri buruzko modulua TPn: 4

6. Errendimendu digitala

Ekintza-ildo nagusia: Zure hizkuntzan gauzak digitalki egiteko aukerak sortzea

- » Eskualde-hizkuntza eta hizkuntza gutxituetako Interneteko zerbitzuen eskaria sustatzea
- » Softwarea eta erabiltzaile-interfazeak lokalizatzea
- » Itzulpen automatikoko zerbitzuak
- » Domeinu espezifiko bat lortzea

6.1 Interneteko zerbitzuen eskuragarritasuna

R10 Interneteko zerbitzuak zure hizkuntzan erabiltzeko aukerak zabaltzea

Hizkuntza batek mundu digitalean funtzionatuko badu, ezinbestekoa da sarean “gauzak egin” ahal izatea. Besteak beste, zerbitzu hauei buruz ari gara: banku elektronikoa, osasuna, erosketak, turismoa, kultura eta albisteak.

R10.1 Zure eskualde-hizkuntza edo hizkuntza gutxitua erabiltzen dutenen komunitatetik informazioa eta esperientziak biltzea, zehazteko zer zerbitzu diren garrantzitsuenak eta erabilienak

Jotzen da 4. mailan dagoen hizkuntza batek baduela nolabaiteko funtzionaltasuna Interneten. Hizkuntzak nolabaiteko aitortpen ofiziala badu, litekeena da administrazio publikoak onlineko zerbitzuak eskaintzea hizkuntza horretan. Lan-ildo bat izan daiteke arlo hori zabaltzea eta indartzea, baina, aldi berean, lan egitea sektore pribatua kontzientziatu dadin komeni dela bere bezeroei eta erabiltzaileei eskualde-hizkuntza edo hizkuntza gutxituan zuzentzea. Nolanahi ere, hiztunen beharretara bideratuta egon behar du gure jardunak. Beraz, oso garrantzitsua da hautematea zer zerbitzu erabili nahiko lituzketen hiztunek beren eskualde-hizkuntza edo hizkuntza gutxituan.

Horrenbestez, ekimen bat abiarazi baino lehen administrazioa edo enpresak bultzatzeko beren zerbitzuak zure hizkuntzan eskaintzera, informazio zehatza eduki behar da benetako egoerari buruz eta erabiltzaileen beharrei buruz, lehenetsunak ezartzeko.

Informazio hori lortzeko modu bat da inkesta bat egitea erabiltzaileen artean. DLDP proiektuan egindako galdetgia har dezakezu abiapuntu gisa, eta zure hizkuntzara itzuli.

Inkestaren emaitzak jaso dituzunean, kanpaina bat antola dezakezu hiztunek gehien nahi dituzten zerbitzuak zure hizkuntzan edukitzeko. Zerbitzu gehienak tokiko agintarien edo agintari zentralen ardurapean daude. Jarri harremanetan zerbitzu horien arduradunekin. Galdetu ea badagoen hizkuntza-arduradunen bat.

Zerbitzuen arduraduna kontzientziatzeko, nabarmendu behar zenuke zeinen garrantzitsua den tokiko hizkuntza jendearen ongizaterako eta gizarteratzerako.

Hartzaileak: kolektiboak, erakundeak.

Adibideak:

- » [The ‘Survey on Digital Fitness’ | the Digital Language Diversity Project](#)
- » Paricio Martín, S.J. & Martínez Cortés, J.P. (2014). [El uso del aragonés en Internet y las nuevas tecnologías: herramientas y repercusión](#). In *Actas II Jornadas Aragonesas de Sociología*. pp. 105-120. Zaragoza, 2014.

Gehiago irakurtzeko:

- » [Why language matters for the Millennium Development Goals - Unesco](#)

R10.2 Aztertzea zenbat balio duen negozioetan erabiltzaileen hizkuntza erabiltzeak

Globalizazioak are gehiago nabarmendu du hizkuntzak merkataritzan duen garrantzia. Gero eta jende gehiagok uste du herrialde gehienetako onlineko kontsumitzaileak erosoago sentitzen direla webguneak beren hizkuntzan baldin badaude. Halaxe gertatzen da eskualde-hizkuntza edo hizkuntza gutxituekin ere, are gehiago kontuan hartzen baditugu faktore emozionalak, zeinek oso eragin zuzena baitute marka batek kontsumitzaileei helarazi nahi dien irudian. Horregatik merezi du hizkuntzaren faktorea kontuan hartzea negozio-estrategia bat diseinatzean.

