



EUSKALTZAINDIA

REAL ACADEMIA DE LA LENGUA VASCA
ACADÉMIE DE LA LANGUE BASQUE

**AHOZKO TRADIZIOZKO HIZKEREN
CORPUSA
(ATHC)**

Burlatan, 2012ko ekainaren 7an

0. Laburpena

Ahozko Tradizioko Hizkeren Corpusak (ATHC), izenak berak adierazten duen bezala, ahozko euskararen corpus antolatu eta hautatu bat eraikitzea du helburu. Helburu honek badu izateko arrazoirik; izan ere, Euskaltzaindiaren betebeharretakoa bat da euskararen inguruko ondare ahalik eta osatuena jasotzea, -hor idatzizkoa ez eze, ahozkoa ere sartzen dela-, eta ondorengo belaunaldiei transmititzeko bidea bermatzea. Egitasmo honen bitartez, beraz, euskara biziaren ordezkotza egoki bat atxiki nahi zaio Erakundeari eta, era berean, oraingo eta geroko ikerketetarako eta erabakietarako materiala prestatu.

Asko eta asko dira euskal kulturaren eta Erakundearen historian zehar Euskaltzaindiaren eta ahozkotetasunaren arteko lotura bermatu duten adituak, hurbileko batzuk aipatzeko, hor ditugu Azkue, Manuel Lekuona, Luis Villasante eta Antonio Zavala, adibidez. 2007. urtean, ordea, lotura hori berretsiz, ATHC egitasmoa aurkeztu zen Etxalarko barne-jardunaldien, eta une horretan Akademiaren jabetzapeko corpora zabaltzeko lan-prozesu bat abian jarri zen.

Lan metodologia zehatz eta emankorra duen Egitasmo honek, bere fruituak lantzen jarraitzen du eta jarraituko du etenik gabeko lanean. Euskal kulturaren, herri ezagutzaren eta lekuan lekuko hizkeren biltegi hautatu eta antolatu hau, etorkizunera begira baloratu beharreko ondarea dugu.

1. Egitasmoaren helburu nagusia

Ahozko euskararen, hizkera mintzatuaren, Corpus antolatu bat, Euskaltzaindiaren jabetzapekoa eta beronek kudeatua izango dena sortzea: hori du helburutzat Egitasmo honek.

Corpus hori, euskara mintzo den leku diferenteetako hizkeren lekukotasunetan oinarritua da. Lekukotasunon corpus hori, hautaketa, transkribatze, etiketatze eta jarraiko kudeatze batez, Euskaltzaindiaren ikerketarako oinarri izango da. Hala bada, CORPUS ANTOLATU BAT eraikitzea du helburu nagusia.

1.1. Zergatik Euskaltzaindia sartu zeregin honetan?

Euskaltzaindiaren zeregina da, besteak beste, hizkuntza bere osotasunean hartzea. Hizkuntzari, ordea, arlo desberdinetatik hel dakiokegu, baina zoritxarrez badugu joera bat, non uki daitezkeen materialak lehenetsi eta ukiezinak diren beste asko eta asko ahaztuta uzten ditugun. Hori horrela izanda ere, belaunaldiz belaunaldi transmitituz joan den ahozko tradizioa ere ondare moduan ulertu behar dugu, hizkuntzaren osotasunaren barnean kokatzen den ondarea.

Manuel Lekuona euskaltzainak 1935. urtean, jada, bere *Literatura oral euskérica* lanean joera hau mintzagai du. Honela adierazten du bere iritzia gai honi buruz: << El hombre de la edad de la imprenta vive fácilmente persuadido de la imposibilidad de la *literatura sin escritura*. El mismo nombre de “literatura” aplicado a la expresión artística del pensamiento, parece suponer la necesidad ineludible de la letra escrita (littera) para la literatura.

Sin embargo el mundo está lleno de temas literarios cuya dependencia de la escritura resulta absolutamente improbable. El folk-lore está lleno de “tradiciones” de innegable valor artístico, independientes en absoluto de toda ingerencia escrita.>> (LEKUONA, Manuel: *Literatura oral*

euskérica, Beñat idaztiak, Donostia, 1935, 14. orr.)

