

AHOZKOA IKERTZEN: BIZKAIERAREN FONOTEKA

Iñaki Gaminde *, Juan M^a Sanchez, Inma Hernáez, Imanol Madariaga, Isaac Amezaga

Elektronika eta Telekomunikazioak Saila, ESI Bilbao
*Hizkuntza eta Literaturaren Didaktika Saila, IUE Bilbao
Euskal Herriko Unibertsitatea
{igaminde, ion, inma, imanol, isaac}@bips.bi.ehu.es

Helburuak

Artikulu honetan Bizkaieraren Fonoteka deritzon bizkaiko ahozko euskararen bariedade dialektalen soinu artxiboa aurkezten da. Datu-base honetan soinu artxiboak gordetzen dira, dagokien informazioarekin batera, eta interneten bidez erabil daiteke. Sistema erabiltzeko helbidea ondorengo hau da: <http://bizkaifon.ehu.es/>.

1. FONOTEKAREN AINTZINKARIAK

Azken urteotan hizkuntz batek soinu materiala eskura izatea gero eta garrantzi handiagoa hartzen ari da bai multimedia aplikazioen garapeneren arloan eta baita hizkuntz teknologien garapenerako ere. Euskarari dagokionez, azken hamarkadetan hainbat helbururekin batutako material kopuru handia dago. Material hori aplikazio askotarako baliozko izan daiteke. Hala eta guztiz ere, dagoen materiala oso anitza denez eta formatuak bateraturik ez daudenez, material gehiena ez dago erabiltzaile potentzial askoren esku.

Beste alde batetik, euskarak, beste hizkuntza batzuekin konparatuz, bereiztasun nabarmena dauka, izan ere, dialekto edo euskalki askotan banaturik dago. Euskara batuaren ezartze eta garatze prozesuak eragin handia izan du aldaera hauen eboluzioan, eta eragin hori batzuetan kalterako gertatzen da, batez ere batutik urrunen dauden bariedadeetan; hala nola, Bizkaiko aldaera askotan. Horretaz gain, euskara batuaren oinarria aldaera zehatz bat ez denez, euskara batua ez dator bat euskara mintzatuaren inolako aldaera dialektalekin. Horrek guztiak, nolabait, hizkuntzaren ahozko aldean irakaskuntza eta ikaskuntza oztopatzen du (ahozkapena, azentua, intonazioa, etab.).

Euskararen bariedade gehien grabaketak daude, eta era ezberdin askotakoak direla kontuan harturik, balio soziolinguistiko handikotzat jo daitezkeen material hau antolatu eta konserbatzeko beharrezkoak sortzen da. Beraz, hemen aurkezten den lanaren helburu orokorrak ondorengo hauek dira:

- Grabaketen formatuak eta haiei erantsitako datuak bateratuko dituen datu-base bat diseinatzea eta garatzea.

- Datuak era selektibo baten bidez arakatzeko software bat garatzea.

Artikulu hau honako era honetara egituratzen da: jarraian datorren atalak Fonotekaren helburuak zerrendatzen ditu, eta hirugarrenak erabiltzaile potentzialen irudi bat ematen du. Laugarren atalean, Bizkaieraren Fonotekan aurkitzen diren materialen motakako sailkapena egiten da. Bosgarrenean, datu-basearen beraren diseinuaren berri ematen da. Seigarren atalera aurrera alde teknikoagoak azaltzen dira: implementazioa, datu berriak sartzeko prozedura eta erabiltzaile batek bilaketak egiteko aurkituko duen interfazearen deskribapena. Azkenik, ondorioak eta esker onak.

2. FONOTEKAREN HELBURUAK

Ondoren soinu artxiboen datu-base honen helburuak zerrendatzen dira:

- Dagoeneko grabatu den soinu materiala pilatu, babestu eta sailkatzea. Horrela, hizkuntzalaritzaren aldetik balio handia duten grabaketak galtzea eragotziko da.
- Helburu desberdinekin grabatutako soinu dokumentu berriak barneratu ditzakeen egitura irekia garatzea.
- Bizkaiko euskararen aldaera lokalak hedatzeko gune garrantzitsuak hornitzea. Horrela, euskara batuaz gain, Bizkaiko aldaera gutxituenak ere zabalduko dira.
- Bizkaiko euskarari buruzko interesa handia duten hizkuntza ikertzen dutenentzako balio handiko lanabesa telematikoa argitaratzea. Horrekin ikerketak bultzatuko dira, bai hizkuntzalaritzaren aldetik, eta baita gizarte-zientziaren aldetik ere.