Hartzaileak: enpresak, erakundeak.

Adibideak:

- » [Let languages shout out your business benefits](#)
- » [The Benefits of Translating into Minority Languages – Translation](#)

Gehiago irakurtzeko:

- » [Language Means Business](#)
- » [The Importance of Language in Global E-Commerce | TransPerfect](#)
- » [The Value of Language in e-Commerce white paper](#)
- » Cunliffe, D., Pearson, N. & Richards, S. (2010). E-commerce and Minority languages: a Welsh perspective. *Language and the Market, Palgrave Macmillan, Basingstoke*, pp.135-147.

R10.3 Smartphoneetarako app-ak garatzea

Smartphoneetarako app-ak modu erakargarria dira hainbat xedetarako edukia helarazteko. Jakinik smartphoneak hain hedatuta daudela, app bat modu nahiko erraza izan daiteke hizkuntza bat jende gehiagok balioesteko.

Planteatu behar zenuke, adibidez, hizkuntzak ikasteko doako app-ak garatu daitezzen sustatzea. Arlo horretan, adibide polit eta erreplikagarri ugari daude, hala nola Wahzhazhe (iOSerako eta Androiderako), eta Speak Mohawk (iOSEko eta Androideko bertsioak), First Nations-en eskutik, edo rawang-hiztegiaren app-a.

Hitzak, esaldi sinpleak eta agurrak egitea modu dibertigarria da norbait hizkuntza eta kultura batean murgildu dadin, eta hizkuntza horretatik datorren kultura jakinaren berezitasuna ezagutzen hasteko modu bikaina izan daiteke. Beste funtzio batzuk izan daitezke hitzen eta esaldien grabazioak sartzeko, erabiltzaileak hizkuntza praktikatu dezan. Ikasleei beren gaitasunak ebaluatzeko aukera ematen dieten jokoak eta testek harrera ona izan ohi dute. Zenbat eta gehiago murgildu eta jardun esperientzian, hobeto.

Zoritxarrez, app-ak garatzea garestia da webguneen aldean. Baina badaude aukera batzuk *txartel mnemoteknikoen* app-ak sortzeko, plataforma hauek erabiliz, besteak beste: Memrise eta Anki. Horiei esker, hizkuntza-komunitateak askoz gutxiago inbertitu behar du. Metodo horrek eduki interesgarria sor dezake mugikorrarekin ikasteko, baina, agian, app independente baten osperik gabe. Memrise-ko lakota-ikastaroen adibidea: <https://www.memrise.com/courses/english/lakota/>

Hartzaileak: partikularrak, kolektiboak, erakundeak, software-garatazaileak.

Adibideak:

- » [How Technology is Saving Native Tribe Languages](#)
- » [Wahzhazhe, an Osage language app for phones and tablets](#)
- » [Six Nations school launches app that teaches people to speak Mohawk](#)

6.2 Sare sozial lokalizatuak

R11 Sare sozialen erabiltzaile-interfazeak lokalizatzeko ekimenak

Arestian esan dugu zeinen garrantzitsuak diren sare sozialak hizkuntza baten bizitasunean. Nahiz eta litekeena den erabiltzaile-interfazea zure hizkuntzan egotea, horrek ez du zertan nahitaezkoa izan zuk sare sozialetan hizkuntza hori erabil dezazun, baina faktore horrek nolabaiteko eragina du hizkuntzaren erabileran, eta lagungarria izan daiteke erabiltzaile gehiago hizkuntza hori erabiltzen hasteko.

Zoritxarrez, Facebookek, Twitterrek, Googlek eta antzeko enpresek azken urteotan baino aukera gehiago ematen zituzten lehen beren interfazeak lokalizatzeko (ikus, behean, Scannell, 2012).

Nabarmendu beharra dago, halaber, itzulpenaren kalitateak garrantzia duela hizkuntzaren ospean eta horrek erabiltzaileentzat duen balioan, eta litekeena da, orobat, hizkuntzaren erabilera areagotzen laguntzea. Horrez gain, helburu bat izan behar luke itzulpena amaitzea.