Ahozko transmisio bidea, ordea, eten egin da neurri handi batean, batez ere gizartea aldaketa sakonetan murgildu den unetik. Zehatzagoak izateko, transmisio bideak eurak ere aldatu egin dira denboraren joanean. Guzti honen ondorio argi-argia dugu belaunaldiz belaunaldi gorde den herri-pentsamoldea osatu duten elementuak desagertzeko zorian geratu izana.

Asko dira transmisioaren etena dakarten arrazoiak, baina jakin badakigu gure hizkuntzaren egoera eta posizionamendua zer nolakoa izan den gizarteak aldaketak jasan dituen neurrian. Horrezaz gainera, aldaketok bizimodua erabat eguneratzea ere ekarri dute, horrela bizimodu tradizionala itoz, zenbait ogibide desagerraraziz, sinesteak hutsean utziz, ohiturak ahaztuz eta, azken batean, hizkuntzaren barruan hauek betetzen zuten gunea hutsituz.

Ahozko hizkuntzak, zorionez, balio lexikografiko asko erakusten dizkigu, seguru aski, idatzizkoan aurkitu ezin ditugun balioak, eta hori gutxi balitz bezala, idatzizko historia sortzen den momentutik bertan datzan era berean, ahozkoa berri-emaile bakoitzaren ahoan berriidatzi egiten da horrek duen balio erantsia agerian utziz.

Laburbilduz: Ondare bizia da ATHC¹ honek bildu nahi duena, etorkizunean biltzea ezinezkoa izango dena, eta horretarako transmisioa arriskutan dagoen gunek lehenetsi dira. Argi dago idatzizko ondarea eta ondare bizia –ahozkoa– oso diferenteak direla, eta azken honen balioa etorkizunera begira baloratu behar da; gaurko lanaz baliatu, biharko fruituak bildu ahal izateko.

Ahozko testuen corpus hautatu eta antolatu bat, euskara biziaren eta osoaren ordezkotza egokia litzatekeena Akademiak bere esku izateak sekulako balioa du, azken finean, oraingo eta geroko ikerketetarako eta erabakietarako materiala da lantzen dena eta, esan bezala, Akademiaren

¹ Ahozko Tradizioko Hizkeren Corpora aurrerantzean horrela izendatuko da.

izate-arrazoietakoa bat da euskararen inguruko ondare ahalik eta osatuena jasotzea, –hor idatzizkoa ez eze, ahozkoa ere sartzen da–, eta ondorengo belaunaldiei transmititzeko bidea bermatzea.

2. Erakundearen nondik norakoak ahozkotasunean eta ATHCren sorrera

Denboran atzera jo beharra dago, Euskaltzaindiak ahozko tradizioarekin duen eta izan duen lotura estuaren nondik norakoak aipatu nahi izanez gero. Resurrección Maria Azkue, Akademiaren lehenengo buruak, herriaren ezagutza eta tradizioa zuzenean biltzeak zuen garrantzia dela-eta, bide hau jorratzeari ekin zion. Etengabeko lan nekaezin honen fruitua dugu Azkueren *Euskalherriaren Yakintza* (1935) bilduma. 1932an hasi zen Azkue aurretik bildua zuen materiala atontzen, eta 1935 eta 1947ra bitartean kaleratu zituen lau liburukiak. Garai bateko gizartearen testigantza den bilduma horretan folklorea, ipuinak, errefrauek, ohiturak eta abar batu zituen Azkuek, eta ezin uka daiteke itzelezko bilketa lana egin zuela.

Manuel Lekuona euskaltzainburuak ere izan zuen zeresanik ATHC lantzen ari den esparru honetan. Azkue eta Barandiaranekin batera, Euskal Herriak eman duen beste biltzailetako bat da eta Jose Ariztimuño “Aitzol”ek bultzaturik, orduko euskaltzaleek aski bazterturik zuten ahozko literatura goratzen saiatu zen. Horren egiaztagiri ditugu, besteak beste, *La poesia popular vasca* (1930), *Literatura oral euskérica* (1935) eta *Literatura oral vasca* (1964) lanak.

