

EUSKALTEL ETA EUSKALTZAINDIAREN ARTEKO HITZARMENA ETA XX. MENDEKO EUSKARAREN CORPUSAREN AURKEZPENA

Zamudio, 2002-III-08

*Alberto Agirreazaldeg
UZEIko kudeatzailea*

Nire aurrekoek esandakoak entzun eta gero, zaila da hemen gauza berriak esatea. Nik labur hitz egingo dut, gero aurkezpen tekniko egiteko beta izan dezagun.

Azken mende osoko euskara idatziaren erabileraren eta eboluzioaren isla da gaur aurkezten dugun hau. XX. mendean euskaraz argitaratu diren testuak Internet bidez ikus daitezke. Baina ez dira testu soilak bakarrik ikusten. Euskaraz idatzi den corpusa kontsultatzeko aplikazio informatiko aurreratu bat prestatu da eta internetez egin daiteke gainera kontsulta hori.

Beraz, XX. Mendeko euskal corpusa, edonork kontsulta dezake orain, internet bidez, webgune honetan: www.euskaracorpora.net

Proiektu handia eta luzea

EEBS (Eguno euskararen bilketa-lan sistematikoa) izenarekin jarri genuen martxan proiektu hau, 1987an, Euskaltzaindiaren eskariz. Helburu zehatz batekin sortutako corpus baten aurrean gaude, beraz, hain zuzen ere biharko Hiztegi Hiritar Arauemailea izango denaren oinarritzko material gisa sortu baitzen.

Beraz, urretik esan den bezala, orain dela 15 urte, UZEIk akordio bat sinatu zuen Euskaltzaindiarekin eta Herri-Agintekin proiektu honetan hasteko. Orduetik 7-8 laguneko lantaldea aritu da lanean eta, beraz, gaur aurkezten den proiektua ez da atzokoa, azken 15 urtetan egindakoa baizik.

Gaur egun corpusen garrantzia gero eta handiagoa da hizkuntza guztietan eta corpus erraldoiak dira gainera. Ezagunak dira RAEren CREA (Corpus de Referencia del Español Actual), IECen CTILC (CORPUS Textual Infor-

matizat de la Llengua Catalana) edo galegoen CORGA (Corpus de Referencia del Galego Actual) gure inguruko hizkuntzetakoak aipatzeagatik.

Euskaldunok ere, hauen parean, euskararen erreferentzi corpus erraldoia beharko genuke, oraingo hau corpus estatistikoa baita.

Ezaugarriak

- Sarean kontsultagai dagoen lehen corpora.
- Sailkatua: garaia, euskalkia eta generoa.
- Ia 5.000.000 testu-hitzekoa.
- Lematizatua: aldaera guztiak forma estandarrean batuak bilaketa errazteko.
- 100.000 lema desberdin: sarrera arruntak + hitz anitzelko unitate lexicakal (hitz elkartuak, eratorriak, esapideak...).
- Estatistikoa: obra zati asko euskal argitalpenen inbentarioan oinarritua.
- SGML (Standard Generalized Mark-up Language) formatu estandarrean kodetua.
- Autoreak eta izenburuak ere, ematen dira.

Baliabide informatikoak

- Oracle datu-basea.
- Kontsultarako Intermedia erabiltzen da.
- Programaziorako Java erabili da.
- Aplikazio mota: Servlets

Erreferentzi corpusaren beharraz

Euskaltzaindiak aspaldi egin zuen corpusaren aldeko apustua: tradizioa jasotzen duen *Orotariko Euskal Hiztegia*-ren oinarrizko obrak alde batetik, eta XX. Mendeko euskara biltzen duen *EEBS* bestetik. Hain zuzen lan hauetan oinarritu dira *Orotariko Euskal Hiztegia* eta *Hiztegi Batua* besteak beste.

Lehen esan bezala, euskararen erreferentzia corpus erraldoia beharko genuke osatu lehenbailehen. Inguruko hizkuntzek horrela egin dutelako, testu

osoak biltzeko erraztasunak gero eta handiagoak direlako, euskarri informatiboan eskura daitezkeelako eta informazio zabalena edukitzeak lan tresna fidagarriagoa egiten dutelako.

Bestalde, Euskaltzaindia mintzatu da corpusaren beharraz –ikus XV. Biltzarreko ondorioak– eta aitortua dio UZEIri gai honetan duen eskarmentua. Hau guztia kontuan hartuta, corpus erraldoia osatzeko irizpideak aztertzen duen proiektua prestatzea da aurtengo UZEIren egitasmoetako bat eta UZEIk aurkeztuko dio proiektua Euskaltzaindiari, honek aurten gogoeta egin dezan. Nolanahi ere, Euskaltzaindiak badaki UZEI prest daukala proiektu honetan lankide izateko eta hemendik bertatik eskerrak ematen diot Euskaltzaindiari guri eman digun konfidantzagatik eta gure lankidetzak eskaintzen diot behar duen guztirako.

Bukatzeko, UZEIk begi honez ikusten du Euskaltzaindiaren eta Euskaltel-en arteko *Hitzarmena*. Lehenago ere UZEI eta Euskaltelen lankidetzari esker Canal 21 Interneteko ataria sortu zuenean UZEIren Euskalterm datu bankua eskaini zuen, orain Eusko Jaurlaritzaren Euskadi.net-etik eskaintzen bada ere. Gero, corpora eskaintzeko ere lehen harremanak izan genituen, Jose Antonio Ardanzarekin eta harreman hura gaur hemen sendotua atera da. Eta Euskalteli esan nahi diot ez dela damutuko. Estatu mailan hizkuntzaren inguruko proiektu erraldoiak babesten diarduela Telefonica ere, RAEkin egindako hitzarmenei esker, eta Euskal Herrian, bide bera jarraituz, Euskaltzaindiari eta euskararen inguruan sortzen ari diren proiektu erraldoiei aurre egiten lagunduz prestigio handia ematen diola hizkuntzari eta berak ere gizartearen aitormen zabalala lortzen duela. Eskerrak beraz biei ekimen zabal hau bultzatu dutelako.

Orain, Miriam Urkiak, proiektuaren arduradunak, ordenadore bidez Kontsultaren nondik norakoak azalduko dizkigu eta, ondoren, Jose Ramon Zubimendik Jagonet-en aurkezpena egingo du.