Hartzaileak: kolektiboak, erakundeak, instituzioak.

Adibideak:

- » [Sa traduzione de Facebook in sardu](#)

Gehiago irakurtzeko:

- » Losse, K. (2008). [Facebook: Achieving quality in a crowd-sourced translation environment](#)
- » Scannell, K. (2012). [Translating Facebook into endangered languages](#). In Proceedings of the 16th Foundation for Endangered Languages Conference (pp. 106-110).

Horri buruzko modulua TPn: 3

6.3 Software lokalizatua: sistema eragileak eta oinarrizko softwarea

Softwarea zure hizkuntzan erabili ahal izatea ospe onaren seinale da; adierazten du bizi zaitezkeela mundu digitalean zure hizkuntzan, eta horrek guztiak asko balio du, erabiltzaileei gogoia eman baitieziaieke beren eskualde-hizkuntza edo hizkuntza gutxitua hautatzeko komunikaziorako eta lanerako hizkuntza gisa.

R13 Hizkuntza-komunitatean gehien erabiltzen den xede orokorreko software librea edo jабeduna lokalizatzeko ekimenak indartzea

Ziurrenik, garabidean dagoen hizkuntza batean, lokalizatuta egongo dira, gutxienez, idazmahaiko sistema eragile bat eta sistema eragile mugikor bat (irekia edo komertziala), bai eta software orokorren bat ere (testu-tratamenduko programa bat eta nabigatzaile bat).

Garaia da oinarrizko aplikazio horietaz harago joateko, eta aukera gehiago sortzeko hizkuntza horretan lokalizatutako softwarea erabiltzeko, izan kode irekiko software librea edo software jабeduna.

Eraginkortasunik handiena lortzeko, eta ahalegina gehien erabiltzen den softwarea lokalizatuzera bideratzeko, garrantzitsua da jakitea zer premia dituzten eskualde-hizkuntza edo hizkuntza gutxituetako erabiltzaileek, eta eskualde-hizkuntza edo hizkuntza gutxitu aurreratuagoek egin duten bidetik ikastea.

Kode irekiko softwarearen kasuan, ahalegina eska diezaiokezu komunitateari hizkuntz komunitatean gehien erabiltzen den software libre orokorra lokalizatzeko. Halako ekimenen adibide batzuk dira [Softcatalà](#) (katalana), [Meddal](#) (galesa), [Librezale](#) (euskarara), [An DROUIZIG](#) (bretoiera) eta [Softaragonés](#) (aragoiera). Software komertziala lokalizatzeko (Windows, MacOS, MS Office), tokiko agintariek funtsezko papera jokatzen dute enpresekiko solaskide gisa.

Hartzaileak: kolektiboak, erakundeak, instituzioak.

Adibideak:

- » [How to Localize Software: 10 Dos and Don'ts for a Watertight](#)
- » [Software Localization Process](#)

Gehiago irakurtzeko:

- » [Localization 101: A Beginner's Guide to Software Localization](#)
- » [How to Localize your Software, App or Game : 7 Best Practices](#)
- » [How To Localize an Android Application](#)

R14 Bideo-jokoak aintzat hartzea biziberritze-aukera baliotsu gisa

Bideo-jokoek hizkuntza batekiko elkarreragin- eta, horrenbestez, inplikazio-maila altua ematen diete erabiltzaileei. Bideo-jokoak tokiko hizkuntzetan garatzea edo lokalizatzea modu bikaina da hizkuntza horiei bestelako bultzada bat emateko, eta belaunaldi berriei erakusteko hizkuntza bizirik dagoela, eta sasoiari dagoela mundu modernorako.

Pentsa daitekeenez, gutxiago erabiltzen diren hizkuntzetako bideo-jokoek merkaturatu nahiko mugatua da. Enpresa handiek ez dute inbertitzen merkatu txikietan; beraz, aktibistek eta grinatsuek hartzen dute ardura berriro ere, hala nola Gwenn Meynier-ek, zeinak Steredenn joko bretoierara itzuli baitzuen, edo Frédéric Antonpieri-k eta Fabien Mariani-k, zeinek Winterfall joko korsikeraz sortu baitzuten. Euskal Herrian, [Game Erauntsia](#) euskal bideo-jokalariekin talde bat da, zeinek bideo-jokoek munduan euskara erabiltzearen aldeko kanpaina egiten baitute.