Euskaltzaindikoa buru izandakoekin aurrera eginik, Luis Villasanteren izena ere saihestezina da ahozkotasunaren eremura ekarri baliotsuak egin dituztenen artean. Hainbat etnotestu balioko ditu han-hemen argitaraturik, batez ere *Arantzazu* aldizkarian, eta hark bultzatu eta atonduak dira herri-hizkera batzuetako zenbait berri-emaile kualifikaturen lanak: Graziano Anduagarena lekuko.

Akademiarekin zerikusi estua izan duten hainbat adituren aipuak egin daitezke alor honi buruz. Sebero Altube berak *Erderismos* (1930) liburuan, herri hizkeren aldeko jarrera agertu zuen. Lan honen hitzaurrean zionez, euskal idazleen erderakadarik txarrenak joskerari dagozkionak dira eta alde horretatik euskararen nolakotasunik argienetakoa, esakuneei eman dakiekeen esangura desberdina zela aipatzen zuen. Hori horrela, honi irtenbidea aurkitzeko asmoz Altubek bere teoriako langaia bilatzeko herri euskarara jo zuen, euskaldun zaharren euskara bizia eredutzat hartuz, edota hurbilagoko bat aipatzeko.

Hurbileko beste euskaltzale eta euskaltzain bat aipatzeko, Juan San Martinen *Zirikadak* (1960) eta *Eztenkadak* (1965) bildumak ere hor ditugu. Bi lan hauetan hainbat ipuin eta pasadizo barregarri biltzen ditu San Martinek eta guztiak, nola ez, ahozketasunean oinarriturikoak.

Antonio Zavala ere ezin utz genezake puntu honetan aitatu gabe. Euskal herri-literaturaren ikerlari nekaezin honek hainbat urtetako lana eskaini zion ahozketasunari, euskal kultura biziari eta folkloreari. Horren egiaztagiri ezin hobea dugu *Auspoa* bilduma. *Auspoan* mota aski desberdinetako liburuak argitaratu ziren: hor agertu dira bertsolarien bertso bildumak, biografiak eta pasadizoak; herriko jendearen kontakizunak; garai bateko bizimoduaren berri ematen duten liburuak; erromantzeak; narrazioak... Euskal herri-literaturaren gordailurik oparoena da gaur egun *Auspoa* liburutegia, eta bertsolaritzaren ikuspegitik ez ezik, antropologiaren, historiaren, soziologiaren eta literaturaren ikuspegitik ere berebiziko garrantzia du. Biltzaile, ikertzaile eta eragile izan da Zavala izan ere; biltzen eta ikertzen ez ezik, lan handia egin du zer kontatua bazuen jendea bilatzen eta horiek idaztera bultzatzen.

Idatzizko tradizioaren corpus handiak ere Orotariko Euskal Hiztegiak (OEH), zeharka baino ez bada ere, ahozko tradizioa bildu egiten du; horra hor hiztegi honen aberastasun iturri agorreztetako bat.

Hori horrela, ozen esan daiteke Euskaltzaindiak bere hastapenetatik baduela lotura zuzena ahozko tradizioarekin, eta orain, 2007. urtean Etxalarko akademiaren barne-jardunaldietan Adolfo Arejita euskaltzainak egitasmo honen aurkezpena egin zuen unetik, eboluzionatuz, garatuz eta onduz joan da lotura hori.

3. Lan metodologia

Egitasmo bat martxan jartzeak aurretiaz lanerako ordena eta produktuaren tratamendua zehaztu eta berrmatuko duen lan-eskema bat zehaztea suposatzen du. Hori horrela, zortzi ataletan antolatu du ATHC-k bere lan prozedura; bilketan hasi eta dibulgaziora arteko bidea alegia.