- Mundu osoko ikertzaileak, ingeniariak eta hizkuntzalariak elkartzeko topagune bat eskaintzea, haiek hizkuntzaren alde fonetiko eta akustikoetan interesa badute.
- Web orria denez, Internet-ek eskaintzen dituen abantailak ere baditu: dokumentuak azkar kokatu eta kargatzea, denbora eta lekuarekiko independenteki, multimedia inguru batean.

3. ERABILTZAILE POTENTZIALAK

Fonotekaren helburuetariko bat, beharbada garrantzitsuenak, ahozko testuen ondarean gordetzea eta erabiltzaileen esku jartzea izan behar da. Helburu hau gogotan erabilita hiru eratako erabiltzaileak ikusten dira: profesionalak, ikasleak eta jende arrunta

3.1. Profesionalak

Fonotekaren erabiltzaile profesionalen artean hizkuntzalariak, etnologoak, irakasleak eta teknologiaren esparrukoak aurki daitezke.

Hizkuntzalarien artean Fonetika, Fonologia, Morfosintaxia, Lexikoa, Testu-Gramatika eta Etnolinguistika lantzen dituztenak dira erabiltzaile potentzialak. Era berean Dialektologiaz, Soziolinguistikaz eta aldaketez arduratzen diren ikertzaileak. Fonotekaren eduki batzuk kulturari dagozkionez gero, material interesgarria izan daiteke etnologo eta antropologoentzat ere bai.

Profesionalen artean, irakasleak dira erabiltzaile behinenak. Era berean, proiektu hau hezkuntza alorrean eskolatik unibertsitateeraino erabili ahal izango da.

Bai eskolan hizkuntza erregistroak lantzeko materialak edukitzeko, bai ipuinak eta kantak entzun ahal izateko, zein kultura hedatzeko, tresna lagungarria izan daiteke. Hizkuntza bera ere lantzeko erabilgarria da, adibidez Institutu-unibertsitateetan hizkuntza lantzeko material garrantzitsua ematen da; testuak, euskalkiak, ortoepia, e.a.

Teknologiaren garapenean lan egiten duten profesionalak materiala edukiko dute euren aplikazioetan erabili ahal izateko.

3.2. Ikasleak

Fonoteka erabiliko duten ikasleen artean mota nagusi bi bereiz daitezke; ikasle arruntak eta hizkuntza-ikasleak.

- Ikasle arruntek fonoteka honako ekintzetarako erabili ahal izango dute.
 - Entzunaldiak: entzun praktika egiteko.
 - Erkaketak: barietate desberdinak ezagutzeko.
 - Ariketak egiteko aukera profesionalak lantzen dituzten gaiak lantzeko.
 - Ingurunearen ezagutzarako materiala kontsultatzeko.
 - Herri kultura lantzeko materiala kontsultatzeko.
 - Ahoskera ariketak egiteko.
- Hizkuntz-ikasleen erabileretan honako hauek ikus daitezke:
 - Entzun praktikak egiteko.
 - Ahoskera lantzeko.
 - Hizkuntzaren gainerako atalak lantzeko.

3.3. Jende arruntaren erabilerak

Fonoteka edozeinek erabiltzeko moduko tresna da, bai norberaren kuriositatea asetzeko edo zerbait ikasi gura duelako. Jende askok bere herrian zer nolako euskara egiten den entzuteko edo norberarena besteenarekin erkatzeko erabil dezake.

4. MATERIAL MOTAK

Fonotekan edozein grabaketa gorde badaiteke ere, proiektu honen helburua, honi dagokionez, bizkaieraren aldaera dialektalak pilatzea da; antolatzea, sailkatzea eta etiketatzea garrantzitsuenak izanik. Zehazki, gutxienez (aztertu gabeko material asko dagoen arren) 100 grabaketa ordu daude eskura, 85 herritan grabatutakoak, Bizkaiko 8 guneko geografikoak ordezkaturak daude. Dagoen material hori abiapuntu moduan hartuta, ondorengo lau material motak bereizi dira:

- *Berba isolatuak*: Banan-banan ahoskatutako berbak, aldaera dialektalaren hizkuntz alderen bat aztertzeko grabatu zirenak: lexikoiak, azentuari buruzko ikerketak, etab. Orokorrean, material hori dagoeneko digitalizatuta zegoen eta askotan ere, transkribatuta (soinuak eta transkribapenen arteko loturarik ez bazegoen ere).
- *Esaldiak*: Banan-banan ahoskatutako esaldiak, laborategian edo herrietan grabatutakoak, eta

askotan irakurritakoak izan direnak. Berben kasuan gertatzen den moduan, askotan transkribapena eskuragai badago ere, soinua eta testuaren arteko denbora loturarik ez dago.

