

# Bertsolaritzarako errima-aurkitzaile informatikoa

**Iñaki Alegria, Xabier Arregi eta Bertol Arrieta**

CAF-Elhuyar Sariak. Lehen saria

**Internet sarea, posta elektronikoa, itzultzaile automatikoak, ahotsa ezagutzeko aplikazioak, zuzentzaile ortografikoak... Nor ez da gaur egun informatikaren aurrerapenez baliatu? Edo nork ez du gutxienez, horien berri izan? Bertsolaritzaren ingurunean ere, ezin al gara aurrerapen horietaz baliatu?**

ARGI DAGO XAKEAN KASPAROV-I IRABAZI ZION "Deep Blue" hura bezalakoren bat sortzea mementuz ezinezkoa dela bertsolaritzan. Argi dago oraindik bertsolari birtualik ez dugula lortuko, ez dugula Andoni Egaña baten mailako bertso-makinarik sortuko. Baina bertsoak egiteko garaian lagunduko duen aplikazio bat, zergatik ez? Artikulu honetan aurkezten den errima-aurkitzaile honek bertsopaperetan hain iaioak ez direnentzat laguntzatxo bat edota etorkizunean bertsolari izateko ikasten ari direnentzat euskarri bat izan nahi du.



Izan ere, lehen ordenagailua sortu zenetik, duela 40 bat urte, informatikak izan duen garapena ikaragarria izan da. Gaur egun edozein esparru edo arlotan lan egiteko ia ezinbestekoa zaigu aplikazio informatiko bat edo beste. Gauzak honela, zergatik ez erabili informatika, maite ditugun beste hainbat alorretan? Eta bertsolaritza zergatik ez da bada, alor horietariko bat

izango? Argi dago Egaña edo Amuriza birtual bat lortzetik urrun gaudela, alde batera utzi baitira mementuz, informatika sortu zenetik amets izan diren makina pentsalaria edota antzekoak sortzeko helburu handizaleegiak. Hala eta guztiz ere, helburu xumeago batzuk betetzeko ez al gara ba dagoeneko gai izango? Bertsozalearentzat lagungarria izan zitekeen aplikazio oro-

kor bat egiteko ideiatik jaiotako errima-aurkitzaile informatiko honek, hori besterik ez du izan nahi: errimak aurkitzeko aplikazio lagungarri bat.

Azalpen sakonagoetan hasi aurretik ordea, koka dezagun gure lana dago-kion informatikaren alorrean, adimen artifizialean hain zuzen —hizkuntza-ingeniaritzan zehazkiago—, eta azter dezagun alor horrek izan duen bilakera, makina pentsalaria helburu zuela jaio zenetik gaur egunera arte.

## Makina pentsalaria: hura bai astakeria!

Horixe baitzen adimen artifizialaren helburua (ametsa esan beharko genuke), 50eko hamarkadan informatikaren alor hori hitz horiekin bataiatu zutenetik. Makina adimentsua sortzea; pentsa zezakeen makina.

60ko hamarkadan jasandako hainbat porroten ostean ordea, desilusio antzeko bat etorri zen 70eko hamarkadan, eta honekin batera, helburuen xumetze bat. Helburuak xumetu izanaren ondorioz, "sistema adimendunen" lehen prototipo garrantzitsuak sortu ziren; gai zehatz eta mugatueta erantzunak emateko gai ziren sistemak. Hala nola, MYCIN (Shortliffe, 1976) "sistema adimenduna", gaixotasun infekzioen tratamendua eta diagnostikoa burutzen zuena, edo PROSPECTOR (Duda eta beste zenbait, 1976) esplorazio geologikoak egiteko leku egokiak aukeratzeko zituen "sistema adimenduna".

