

## **EEBS PROIEKTUAREN 2. FASEA:** **Abiaturua, helburu estrategikoak eta epe laburreko asmoak**

*Laurgain-Aia, 1991-XI-28*

*Andoni Sagarna*

### **1. Abiaturua**

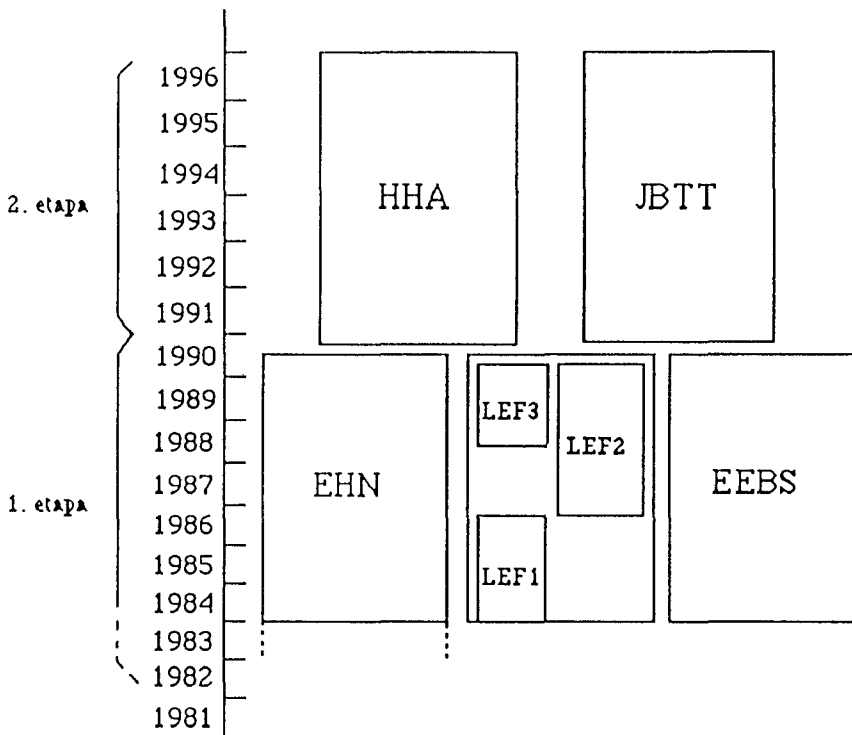
Dakizuenez, EEBS proiektua aurrera eramateko asmoa Hiztegitzinta-batzordearen epe luzerako planean jaio zen. Egungo Euskararen Bilketa-lan sistematikoa, *Orotariko Euskal Hiztegia* eta Lexikologi Erizpideak Finkatzeko batzordearen lanak, ondoren egin beharreko Hiztegi Hiritartu Arauemailearen eta Jakintzagai eta bizigune desberdinetako terminologia teknikoen oinarri izan behar zutela esaten zen egitasmo hartan.

Ez da alferrikakoa izango, agian, puntu horretaraino iristeko, Euskaltzaindiak egin zuen bidea gogoratzea. Oraintxe hamar urte, 1981eko abenduaren 29an, Hondarribian bildurik, Euskaltzaindiak adierazpen bat egin zuen, hiztegitzintzari garrantzi berezia emanez. Bertan esaten zen hizkuntzarekiko plangintza orokor batean sartua aurkitzen dela hiztegitzinta, eta Euskaltzaindiak erantzun egokiak eman nahi zizkiola honi. Azalpen honen ostean maila eta lehentasunak bereizten zituen hiztegitzintzan. Hiztegi Orokorra, orain OEH deitzen duguna alegia, ahalik eta lasterren burutzeko asmoa zuela, "hiztegi hiritartua" martxan jartzeko batzorde bat sortzea beharrezkoa ikusi zela eta, azkenik, hiztegi espezializatuak egitea eta banan banako azterketa egitea ez dagokiola Euskaltzaindiari esan zuen.