- *Testuak*: bere edukien arabera, mota honi azpi-sailkapen bat ezarri zaio:
 - Testu etnografikoak: lekuko bizitzari buruzko informazioa eskaintzen dute.
 - Testu linguistikoak: hizkuntzaren alderdiren bat aztertzeko grabatutako testuak: sintaxia, intonazioa, etab.
 - Narrazioak: Herriko kontaerak sailkatzen dira horrela, pasadizoak, istorio barregarriak.. egiazko bezala kontatutako istorioak.
 - Ipuinak: gertatu ez diren (edo gertatutzat hartzen ez diren) istorioak.

Material hau normalean sailkatu, markatu eta transkribatu gabe aurkitu da, fonotekan barneratzeko ekintza guzti hauek burutu behar dira.

- *Herri Literatura*: Herriaren jakituria, kantak edo bestelako herri kulturaren adierazpenak zabaltzen dituzten artxiboak. Hauetan ere azpi-sailkapen bat egiten da:
 - Bertsoak: gertaera zehatz baten inguruan, bat-batean botatako konposizio poetikoa, ondoren herri jendearen oroimenean geratu dena.
 - Errezitatuak: ahoskatutako edo irakurritako testu ezagunen bat.
 - Esaldi eginak: euskal folklorea islatzen duten esaldi laburrak.
 - Igarkizunak: testu baten ezkutuko esangura igartzeko jokoak.
 - Kantak: herriak mu sikarekin abesten duena.
 - Olgetak: umeekin jolasteko erabiltzen diren formak.
 - Otoitzak: elizan erabiltzen diren esanak, eta bestelako errezoak.

5. DATU-BASEAREN EGITURA

Datu-basea ondoren adierazten den moduan oinarritzen da: soinu artxiboak, direktorio desberdinetan kokatuak, eta, soinu artxibo bakoitzerako, TEI Consurtiumaren gomendioaren arabera datuak gordetzen dituen SGML testu-artxibo bat (Text Encoding Initiative, 1994) (Sanchez et al., 2001). Horrez gain, SGML testu-artxiboko datu berak artxibo bitar batean gordetzen dira, honek datuak azkarrago ateratzea bermatzen baitu.

5.1. Soinu artxiboen formatua

Grabaketak Windows PCM (*wav*) formatuko audio artxiboetan gordetzen dira. Soinu artxibo hauek guztiak 16 biteko laginak erabiltzen dituzte, eta sistemak laginketa maiztasun anitzak onartzen ditu. Grabaketa guztiak kanal bakarrekoak dira (*mono*).

5.2. Testu artxiboen formatua

Testu artxibo hauek ondorengo datuak gordetzen dituzte, TEI-ren P4 gomendioaren arabera (Text Encoding Initiative, 1994):

- Soinu artxiboaren deskribapen tekniko bat, ondorengo datuez osaturik:
 - Audio artxiboaren izena edo *path*-a.
 - SGML artxiboaren eguneratze data.
 - Bildumaren izenburua (erabilitako datuak aurretik pilatutako bilduma batetik hartu badira).
 - Bildumaren identifikazio kodea (bildumaren izenaren arabera).
 - Fitxategiaren izenburua (material mota *Herri Literatura* bada).
 - Etiketatze prozeduraren arduradunaren izena.
 - Argitaratzailearen izena.
 - Copyright-ari buruzko informazioa.
 - Grabazio prozeduraren arduradunaren izena.
 - Grabaketa gauzatzeko erabili den ekipamendua.
 - Grabaketaren data.
 - Material mota.
 - Azpimota (material mota *Testuak* edo *Herri Literatura* bada).

- Hiztun desberdinen identifikazio kodeak.
- Hiztun desberdinen generoa (gizonezkoa, emakumezkoa edo ezezaguna).
- Hiztun desberdinen izenak (ezagunak badira).
- Grabaketaren lekua.
- Grabaketaren eskualdea.
- Gaia irudikatze hitz gako batzuk (bi gutxienez, bost gehienez).