Lehenbiziko prototipo horien arrakastak enpresa-munduaren inplikazioa ekarri zuen, eta horrekin batera, adimen artifizialaren sendotzea eta garapena etorri zen 80ko eta 90eko hamarkadetan. Horrela, adimen artifiziala informatikaren barruan ikerketa-lerro garrantzitsuenetako bat izatera iritsi da gaur egun.

Lengoaia naturalaren prozesamendua (LNP), edo hizkuntza-ingeniaritza nahi bada, adimen artifizialaren esparruan

"Gaur goizean pasa det ikararen marra andrea han zebilen tirri ta tarra gosaria prestatu ez zegoen txarra bapatean esan dit "adio senarra" musu bat suharra harretzek zun garra barreneko harra zure irrifarra horixe da inporta zaidan bakarra" (bis)



Andoni Egaña.

A. Egaña.

1. irudia.

kokatzen den alor zehatza da. Hizkuntzaren tratamendu automatikoa aztertzeko duen alor horren emaitzen artean, honako aplikazio informatikoak aurki ditzakegu; zuzentzaile ortografikoak, sintaktikoak, itzultzaile automatikoak, ahotsa ezagutzeko sistemak eta abar. Hauek lortzeko kontuan eduki behar da, ordea, hizkuntza bakoitzak bere berezitasunak dituela. Beraz, prozesua, hizkuntza guztientzat antzekoa izan arren, hizkuntza bakoitzari egokitu behar zaio.

*"errima-aurkitzaile  
honek  
bertsopaperetan  
hain iaioak ez  
direnentzat  
laguntzatxo bat  
izan nahi du"*

Baina proiektu honen mamia azaldu aurretik, kontatu dezagun zerk bultzatu gaituen lan hau burutzera; zein izan diren lan hau egiteko garaian gogoan izan ditugun motibazio eta helburuak, alegia.

## Motibazio eta helburuak: konplikazio seguruak

Esan beharrik ere ez dago zein altxoria den bertsolaritza. Herrialde gutxiak eusten diete hain bizi halako ahozko adierazpideei. Eta honelako bertsoek adierazten dute (ikus 1. irudia) bat-bateko bertsolaritzaren funtsa baita aberastasuna ere.

Sarritan ordea, ez gara ohartzen bertso bat egiteak duen zailtasunaz. Izan ere, jaiotzatik bertsolari-sena edo izatea garrantzitsua izan daitekeen bezala, are garrantzitsuagoa izan daiteke ikaste-prozesua bera. Ikusi bestela gaur egungo bertsolari guztiak nola atera diren bertso-eskola batetik edo bestetik. Doinuak, neurriak, errimak... Ez da hori guztia jakinda jaio den gizakirik. Ondo dakigu guk, zaleok, hainbestetan nahi eta ezin ibili garenok.

Pentsamendu hauek buruan genituela hasi ginen lanean. Bertsolaritza, bertso-eskolak, bertsotan egiten nola ikasi, nola erakutsi... Eta horiek guztiak gogoan, bertsolari hasi berriei, edo bertsozaleei oro har, laguntzeko aplikazio oso bat egiteko ideia atera genuen. Neurriak kontrolatzeko, doinuak ezagutarazteko, bertso zahar eta

berrien liburutegia izateko, errimak aurkitzeko... aplikazio bat egitea izan zen hasierako ideia. Nonbaitetik hasi beharra zegoen ordea. Eta nondik hasi erabakitzeko guri, zaleoi, zailena egiten zaigunetik hastea egokia izan zitekeela otu zitzaigun. Eta zer egiten zaiagu bada zailago, bertso batentzat errima egokiak aurkitzea baino?

<b>zoritzarrez</b>	("zoritzarrez" ADB)
	("zoritzar" IZE + DEK INS MG)
	("zoritzar" IZE + DEK INS NUMP MUGM)
<b>dena</b>	("izan" ADL A1 NOR3 + ERL ERLT + DEK NOM NUMS MUGM)
	("izan" ADL A1 NOR3 + ERL KONP)
	("dena ZNB + DEK NOM NUMS MUGM)
	("dena ZNB)

3. irudia. Análisi morfológicoaren bi adibide.