Ejemplos:

- » [Steredenn](#)
- » [Winterfall](#)

Para leer más:

- » Get inspired by the story of Kisima Ingitchuna, the first video game in Inupiaq language that, in addition, drawn from indigenous culture for its story and characters: [Showcasing Alaska's Inupiat culture through gaming](#)
- » [PES 2016 Introduces the First Welsh Language Video Game Box Art](#)
- » [The video game: A challenging universe for minority languages](#)
- » [Conquering digital worlds in Scottish Gaelic](#)

6.4 Itzulpen automatikoko zerbitzuak

Itzulpen automatikoko zerbitzuak eskura egoteak adierazten du hizkuntza sarri eta aktiboki erabiltzen dela mundu digitalean, jotzen baita tresna eta baliabide ugari dituela. Hizkuntzaren erabilgarritasun digitalaren ikuspegitik, hizkuntza batek itzultzaile automatiko fidagarria edukitzeak adierazten du hizkuntzak presentzia eta garrantzia digital nahiko altua lortu duela.

Hizkuntza gutxitu batek mugitu egin behar duen eta komunikaziorako tresna gisa bizirauten saiatu behar duen inguruneetan, itzulpen automatikoaren erabilera jakin batzuek eragiten dituzten egoerak tentuz tratatu beharrekoak dira. Gutxieneko kalitatea duen sistema bat erabilgarria eta eraginkorra izan daiteke asimilaziorako, hau da, testu ulergarriak ekoizteko, nahiz eta ez izan egokiak argitaratzeko (dibulgazioa). Baina zaila da, edozein hizkuntzatan eta, bereziki, eskualde-hizkuntza edo hizkuntza gutxituen kasuan, dibulgaziorako behar den kalitatea lortzea, itzultzaile batek testua editatu beharrik gabe. Desberdintasun hori ez bada kontuan hartzen, gerta liteke gehiegi erabiltzea itzulpen automatikoa, eskualde-hizkuntza edo hizkuntza gutxituen beraren kaltetan, zoritxarrez. Itzulpen automatikoaren erabilera oker horrek zenbait esperientzia txar eragin ditu (adibidez, [txerokierazko Wikipediaren kasua](#)), bai eta zenbait kritika ere (ikus M.

Bauer-en artikulu hau: [When things are way, way, WAY worse than you thought they might get](#); eta M.B. Měchura-ren beste hau: [Do minority languages need machine translation?](#)). Hori dela eta, funtsezkoa da itzulpen automatikoaren erabiltzaile potentzialek muga horien berri izatea, teknologia hori ondo erabil dadin. Ikus, esaterako, [Itzulpen Automatikorako Ohar hau](#), non ideia hori nabarmentzen baita galesaren kasurako.

Bizitasun digitalaren 3. mailan dagoen hizkuntza baten kasuan, ez da harrizkoa itzulpen automatikoko zerbitzuak emateko behar den teknologia garatu gabe egotea, edo behar den heldutasun-maila lortu gabe egotea. Hala ere, komeni da etorkizunean urrats hori egiteko oinarria prestatuko duten mekanismo jakin batzuk abiarazten hasia.

Bizitasun digitalaren 4. mailan, oso litekeena da hizkuntzak onlineko zerbitzu edo tresna bat edukitzea, gutxienez hizkuntza bat duela eskualde-hizkuntza edo hizkuntza gutxituarekin parekatua, eta, gutxienez, norabide batean.