- Markaketa maila bakoitzerako denbora markak. Markaketa maila anitzak onartzen dira: testu maila, esaldi maila, berba maila, difonema maila... Hala ere, sistemak maila berriak erregistratzea onartzen du, horrela SGML artxiboetan maila berriei dagozkien markak sartzeko.
- Soinu artxiboaren transkribapena edo transkribapenak. Batuzko transkribapen ortografiko bat derrigorrezkoa da, *Herri Literatura* izeneko material motako artxiboetarako izan ezik (hauek batuzko izenburu baten bidez sailkatuko baitira). Horrez gain, beste transkribapen batzuk ere onartzen dira: bizkaiera fonetikoak, bizkaiera literarioak... Onartuak dauden transkribapenez gain, sistemak transkribapen mota berriak erregistratzea onartzen ditu, SGML artxiboetan sartu ahal izateko.

Lehenengo irudian SGML artxiboaren sintaxiaren adibide bat ikus daiteke.

```

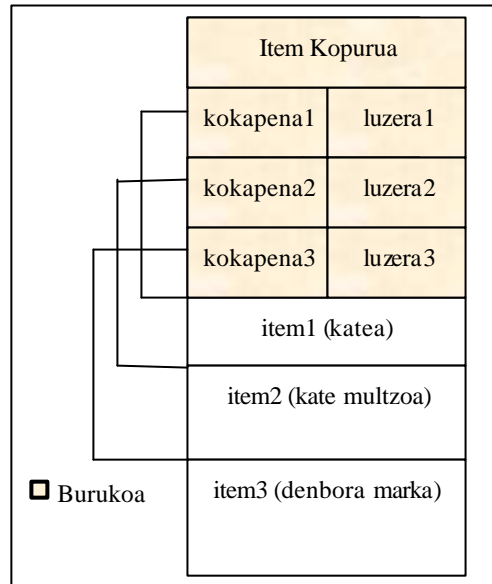
<!DOCTYPE TEI.2 PUBLIC "-//TEI Consortium//DTD TEI P4//EN" [
<!ENTITY % TEI.corpus 'INCLUDE'>
<!ENTITY % TEI.spoken 'INCLUDE'>
<!ENTITY % TEI.linking 'INCLUDE'>
<!ENTITY % TEI.figures 'INCLUDE'>
<!ENTITY % ISOLat1 PUBLIC "ISO 8879:1986//ENTITIES Added Latin 1//EN">
%ISOLat1;
<!NOTATION wav PUBLIC
'-//Microsoft, Inc.//NOTATION WAV: Windows PCM Audio File//EN' >
<!ENTITY audiofile system
"larras08.wav" NDATA wav> ]>
<tei.2>
<teiHeader date.updated="08/02/01">
<filedesc teiform="fileDesc">
<titlestmt id="IGA" teiform="titleStmt">

```

Lehenengo irudia: SGML artxiboaren sintaxiaren itxura.

5.3. Artxibo bitarrak

SGML formatuko testu-artxiboaren erabilera oso hedatua bada ere, zenbait erabilpenetarako ez dira egokiak, bertatik datuak ateratzeko interpretatu egin behar baitira, eta eragiketa horrek prozesu koste handia behar du. Bilaketak eraginkorrak izan daitezten, SGML artxiboko datu berberak artxibo bitarretan kopiatzen dira, azkar aztertu ahal izateko. Artxibo hauen egitura bigarren irudian erakusten da.



Bigarren irudia: Artxibo bitarren egitura.

Irudian ikusten den moduan, artxibo bitarrak buruko bat dauka. Bertan elementu kopurua agertzen da eta, elementu bakoitzerako, kokapena eta luzapena. Burukoaren ondoren elementuak heurak agertzen dira. Hiru elementu mota definitu dira:

- Karaktere kateak: SGML eremu bati dagokion testua gordetzen denean. Adibidez, grabaketaren lekua.
- Kate multzoak: SGML elementu baterako testu desberdin batzuk gorde behar diren kasurako. Hitz gakoak, adibidez.
- Denbora markak: denbora marka bakoitzerako, hasieraren unea, iraupena, eta transkribapen bat izan daitekeen etiketa bat gordetzen dira.

Artxibo bitar huetan datu bat kokatze a askoz azkarragoa da SGML artxiboak interpretatzea baino. Testu artxiboetatik artxibo bitarrak sortzea, *off-line* egiten da, era automatikoan.