3.1. Bilketa

Lehenik eta behin grabaketak, nori, non, zertaz, eta nola egin zehaztu beharra dago. Aipatu den bezala, transmisioa arriskuan dagoen guneak dira ATHC egitasmoak lehenesten dituenak, baina beste alde batetik, hiztun egokiak ere erreferentzia legez ulertzen dira Egitasmoaren helburua ikuspegi literario eta narratiboa duten testu ahozko kalitatezkoak biltzea baita. Orokorrean adin batetik aurrerako berri-emaien profila da bilatzen dena; izan ere, berauek dira aurreko garaietako ofizioak ezagutu dituzten pertsonak, bizimolde desberdinean bizi izan direnak eta, denborari esker, lekuan lekuko euskararen gordailu bihurtu direnak.

Bakoitzarekin zeri buruz hitz egingo den ere aurretiaz finkatu beharra dago. Hori horrela, galdetegiak prestatu edo behintzat aurretik sortutakoak egokitu egin behar dira kasuan kasurako. Bada J. M. Barandiaranen galde zerrenda etnologiko bat, oso ezaguna, baina hori oinarri legez erabiltzeaz aparte, berri-emaiak bakoitzak bere egoerari, inguruari eta bizipenei egokitzen zaien

galde-sorta bat merezi du. Materialen kalitateak kantitateak baino pisu gehiago du ATHC-n, edo hala izatea nahi genuke, behintzat. Horrela bada, informante izango diren norbanakoen profila ezagutu ostean sortzen edo taiutzen ditu ATHC-k galdetegiak.

3.2. Materialen tratamendua

3.2.1. *Formatua eta oinarrizko etiketatzea*

Gaur egungo material guzti hauek, nola ez, modu digitalean jasotzen dira, baina beste jatorriren bat duen grabazioaren bat jasoz gero, kaseteak nagusiki, digitalizazio prozesu batetik pasatzen dira modu horretan materialaren erabilgarritasuna eta iraunkortasuna bermatuz.

Bildutako materialok formatu aldaketa bat jasaten dute Akademiaren zerbitzarian sartu bezain laster. Material pisutsuak dira izatez, oso kalitate handikoak (.WAV), baina ATHC-ren helburuetarako eta lanetarako deserosoegiak. Horregatik, Audacity software libreaz baliatuz, .WAV artxibok .MP3 bihurtzen dira, baina etorkizunean etorri daitezkeen ikerketa edo bestelakoak ahortzi gabe, audioaren lehen bertsioa ere -WAV²- gorde egiten da.

Audacity-rekin egin beharrekoak, ordea, ez dira hor amaitzen. Lan materiala kudeatu ahal izateko identifikatu egin behar da. Oinarrizko identifikazio etiketak —metadatuak— ezartzen zaizkio audioari programa honen bitartez, eta horrela grabazioa burutu den leku, data, inkestagile eta inkestatuari buruzko datuak txertaturik gelditzen zaizkio geroako bilketak errazteko asmoz.

Artxiboa edo saioa bere osotasunean izendatzeko, Euskaltzaindiko onomastika sailak bere lanetarako zehatutako herriz herriko kodeketa sistema erabiltzen du ATHC-k. Kode hauen bitartez lurralde historikoa, eskualdea eta herria zehazturik gelditzen dira. (Adb.: 5.2.11.03 Etxalar;

² MP3 formatua, adibidez, ikerketa fonetikoetarako ez da oso egokia; izan ere, materialaren pisua murrizteko helburuarekin hainbat mozketak egiten dizkio soinuari, gizakiaren belarriak somatzen ez dituen mozketak, baina laborategi fonetiko baten berehala ageri direnak.

5.2.08.01.10 Orokieta; 7.3.00.16 Ürrüstoi-Larrabile)

3.2. 2. Hautaketa eta zatiketa

Kasurik gehienetan ordubetetik gorako iraupena duten grabaketak dira manipulatu beharrekoak, eta nola ez, ordu horietan guztietan hainbat eta hainbat gai lantzen ditu berri-emaileak. Egitasmoak, ordea, hautaketa prozesu bat egin beharra du material honen guztiaren artean. Izan ere, bere helburu argietako bat da kalitatea bermatzen duten materialak bakarrik lantzea: kalitatea tematikari dagokionez, kalitatea estilistika alderditik, eta azken baten, kalitatea hizkuntzari dagozkien alor guztietan.