Helburua kokatua geneukan bederen. Errima-aurkitzaile informatikoa egin nahi genuen. Hitz-bukaera bat eman eta bukaera horrekin bat zetozen euskarazko forma guztiak emango zizkigun aplikazioa. Baina nola lortu hitz-bukaera bat emanda, hitz-bukaera hori duten euskal hitz guztiak? Eta guztiak diogunean, hiztegieta agertzen diren hitz guztiez gain, hitz eratorri eta deklinatuez ere ari gara, aditz trinko eta jokatuez ... Nola lortu guztiak?

zakur	handi
zakurra	handiago
zakurrak	handien
zakurrari	...
zakurraren	handia
zakurrarentzat	handiak
zakurrarekin	...
...	handiagoa
	handiagoak
	...

4. irudia. Sorkuntza morfológicoaren bi adibide.

### Euskararen alderantzaketa: hau da hau nahasketa!

Ezer baino lehen, ikus dezagun zertan oinarritu garen lan hau burutzeko garaian, ez baikara inolaz ere, hutsetik abiatu.

IXA taldeak ([www.ixs.si.ehu.es](http://www.ixs.si.ehu.es)) euskararen deskribapen morfológicoa digitalizatua dauka (Agirre eta beste zenbait, 91) K. Koskenniemi-ren bi mailatako morfologiaren ereduaren (Koskenniemi, 83) arabera. IXA taldearen deskribapen morfológico horretan sarrera bakoitza euskal morfema bat da. Morfemak, aldiz, multzo desberdinetan banatuak daude eta

*"IXA taldeak euskararen deskribapen morfológicoa digitalizatua dauka"*

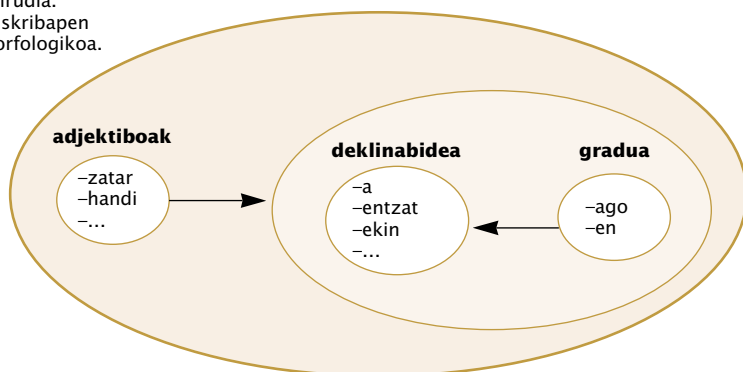
multzo bakoitzak definitua dauka jarrai diezaiokeen sarrera-multzoen zerrenda. Beste hitzetan esanda, morfema bakoitzari zein morfemak jarrai diezaiokeen gordetzen da.

2. irudiko adibidea aztertuz gero, *adjektiboak* multzoari *deklinabidea* eta *gradua* multzoak jarrai diezaioketela ikus daiteke, eta era berean, *gradua* multzoari *deklinabidea* multzoak jarrai diezaiokeela.

Euskararen deskribapen morfológico hori da euskararako analizatzaile-sortzaile morfológicoaren oinarria. Analizatzaile-sortzaile morfológico horrek, alde batetik, hitzen análisis morfológicoa egiten du (ikus 3. irudia) eta, bestalde, lexikoko sarrera bat emanda, lexikoko sarrera horretatik abiatuz sor daitezkeen hitz guztiak lortzen ditu (ikus 4. irudia).

Guk behar duguna ordea, hitz-bukaera batetik abiatuz, atzekoz aurrerako sorkuntza egingo lukeen sortzaile morfológicoa da (ikus 5. irudia).

2. irudia. Deskribapen morfológicoa.