6. SISTEMAREN IMPLEMENTAZIOA

Bilaketa-motorea eraginkor gauzatzeko, bai datu-basea eta baita bilaketa eta kudeaketa softwareak ere Linux makina batean kokatzen dira, bilaketa interfazea *Apache* web zerbitzariaren bidez eskura dago.

Ondorengo ataletan inplementazioaren xehetasunak ematen dira: datu berriak sortzeko (eta sartzeko) prozedura azaltzen da, eta ondoren, web interfazearen deskribapena eta erabilpena zederritzen dira.

7. DATU BERRIAK SARTZEKO

Bizkaieraren Fonoteka datu berriaz hornitzeko ondorengo ekintzak gauzatu behar dira:

- Grabaketa
- Digitalazioa
- Aukeraketa
- Transkribapena
- Segmentazioa
- Etiketazioa

Beraz, grabaketa berri bat gauzatu (eta digitalizatu) denean, dagokion informazio guztia sortu behar da. Datu horiek guztiak bi urratsetan sortzen dira: lehenengoan, grabaketa hautatu, transkribatu eta segmentatu egiten da (Llisteri, 1997). Horretarako, *AhoTools* deritzon seinale edizio software berezia erabiltzen da. Transkribatzerakoan, kontuan hartu behar da derrigorrezkoa dela batuzko transkribapen bat gauzatzea (Herri Literatura material motan izan ezik).

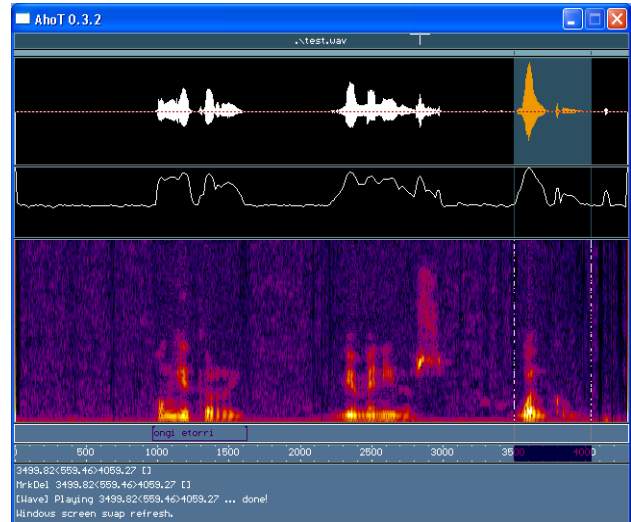
Transkribapenak burutu direnean hurrengo urratsa SGML formatuko testu-artxiboa sortzea da. Horretarako ere software berezia erabiliko da: behar dituen datuak eskatzen duen programa bat, eta haiekin eta lehenengo urratsean garatutako markekin SGML kodea sortzen duena.

Datuak sortzean, azkeneko ekintza artxibo bitarra sortzea da, SGML artxiboko datuak erabilita.

7.1. Segmentazioa

Lanaren atal hau gauzatzeko laborategian garatutako seinale analisirako softwarea erabiltzen da, *AhoTools* izenekoa (Etxebarria et. al., 2000). Hirugarren irudian ikusten den bezala, programa honen bidez seinalearen zenbait ezaugarri ikus daitezke: soinu uhina, energia, espektrograma, etab.

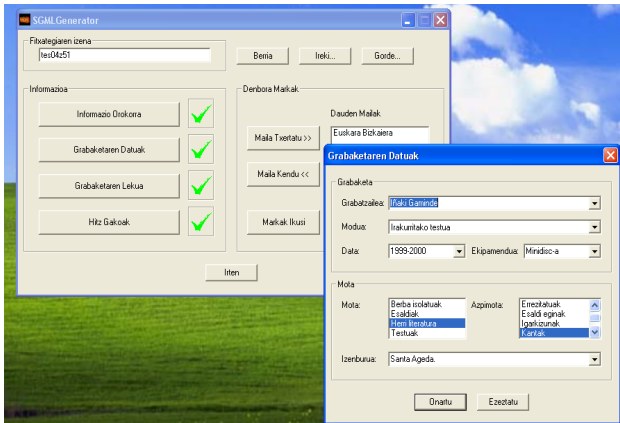
Elementu hauek guztiak balio handikoak dira segmentazioa eta transkribapenak gauzatzeko. Bai bata eta bai bestea gauzatu denean, hurrengo urratsean erabiliko diren marka artxibo batzuk gordetzen dira, SGML artxiboan baneraturako direnak.



Hiru garren irudia: *AhoTools* programa.