Hondarribiko adierazpena ez zen hitz hutsetan geratu eta 1982ko martxoaren 26an Euskaltzaindiak zenbait erabaki hartu zituen, bai Hiztegi Orokorrak lantokia, tresnak eta lantalde bat izateko bidea jarriz eta bai Hiztegi hiritartuaz arduratuko zen batzordea izendatuz ere.

Hurrengo urrats inportanteak, 1983ko apirilean, Euskaltzaindiak Iker Sailaren barnean Hiztegitzinta-Batzordea izendatzen duenean, eta 1984. urtean epe luzerako plana onartzen duenean, ematen dira. Azken hau inportantzia handikoa izan zen, urte batzuetan euskal hiztegitzintzan Euskaltzaindiak egin edo gidatu beharreko lanak mailakatu eta denboran programatzen zituen erabaki akademikoa hezurramitzen zuelako.

*Euskera* aldizkarian (XXXI zenbakia (2. aldia), 130 or.) argitaratu zen grafikoa dakargu hona epe luzerako lan-programaren adierazgarri:



*Orotariko Euskal Hiztegia*, euskal lexikoa mendetan zehar nolakoa izan den agertzen ari zaigu, Lexikologi Erizpideen Finkapenerako Batzordea, euskal hitzak eratzeko baliabideen azterketa xehea egiten, h. d. maileguzko hitzen idazkerak dituen arazoak edo hitz-elkarketak eta eratorpenak eskaintzen dituzten aukera aberatsak argitzen.

Egungo Euskararen Bilketa-lan sistematikoak —Larraun-Lekunberrin 1987an eginiko gogoeta-saioan ezarritako mugen arabera— XX.mendeko euskara idatziaren berri ematen duen corpus bat eratzea izan du helburu azken bost urteotan. Bertan 2.000.000 testu-hitz aurkitzen dira beren testuinguruekin, kokapen-erreferentziekin eta lemekin. Bilduma honek corpus zerrendatu baten itxura du, ondoko irudiko adibidean ikus daitekeenez:

hitza-erref.	esaldia	burdina	lefn8
33 054017308106 0001		burdin hari galbaniztuz egindako ardallak.	burdin hari
33 054017514105 0045	...a liria horretarako :	burdin berriak , gur trinak , ibi gurak eta..	burdin hari
33 054017514114 0079	... altzairuak eta	burdin hari galbaniztakoak	burdin hari
33 054017517271 0018	Garrantziakoa :	burdin industria hartun suberak izan dir..	burdin industria
33 054015415913 0012	...kastapereke	burdin industria horiei burdindolak deitena ..	burdin industria
33 054004209806 0048	...i Euzkoal Herria :	burdin neatz eta .	burdin neatze
33 054004209803 0045	...a zabalait elasoitu :	burdin neotogiak ( Arrapeta , Logandi , et..	burdin neotogi
33 054006123117 0050	...xiotetan nagusia eta :	burdin metalak , zere oxidazio zentralatik ..	burdin metal
33 054004209804 0106	... , bidualaketa zuzi	burdin produktua XI, ) senda bestapereki I..	burdin produktua
33 054005064009 0089	...lo 5 , 3000-6000 tona	burdin produktuen selarik , hots , europ..	burdin
33 054018802716 0016	... - krontzea eta	burdin urtua eg .	burdin
33 054003097010 0052	...i burdin erraira eta	burdin-neatze produktuk dira , produktu z..	burdin neatze
33 054004133118 0026	...i liria erro baretake	burdin-urak .	burdin ur
33 054010001717 0001		burdin-urra enaten du eta " Pol-pol ) * du ..	burdin ur
33 054016009711 0025	...aitok : Sufra ( S )	burdina ( Fo ) Prozedura [kurts] : , ) Pasa..	burdina
33 054006018703 0016	...ura , barris ,	burdina , erlojuak , gure iagruan ikusten ..	burdina
33 054014601111 0006		4 , ) burdina , { Fo } .	burdina
33 054014600404 0079	...egilua handia dira :	burdina , kobrea , kobaltak , lodoa , flua..	burdina
33 054016009711 0028	...open pisa itzazu 6 g	burdina , berazren koloreak aguntatuz , etl..	burdina
33 054018701518 0087	...a liriola itxakita ( burdina ,	alkantira eta urtua ) .	burdina
33 054004014318 0140	...ste berrea eta berrea	burdina .	burdina
33 054004014309 0065	... iklunatu , bal eta	burdina oro erabaten erabaten elandiar bl..	burdina
33 054006182421 0085	...ik botetuen badozu ,	burdina metalikoa [forn] burdina(II) ibia..	burdina
33 054019814704 0027	...iri* [dn] ( subegil :	burdina [kurts] berotzeko [kurts] su [kurts]..	burdina
33 054014600213 0001		burdina [bit] ( Fo ) .	burdina
33 054006123321 0110	...ina metalikoa [forn]	burdina(II) izen bihurtuko zaigu .	burdina
33 054006123321 0007		ad , ) burdina- apurten gindan kezo(II) -eko dl..	burdin apur