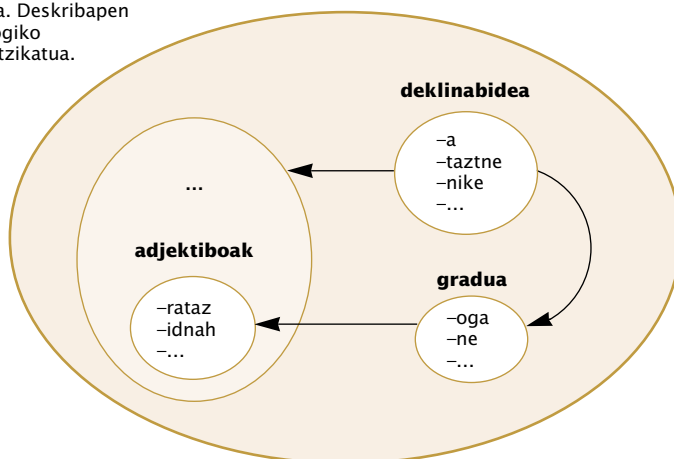


Horretarako, euskararen deskribapen morfologikoa nolabait alderantzikatu beharra dago, sarrera multzo baten atzetik joan daitezkeen sarreren multzoak definitu beharrean, aurretik joan daitezkeenak definitzeko. Beste hitzetan esanda, morfema bakoitzaren aurretik joan daitezkeen morfemen multzoak lortu nahi dira.

2. irudiko adibidearekin jarraituz, deskribapen morfologiko hori alderantzikatu ondoren, lortuko litzatekeen deskribapen morfologikoak honakoa adierazi beharko luke: *deklinabidea* multzoko sarreren aurretik, *gradua* multzoko sarrerak edo *adjektiboak* multzokoak joan daitezkeela, eta *gradua* multzoko sarreren aurretik, aldiz, *adjektiboak* multzokoak.

Hori lortzeko ordea, sortzaile morfologikoa nola dabilen kontuan izan behar da. Sortzaile morfologikoak sarrera multzoen aurreranzko kateaketa soilik egin dezake, deskribapen morfologikoak adierazten dion ordenaren arabera sortzen baititu hitzak. Beraz, ezin dezake atzeranzko kateaketarik egin. Eta sorkuntza hitz-bukaera batetik abiatzen denez, deskribapen morfologikoa alderantzikatzeke, morfema bakoitza zein morfemaren ondoren doan adierazteaz gain, morfemak banan-banan alderantzikatu beharko dira (ikus 6. irudia), eta, beraz, sorkuntza abiatzen duen hitz-bukaera ere bai. Hala bada, hitz-bukaera horrekin bat datozen hitz guztiak lortuko dira, baina atzekoz aurrera. Hori horrela, azken alderanzketa bat egin beharko da hitzak erabiltzaileari modu egokian aurkezteko. Prozesu osoa hitz-bukaeratzat 'ekin' hartuta, 7. irudian ikus daiteke.

6. irudia. Deskribapen morfologiko alderantzikatu.



Bada, hala ere, beste alderdi bat, orain arte aipatu ez izanagatik, erabili dugun adibidearekin begi-bistan geratu dena; aldaketa morfologikoa, hain zuzen. Euskararen aurretik atzerako

karaktareaz bukatzen den hitz bati bokalez hasten den forma batek darraionean, 'r' bigun hori 'r' gogor bihurtuko duen erregela bat definitu beharko da:

$$r + bokala \rightarrow rr + bokala$$

[ “euskararen deskribapen morfologikoak alderantzikatu egin da” ]