7.2. SGML Etiketazioa

TEI Consortium-aren gomendioarekin bat datorren SGML kodea erabiltzen duten artxiboak era azkar eta errazean garatzeko, software berezia garatu da. Windows inguru adiskidetsuan, etiketatzaileari beharrezko datuak galdetzen zaizkio. Posible denean, sartu daitezkeen datuetan programak balio zuzenak baino ez ditu onartuko (adibidez, ezin da Bizkaian ez dagoen herri edo eskualdearen izenik sartu). Datu guztiak sartu direnean, aurreko urratsean garatutako marka artxiboak ere sartzten dira, SGML artxibo zuzen oso bat gauzatzeko.



Laugarren irudia: SGML artxiboak sortzeko erabiltzen den programa, Windows inguruan.

Laugarren irudian SGML artxiboak sortzeko erabiltzen den programaren itxura ikus daiteke.

7.3. Lanabesa osagarriak

SGML testu-artxiboa sortu denean, datu-basearekiko bateragarritasunari eusteko, beste ekintza batzuk burutu behar dira. Garrantzitsuen artxibo bitarra sortzea da. Baina beste batzuk ere badaude: edukien indizea sortu behar da, adibidez. Ekintza horiek guztiak, eta sistemaren administraziorako behar diren beste batzuk ere, gauzatzeko Linux inguruan lanabes batzuk garatu dira. Ondorengo ekintzak lanabes hauen bitartez burutzen dira:

- SGML testu-artxibo batek dituen datuekin, dagokion artxibo bitarra sortzea.
- SGML testu-artxibo batean aldaketaren bat gertatzen bada (datuaren bat zuzendu behar delako adibidez), artxibo bitarreko datuak eguneratzea.
- Datu-basean dagoen grabaketa kopurua jakitea (material motaren eta eskualdearen arabera).
- Ondorengo atalean deskribatuko den web interfazearen itxura aldatzea.
- Grabaketa bakoitzerako sortu behar diren artxibo guztiak (SGML-a eta bitarra) agertzen direla egiaztatzea.
- TEI Consortium-en P4 gomendioaren arabera SGML testu-artxiboetako kodea zuzena dela egiaztatzea.

8. BILAKETAK EGITEKO

Web interfazeak hiru funtzio bete behar ditu:

- Bilaketa formularioak sortu, bidali, eta, betarik bueltatzen direnean, interpretatu.
- Bilaketarako ezarritako baldintzak betetzen dituen datu-baseko emaitzen zerrenda bat sortu, eta erabiltzaileari bidali.
- Grabaketa bakoitzerako, dagozkion datu guztiak datu-basetik berreskuratu eta erabiltzaileari bidali, datuetariko bat soinu bera da.

8.1. Bilaketa Formularioa

Bilaketa formulario desberdin bi garatu dira, bata arrunta, erabiltzaile berriari bilaketa errazteko, eta bestea aurreratua, erabiltzaile trebatuak bilaketa konplexuagoak egin ahal izateko.

8.1.1. Bilaketa formulario estandarra

Bilaketa formulario honetan, erabiltzaileak bilaketak gauzatzeko laguntza dauka: behin material mota bat aukeratu gero (bilaketa estandarrerako material mota aukeratzeko derrigorrezkoa da eta), formularioak berak erabaki batzuk hartzen ditu, bilaketa errazteko:

- Material mota *Berba Isolatua* edo *Esaldia* bada, batuzko transkribapenetan testu bilaketa bat gauzatzeko proposatuko da.
- *Testuak* mota duten grabaketarako, batuzko transkribapenetan testu bilaketaz gain, azpimotaren arabera eta gaiaren arabera (hitz gakoaren bidez) bilaketak proposatuko dira.
- Material mota gisa *Herri Literatura* hautatu bada, testuaren arabera bilaketa ez da batuzko transkribapenetan gauzatzeko, baizik eta "Izenburua" SGML eremuan. Azpimotaren arabera bilaketa eta gaiaren arabera ere proposatzen dira.

Bosgarren irudia: *Bilaketa estandarretako formularioa.*

Baldintza hauez guztiez gain, bilaketa geografiaren arabera ere murriz daiteke. Bosgarren irudian ikusten denez, eskualdea eta herria hautatzeko laukiak ere badaude.

Formulario estandar honen bidez egindako bilaketa batean, bueltatzen diren emaitza guztiak ezarritako bilaketa baldintza guztiak bete behar dituzte.