Saio oso batetik hainbat eta hainbat “zati” labur atera daitezke. Zati hauetariko bakoitza “pasarte” bezala ezagutzen da eta, normalean, hiru eta bost minutu bitarteko iraupena izaten du. Selektzioa egina dago, beraz: Pasarte batzuk “interesgarri” moduan baloratuak izango dira eta beste batzuk, ordea, “interesgabe” bezala. Pasarte guzti hauek audio osoari txertatutako metadatuak mantendu egiten dituzte eta berauen identifikaziorako, saio osoari emandako kodea gehi zenbaki korrelatibo bana ezartzen zaie. (Adb.: Etxalarren kasuan 5.2.11.03-001; 5.2.11.03-002... 5.2.11.03-026).

Pasarte hauek guztiak –“interesgarriak” zein “interesgabeak”- dagokien herriaren izeneko karpetan gordetzen dira. Emandako azken urrats honekin eta izenean gehituriko zenbaki korrelatiboei esker, .MP3 formatuan ere saio osoa entzungai bihurtuko da nahiz eta zatituriko material legeaz aurkeztu

3.2.3 Etiketatzeta

Hautaketa selektiboa egina dago, beraz. Gordetako pasarte guztien artean, zenbaitzu baino ez dira izango lanerako interesgarriak; hori horrela, interesa agertzen duten horiek markatu eta

zehaztasunez etiketatzean datza hurrengo urratsa. Lan hau burutzeko, ordea, Audacity laga eta ID3-TagIT metadatuak kudeatzeko softwareaz baliatzen da ATHC.

ID3-TagIT honek soinu artxiboei identifikazio zehatzerako baliagarri izango diren zenbait etiketa txertatzeko aukera eskaintzen ditu, beti ere ahaztu gabe oinarrizko datuak, jada aurreko urratsetik esportaturik daudela. Etiketa berriok pasartearen “arloa”, “gaia”, “izenburua”, “pribatutasuna” eta “kalitatea” definituko dute.

“Arloa” eta “gaia”ren eremuetako datuak aurretik zehazturiko *thesaurus* luze eta zabal baten bilduta daude. Zerrenda honek zabal izaera eta tratamendua behar du, kontutan eduki beharra baitago ezinezkoa dela aurretiaz geroko elkarriketetan tratatu daitezkeen eremu tematiko eta gai guztiak zehaztea, beraz, zerrenda itxi bat sortzea bilaketa sistema hanka-motz uztea litzateke. Une honetan hamabi dira ATHC-k zehazturik dituen “Arloa”k, eta hauen barnean maneiatzen dituen gaiak beste ehun eta hoge inguru.

“Izenburua”ri dagokionez, “arloa” eta “gaia” zehaztu ostean, erabiltzaileari pasarte horretan lantzen den tematika zein den hobekiago identifikatzen lagunduko dion titular laburra dela esan daiteke; hiruzpalau hitzetan pasarte laburbiltzea da esparru honen helburu nagusia.

Materialaren “kalitatea” nolakoa den ezagutzea ere ezinbestekoa da. Baliteke urteen poderioz kalitatea galduz joan den materialen bat aurkitzea, edota zarataz beteriko grabaketa bat landu beharra izatea; hortaz, egoera guzti horiek kalifikatu egiten dira lau mailatako etiketak erabiliz: “Oso txarra”, “txarra”, “ona” edo “oso ona”.

“Pribatutasuna”ri dagokion etiketa, oster, etorkizunari begira txertatutakoa da. Badira kasuak, non pasartean zehar izaera guztiz pertsonala duten gaiak tratatzen diren (datuak, balorazioak, etab.), horiek guztiak markatu egiten dira, azken batean, etorkizunean materiala

argitara ateratzeko orduan arazo legal eta moralik ez aurkitzeko, besteak beste.

Guzti honekin, ID3-TagIT honek eskaintzen digun irudi nagusia “soinu artxiboen biblioteka” bezala ezagutzen dena izango da. Software libre honen bilatzaileaz baliatuz eta ezarritako etiketa guztiei esker, biblioteka guztiko materialen arteko selekzio zehatza egin daiteke: herriz herriko bilaketak, arlo jakin batekoak eta gai jakin bateko bilaketak egin daitezke, besteak beste.