Atzekoz aurrerako sortzaile morfologikorako beharrezkoa da erregela horiek "alderantzikatzea", morfema guztiak alderantzikatuak izateaz gain, morfemen arteko loturak ere alderantzikatuak baitaude. Beraz, erregela berriak, 'r' gogorra lortzeko, ezkerreko testuinguruak bokala eta eskuinaldekoak 'r' karakterea izan dezala eskatu beharko du:

$$bokala + r \rightarrow bokala + rr$$

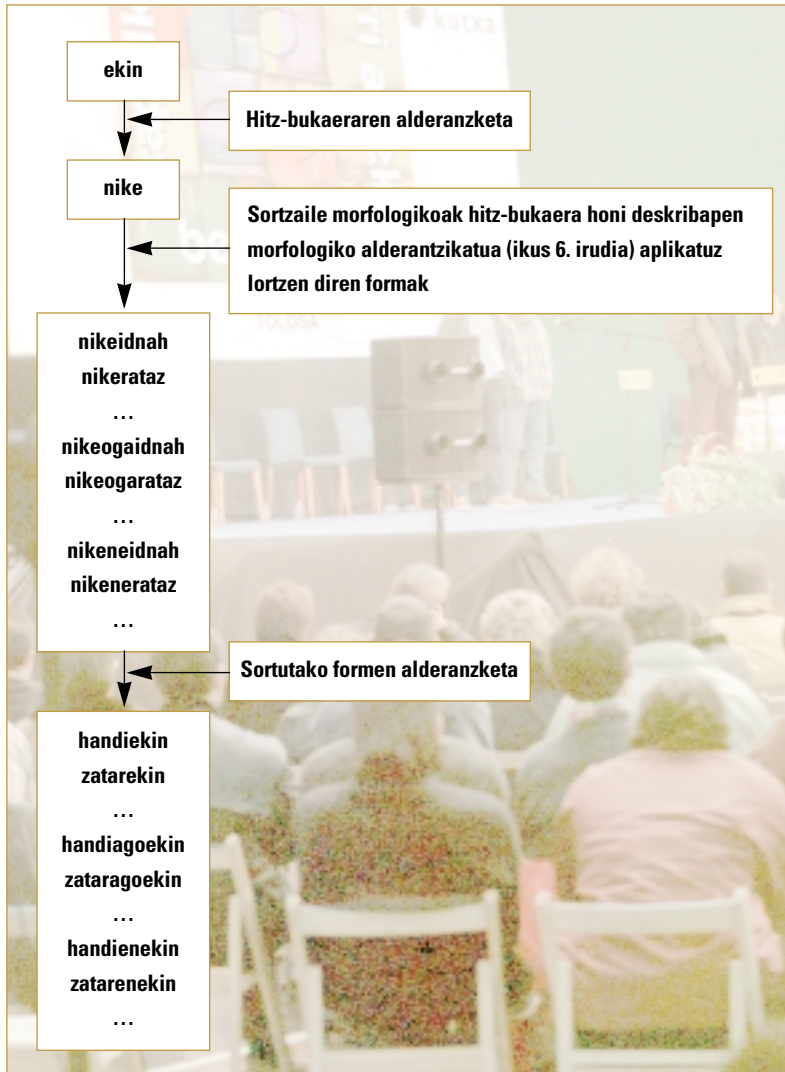
deskribapen morfologikoaren kasuan, 'zatar' hitzarekin hasitako sorkuntzan 'zatara', 'zatarren', 'zatarago' edo 'zatarrena' bezalako formak lortuko genituzke, 'zatarra', 'zatarren', 'zatarago' edo 'zatarrena' bezalakoak lortu beharrean. Aldaketa morfologikoko horiek eragiteko hain zuzen, euskararen deskribapen morfologikokoan 2 mailatako erregelak definitzen dira (hortik dator kio izena eredu honi). Horrela, 'r'

Aurreko adibide berarekin jarraituz, 'a' eta 'rataz' ('zatar' hitzaren alderantzizkoa) elkartzean 'arrataz' lortuko litzateke erregela horrekin. Gauza bera egin da gainontzeko erregela morfologikoko guztiekin.