8.1.2. Bilaketa formulario aurreratua

Beste alde batetik, bilaketa konplexuagoak onartzen dituen formularioa ere garatu da. Formulario honetan, SGML artxiboan dagoen edozein datu erabil daiteke bilaketa parametro gisa, (grabaketa ekipamendua edo hizlarien izena, adibidez), eta konbinaketa booleanoak egitea onartzen du, (AND, OR eta NOT operadoreak erabiliz). Adibide moduan, seigarren irudian formulario horrekin eskatutako bilaketa konplexua ikus daiteke.

Seigarren irudia: *Bilaketa formulario aurreratua erabiliz konposatutako bilaketa.*

8.2. Emaitzen zerrenda

Bilaketa formulario bat osotu eta bidali denean, zerbitzarian prozesatu egiten da. Horren ondorioz, bezeroaren nabigatzaileak dinamikoki sortutako web orri bat jasotzen du, bertan taula bat dagoelarik. Taula horretan zerrenda bat agertuko da baldintzak betetzen dituen artxibo bakoitzerako. Zerrenda horien formatua sistemaren konfigurazio artxiboaren bidez alda daiteke, baina normalean ondoren azaltzen diren datuak edukiko ditu:

- Emaitzaren zenbakia.
- Batuzko transkribapen ortografiko bat (edo, material mota *Herri Literatura* bada, batuzko izenburu bat).
- Material mota.
- Grabaketaren lekua (herria).
- Bi lotura:
 - Lehenengoa, soinua zuzenean entzuteko.
 - Bigarrena, hurrengo atalean deskribatutako datu guztien leihoa ikustarazteko.

Zazpigarren irudian emaitzaren zerrendaren itxura ikusten da.

Bizkaieraren Fonotekea

Bilaketaren Emoitzak

Bilaketa zerbitziariak 387 emoitza aurkitu ditu.

1 (etik 20 (e)ra erakusten.

1 2 3 4 5 6 7 8 9 10 > >>

Testua	Entzuteko	Material Moeta	Herria	Gehiago
1 Aberatsa den moduan dabil hori.		Esaldiak	Beriz	
2 Afaltzen gaudela argia joan da.		Esaldiak	Zaldibar	
3 Afaltzen gaudela argia joan zen.		Esaldiak	Otxandio	
4 Afarria egin dut eta etxea ere garbitu dut.		Esaldiak	Abadiño	
5 Ai bai! hiri ere kantatu dut nik, eutsi hagin zaharra eta ekarzu barria, baia zelai da hasikera ba? Marra goma gorria eutsi hagin zaharra eta ekarzu barria. - Eta zer egiten zenuten? - Eta teiltaturo botatzen genuen hagine.		Testuak	Zornotza	
6 Aitarekin egon garen taberna itxi egin dute.		Esaldiak	Otxandio	
7 Alde batetik heldur naiz baina bestetik joatea nahi nuke.		Esaldiak	Ermua	
8 Altuan San Bernabe.		Herri literatura	Otxandio	
9 Amaltzen duzuen arte joango gara etxera.		Esaldiak	Otxandio	

Zazpigarren irudia: *Bilaketaren emaitzen zerrenda.*

Batzuetan, datu-basean ez dago baldintzekin bat datorren artxiborik. Orduan ezin da emaitza zerrendarik egon, eta zortzigarren irudian azaltzen den orrialdea agertuko da, bilaketa formularioari erantzun moduan.



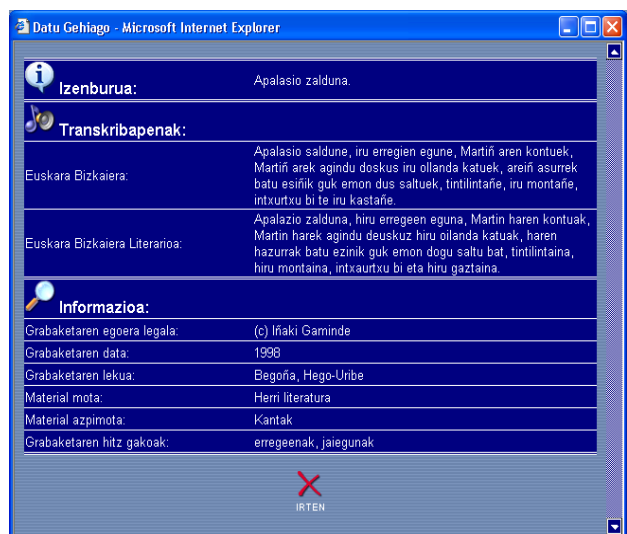
Zortzigarren irudia: *Emaitzarik ez dagoenerako orrialdea.*