Horretaz gain, programa honek hautaturiko materiala entzuteko aukera ere eskaintzen du. Hori horrela, grabazio bateko pasarte guztiak korrelatiboki antolatuta egoteari esker — “interesduna”k eta “interesgabea”k—, grabazioa oso-osorik entzuteko aukera ere eskaintzen zaio erabiltzaileari.

3.3. Transkribatzea

Soinua guztiz antolatu ostean testuak sortzean datza hurrengo urratsa. Hautatu ditugun material egokienak bakarrik transkribatuko dira eta horretarako erabilitako softwarea, librea hau ere, Transcriber izango da.

Transcriber-en aukeraketa klabea da materialaren etorkizuneko dibulgaziorako; izan ere, software honek XML formatua darabil oinarri legez eta, horrek epe laburrean web orrian soinu eta testuak publikatzeko asmoa izanez gero, lana asko aurreratuko luke. Bestalde, software honek baditu beste hainbat ezaugarri eta kalitate positibo: soinua eta testua uztartzen ditu adibidez, eta hori gutxi balitz bezala, hizlari desberdinak eta bakoitzaren txandak ere bereizteko aukera eskaintzen du, besteak beste. Ondorioz, ahozketasuna modu zintzo eta errealago baten aurkezteko aukera eskaintzen duen softwarea da Transcriber.

Ordea, transkribatzeak ere badu aurretik egin beharreko beste lanik. Irizpide eta arau ugari

zehaztu behar dira hasiera-hasieratik, eta ATHC egitasmoak lekuan lekuko euskara ahalik eta modu zintzoenean mantentzeko asmoa agertu badu ere irizpide nagusi bezala, dagokion euskalkiaren forma jasoa hobesten du. Bestalde, badira tokian tokiko hizkeraren funtsezko ezaugarri diren zenbait forma. Kasu horietan forma hori berori mantentzeko apustua egiten da, baina guztiz justifikaturik gelditu behar da zergatia.

Prozesu horren guztiaren ostean, hasieran bildutako audioa trataturik, zatiturik, hautaturik, etiketaturik eta, zenbait pasarte baino ez badira ere, transkribaturik gelditzen da. Material horren dibulgazioa da, beraz, hurrengo urratsa.

4. Corpora

Audio klasifikatua eta testu landuak sortzen joan ahala, Euskaltzaindia bere jabetzapekoa izango duen corpora zabalduz doa. Une honetan soinu artxiboak eta transkribapenak banatuta gordetzen dira, aurretik esan bezala, pasarteak ID3-TagIT-en bitartez kudeatzen direla eta testuak eskuz, herriz-herriko karpetetan.

Momentu honetara arteko material landuarekin corpora muntatzen hastea da, beraz, Euskaltzaindiko egitasmo honen epe laburreko helburuetako bat. Corpus bat, non soinua eta testua uztarturik aurkituko dituen erabiltzaileak, baina horretaz gain, txertaturik dauden metadatuak esker, zehaztasunez bilaketa emankorrak egiteko aukera eskainiko diona.

Denboran zehar elikatuz joan behar den egitasmoa da ATHC, etenik aurrikusten ez duena; izan ere, etorkizuneko ikerketa era erabakien oinarri izango den material antolatu hau, denborak aurrera egin ahala, zabalduz joan behar den corpus bat da.

5. Bibliografia

- AREJITA, Adolfo: “Ahozko ondarea biltzeko egitasmoa. Euskaltzaindia: Dialektologia Batzordea” *in* Euskera LII (2007), 269-274.
- AZKUE, Resurrección M^a: *Euskalerriaren Yakintza*, Bilbo, 1935-1947.
- EUSKALTZAINDIA: *Euskararen liburu zuria 1*, Bilbo, 1978.
- LEKUONA, Manuel: *Literatura oral euskérica*, Donostia, 1935.
- LEKUONA, Manuel: *Literatura oral vasca*, Donostia, 1964.
- MURUA, Imanol eta TORREALDAI, Joan Mari: *Euskaltzaindia ekin eta jarrai*, Euskaltzaindia, 2009.