Esan beharrik ez dago hau guztia modu automatikoan egin dela, erregela guztiak eta lexikoko sarrera guztiak banan-banan eskuz alderantzikatzea lan ikaragarria izan baitzitekeen. ➔

5. irudia. Atzekoz aurrerako sorkuntza morfologikoaren bi adibide.

ago	ena
nago	pena
hago	antena
...	...
handiago	handiena
txikiago	txikiena
...	...



7. irudia. Prozesu osoaren deskribapena.

Beraz, euskararen deskribapen morfologikoa gisa honetara alderantzikatuz lortu da euskararako atzekoz aurrerako sortzaile morfologikoa, hitz-bukaera bat emanda, bukaera horrekin bat datozen hitz guztiak ematen dituen. Aplikazioaren abantaila bat dirudien ezaugarri hau –hitz guztiak ematearena alegia–, buruhausteen iturburu ere izango da ordea, zenbait hitz-bukaerarekin sortzen den hitz-zerrenda handiegia baita.

‘Ena’ hitz-bukaera emanda, adibidez, sortzen den hitz-kopurua ikaragarria da, hitz soilez gain (‘pena’, ‘antena’...), hitz eratorriak eta deklinatuak (‘handiena’, ‘txikiena’, ‘ederrena’..., ‘etxeena’, ‘mutilena’, ‘amarena’ ...), edota aditz jokatuak (‘dituena’, ‘zituena’, ‘genitue-

na’...) ere asko direlako. Horregatik zenbait neurri hartu behar izan dira gehiegizko sorkuntza hori nolabait eteteko edota gutxienez, bertsolariari gehien interesatuko litzaizkioken hitzak lehenbizi azaltzeko.

### Euskararen aberastasunak ekarri dizkigu zailtasunak

Euskararen berezitasunak aztertzen hasita ikus daiteke euskarak badituela halako ezaugarri berezi batzuk, euskal hitzek bukaeran izan dezaketen aberastasunaren berri ematen dutenak.

Hasteko, hizkuntza eranskaria da, hau da, morfema baten ondoren beste morfema bat edo gehiago joan daitez-

ke. Bestalde, aditzaren aberastasuna ere nabaria da, aditz bakoitzak har ditzakeen forma desberdinak asko baitira. Eta gainera flexioan dituen hamalau deklinabide-kasuek hitz batek bukaeran onar ditzakeen forma desberdinen berri ematen dute.

Euskal hitzek bukaeran izan dezaketen aberastasun horren ondorioz hain zuzen, azalatu dira aplikazioak izan ditzakeen arazoak, erabiltzaileak ematen duen hitz-bukaeraren arabera, hitz-bukaera horrekin bat datozen euskal formak milaka izan baitaitezke. Zer egin bada, arazo honi irtenbide bat emateko? Bi hobekuntza nagusi planteatu dira.

*“zenbait hitz-bukaerakin sortzen den hitz-zerrenda handiegia da”*

Bata, aplikazioak itzultitako hitzak kategoriaka multzokatzea. Lehen erabilitako adibidearekin jarraituz, ‘handiena’, ‘txikiena’, ‘ederrena’ eta hauek guztiak azaldu beharrean, adibide eta azalpen bat emango litzateke gisa honetara:

**handiena**  
handi(ADJ) + en(GRAD) + a(NOMIN)

edota,

**handiena**  
handi(ADJ) + en(GENPL) + a(NOMIN)

Argi geratuko litzateke horretara, adjektibo guztiei ‘en+a’ erantsita ‘ena’rekin errima egingo luketen hitzak lortuko liriatekeela, bai graduaren kategoria

aplikatuz gero, baita pluraleko geniti-boa aplikatuta ere.

Errima berdinaren beste kategorizazio bat ondokoa litzateke:

**duena**  
**du(izan-ADI) + en(ERL) + a(NOMIN)**

Kategorizazio hauetatik at dauden hitzak –hitz trinkoak– ordea, guztiak erakutsi beharko lirateke, multzokatze-ko irizpiderik ez baita horientzat: ‘pena’, ‘antena’, ...