Nabarmentzekoa da beharrezkoa dela administrazio publikoak datu elebidunak koordinatzea. Gobernu gehienak ez dira jabetzen zenbateraino izan daitezkeen baliagarriak datu elebidunak itzulpen-tresnak garatzeko, eta, hala, itzulpen-beharrak asetzen laguntzeko (nolabaiteko estatus ofiziala badu hizkuntzak). Normalean, horren inguruan aholkatzeko eskatzen zaie hizkuntza-teknologoei, eta nabarmentzeko itzulpenak modu koordinatuan zentralizatzeko beharra, bermatzeko datuak berrerabili egiten direla (*itzulpen-memoriak*) eta datu elebidunak erraz eskuratu daitezkeela itzulpen automatikoko sistemak eraikitzeko. Administrazio publikora bereziki egokituta badago, lor daiteke nahiko kalitate oneko emaitza; hala, esfortzu txikiagoa egin beharko lukete itzultzaileek (eduki errepikakorra editatu, zerotik hasi beharrean).

R17 Hizkuntza-konbinazio kopurua handitzea; saiatzea ingelesa sartzen, jada sartuta ez badago

Bizitasunaren "Garabidean" mailan, oso litekeena da hizkuntzak arauetan oinarritutako itzulpen automatikoko sistema bat izatea. Kasu horretan, hizkuntza-konbinazioak dibertsifikatzeko lehenengo aukera litzateke konbinazio batean garatutako datu eta tresna linguistikoak (hiztegi morfologikoak, etiketatzaile gramatikalak) konbinazio gehiagotarako sistema bat garatzean berrerabiltzea. Horrek, bistan denez, mugatu egiten du gure sistemarekin erabil dezakegun hizkuntza-kopurua, hizkuntza horiek nahiko hurbilekoak izan behar baitute (ez oso desberdinak morfologiari eta sintaxiari dagokienez). Hala ere, garatu izan dira arauetan oinarritutako itzulpen automatikoko sistemak oso hizkuntza desberdinetarako, hala nola gaztelania-euskararako (Matxin).

Bestalde, gerta liteke hizkuntzak corpus paralelo kopuru aski handia izatea; horrek aukera emango lioke hizkuntzari itzulpen automatiko estatistikoaren munduan sartzeko. Itzulpen automatiko neuralaren arloan eskualde-hizkuntza edo hizkuntza gutxitu batzuekin egin berri diren esperimendu batzuek emaitza itxaropentsuak eman dituzte. Agian, modu horretan egin dezake salto kualitatibo bat eskualde-hizkuntza edo hizkuntza gutxituetarako itzulpen automatikoak, eta itzulpen-zerbitzuetan eguneroko tresna gisa erabiltzetik gertuago jarri.

Hartzaileak: ikerketa-taldeak, software-garatzaileak.

Adibideak:

- » [Matxin: an open-source transfer machine translation engine](#)
- » [Automatic translation - Llengua catalana - Gencat.cat](#)
- » [Oersetter: Frisian-Dutch Statistical Machine Translation](#)
- » [You can now translate Wikipedia articles from Spanish into Basque, thanks to an open source machine learning tool](#)

Gehiago irakurtzeko::

- » Popović, M., Arcan, M. & Klubička, F. (2016). [Language related issues for machine translation between closely related South Slavic languages](#). In *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)* (pp. 43-52).
- » Dowling, M., Cassidy, L., Maguire, E., Lynn, T., Srivastava, A. & Judge, J. (2015). [Tapadóir: Developing a statistical machine translation engine and associated resources for Irish](#). In

Proceedings of the The Fourth LRL Workshop: Language Technologies in support of Less-Resourced Languages, Poznan, Poland.

- » Screen, B. (2017). [Machine Translation and Welsh: Analysing free Statistical Machine Translation for the professional translation of an under-researched language pair](#). *Journal of Specialized Translation*, 28.
- » Mayor, A., Alegria, I., De Ilarraza, A.D., Labaka, G., Lersundi, M. & Sarasola, K. (2011). [Matxin, an open-source rule-based machine translation system for Basque](#). *Machine translation*, 25(1), p.53.
- » Gompel, M.V., van den Bosch, A.P.J. and Dijkstra, A. (2014). [Oersetter: Frisian-Dutch statistical machine translation](#). Ljouwert: Fryske Akademy
- » Etchegoyhen, T., Martínez, E., Azpeitia, A., Labaka, G., Alegria, I., Cortes, I., Jauregi, A., Ellakuria, I, Martin, M. & Calonge, E. (2018). [Neural Machine Translation of Basque](#). In *21st Annual Conference of the European Association for Machine Translation* (p. 139).