8.3. Datu osoak

Emaitzen zerrendako lotura egokian klik eginez, (zazpigarren irudian itxura duena hain zuzen), grabaketari buruzko datu guztiak dituen orri bat agertuko da. Orri horretan agertzen diren datuak sistemaren konfigurazio-artxiboan bidez erabaki daitezke, baina normalean ondoren zerrendatzen direnak izaten dira:

- Grabaketaren egoera legala (copyright-ari buruzko informazioa).
- Grabaketaren data.
- Grabaketaren lekua (herria eta eskualdea).
- Material mota.
- Material mota *Herri Literatura* edo *Testuak* bada, azpimota.
- Material mota *Herri Literatura* edo *Testuak* bada, gaia irudikatzen duten hitz gakoak.
- Material mota *Herri Literatura* bada, izenburua.
- Grabaketari dagozkion transkribapenak.
- Hiztunaren araberrako segmentazioa gauzatu bazen, haien izenak (edo identifikazio kodeak, izena ezezaguna den kasuetan) eta lokutore transkribapenak.

Orri honen itxura bederatzigarren irudian agertzen da.



Bederatzigarren irudia: *Datu guztiak erakusten dituen orria.*

Transkribapenak agertzen direnean, testuan bertan klik eginez, testuari dagokion soinu zatia entzun daiteke. Sistemak soinu artxibo bat bidaliko dio erabiltzaileari, MP3 formatuan kodeturik. Formatu hau aukeratzearen zergatia eraginkortasuna da. MP3 kodetze formatuak artxiboak datu-basean bertan gordetzeko erabiltzen den Windows PCM formatukoak baino hamar aldiz neurri txikiagoak dira; beraz, bezeroari bidaltzerakoan Interneten bidezko transmisio denbora ere hamar aldiz murrizten da. Erabaki honek erabiltzaileak MP3 formatuarekin bateragarria den

soinu sistema eduki behar duela inposatzen du, baina hauek oso hedatuak direnez gero, (Windows inguruan, adibidez, *Winamp* dohako programa baino ez da behar, Internetetik <http://www.winamp.com> helbidetik deskargatu ahal dena), ez da arazorik sortzen.

9. ONDORIOAK

Garatutako egituraz sistema irekia sortzen da. Irekia, bertan dokumentu berriak barnatzea erreza delako eta, batez ere, datu mota berriak konfiguratu ahal direlako.

Grabaketei buruz gordetzen diren datu guztiak TEI Consortium-aren gomendioekin bateragarriak diren SGML artxiboetan antolatzen dira.

Erabilera Interneteko orrien bidez egiten. Horrela, web nabigatzaile bat eta MP3 kodetze formatuarekin bateragarria den soinu sistema duen edozein erabiltzailek, datu-basea araka dezake, desiratutako grabaketak entzunez.

Bizkaieraren Fonoteka sistema ren inplementazioa arrakastatsua izan da. Gaur egun datu-baseak 17037 grabaketa gordetzen ditu, ondoren adierazten den moduan antolatuturik:

- 292 herri literatura grabaketa.
- 13394 berba isolatu.
- 809 testu.
- 2542 esaldi.

Grabaketak Bizkaiko herri gehienetan gauzatu dira: 8 eskualdetako 79 herritan hain zuzen.

Garatutako sistema badabil, eta ondorengo web helbidean bisita eta erabil daiteke:

<http://bizkaifon.ehu.es/>

10. ESKER ONAK

Proiektu honek Bizkaiko Foru Aldundiaren dirulaguntzaz finantzatu da.

11. AIPAMENAK

[1] Etxebarria, B. and others. (2000). Tools and Basque language databases developed in the AhoLab

Laboratory. In Workshop Proc. LREC May 2000. (pp. 62-70).

- [2] Listerri, J. (1997). Transcripción, etiquetado y codificación de corpus orales.
<http://liceu.uab.es/~joaquim/publicacions/FDS97.html>
- [3] Sanchez, J. and others (2001). Base de datos oral y textual para el euskera. In Symposium Proc. URSI September 2001 (p. 551).
- [4] Text Encoding Initiative (1994). TEI Guidelines for Electronic Text Encoding and Interchange. Electronic Text Centre at the University of Virginia.