Beste hobekuntza ere ildo beretik dator. Bertsolariak errima batzuk edo besteak berdin estimatzen ez dituenek, forma batzuk besteak baino lehenago pantailaratzean datza pentsatutako bigarren hobekuntza. Horregatik bertsolarietan gehien estimatzen diren hitzak edo formak, gutxiago estimatzen direnen aurretik jartzea erabaki da. Eta gure ustez, estilo aldetik gutxienez, gehiago estimatzen dira hitz trinkoak (‘pena’, ‘antena...’) hitz eratorriak baino (‘handiena’, ‘txikiena’...), nahiz eta oro har, ondo ikusia egon hitz eratorriak erabiltzea, bertso berean kategoria bereko errima bat baino gehiago erabiltzen ez bada behintzat (Amuriza, 81). Dena den, hitzak aurkezteko garaian, lehenbizi hitz trinkoak eta ondoren hitz eratorriak ematea erabaki dugu.

## Hobekuntza bat edo bi, ez dira gehiegi

Aplikazio honek onartuko lituzke, nola ez, hobekuntzak, baita lan osagarriak ere, lan honekin hastean geneuzkan motibazio eta ideien ildotik.

Lehenbiziko hobekuntza, aplikazioak errimak gaika erakusteko ahalmenean datza. Beste hitzetan esanda, erabiltzaileari kategoria semantiko bateko hitz erimatuak bilatzeko aukera emango litzaioke, hau da, erabiltzaileak aukeratutako hitz-bukaerarekin eta gaiarekin bat datozen hitz guztiak erakutsiko lituzke aplikazioak.


Bigarren hobekuntza errima asonanteak edota kontsonante-familien legeak baimentzen dituzten hitzak ere aurkeztea izango litzateke. Bukaera ‘-bi’ balitz adibidez, bukaera honekin errima egiten duten hitz guztiak eskatzean, ‘bi’, ‘di’, ‘gi’, ‘ri’ bukaerak dituzten hitzak aurkezteko ahalmena izatea hain zuzen, erabiltzaileak hala eskatuz gero.

*“Internet sarean argitaratuko da errima-aurkitzaile informatiko hau”*

Hirugarren hobekuntza aplikazioarentzako web interfaze bat egitea eta Interneten argitaratzea izango litzateke, horrela aplikazioa edozeinentzat eskuragarri jarritz. Honekin batera, ez legoke gaizki, aplikazioaren azkartasuna –erantzun-denbora– hobetzea.

Eta azkenik, hobekuntza gisa baino gehiago helburu orokor gisa planteatu beharrekoa ondokoa litzateke: bertso-tan ikasten ari denarentzat euskarri gisako bat osatzea, errimez gain, neurri, doinu eta abarretarako ere, laguntza eskainiko lukeen bertso-irakasle birtual bat edo.

Ondorio gisa, beraz, esan genezake aplikazio hau, beste aplikazio handi-ago baten barruan kokatu beharreko lana litzatekeela. Hala eta guztiz ere, atal hau bere kabuz badabil, eta zenbait hobekuntza egin bezain laster Internet sarean argitaratuko da errima-aurkitzaile informatiko hau, edonork erabil dezan.

Etorkizun hurbilean, beraz, errima-aurkitzaile informatiko honek, edozein bertsozalerentzat euskarri gisakoa izan nahi du, bertso-paperen bat egiteko garaian errimarik gabe inor gera ez dadin behintzat. Edo balio dezala gutxienez, honelako artikuluei izenburu txoro errimatuak aurkitzeko. 



8. irudia.  
Xalbador.

EGUNKARIA

## BIBLIOGRAFIA

AGIRRE, A. ETA BESTE  
ZENBAIT  
B