Proiektu hau lantzeko prestatu den tresneria informatikoari esker, corpus zerrendatu hori modu desberdinetan kontsulta daiteke, hala nola: testuinguru zabalagoa ikusiz, lehen arabera alfabetikoki ordenatuz, testu-hitz edo lema jakin baten bilaketa eginez, testu-hitz baten agerpen-maiztasuna eskatuz, etab. (Ikus *Egungo euskararen bilketa-lan sistemati-koa. EEBS programa I. fasea txostena*).

2. Helburu estrategikoak eta epe laburrerako asmoak

Hiztegi-gintza-Batzordearen esparruan azken urteotan egiten ari garen lanak "azpiegiturakoak" dira. Ez dira ospe eta distira handia ematen duten horietakoak. Gero egin beharreko eraikuntzen zimenduak sendoki finkatu nahian gabilta. Badakigu berehalakoan ez ditugula ikusiko fruitu mardul begi-gozagarriak; epe luzerako lanean ari garela eta inola ere ez goizetik gauera loratu eta hurrengo goizerako zimeldua dagoen lore liluragarri bezain iragankorra aldatzen.

Honelako lanei ekin egin behar zaie, nahiz eta ez jakin zehatz-mehatz haietatik abiatuz luzarora egingo diren lan konkreituak zein eta nolakoak izango diren. Alabaina, egiten den lanaren zentzua gal daiteke, zenbait neurri hartzen ez baldin bada. Hauek lirateke norabidea ez galtzeko, etengabe eta ziklikoki egin beharreko ariketak:

- 1. Azken helburuaz gogoeta egin, uneko ikuspegitik.

2. Azken helburu horretarako bidean hurrengo bide-harria edo tarteko helmuga definitu: zer lortuko den eta noizko.
3. Hurrengo bide-zatia egiteko beharko diren baliabideak zehaztu.
4. Egindako lanari eta hurrengo urratsean egingo denari epe laburrean ere probetxua ateratzeko bideak bilatu.

Ez da ona, ez urrunegi dagoen helburu bat bakarrik, tarteko helmugarik gabe, edukitzea, eta ez berehala probetxua atera nahiak itsutu eta azken helburua galtzea.

Hiztegitza-batzordearen 1985eko epe luzerako planeko lan-programei irizpide hauek aplikatuz gero, ondorio interesgarriak atera ditzakegu.

### 2.1. Planaren lerro nagusiak birpentsatzeko zenbait oinarri

Plan hark bi etapa bereizten zituen, lehena 1984-1990 eta bigarrena 1990-1996, baina idatzi zenetik onartu zenera pasatako denboraren eraginez, planaren hasiera atzeratu egin zitzaigun.

EEBS proiektuaren kasuan, hau burutuko zuen taldea izendatu arte beste atzerapen bat gertatu zen. Lanak ez ziren hasi 1987ko iraila arte. Bukatzeko epemuga, berriz, 1991ko urte-bukaeran ezarri zen.

Eskuetan ditugun datuen arabera, epemuga horretarako ez da zeharo bukatua geratuko lana, corpusaren zati baten lematizazioak 1992ko hila-bete batzuk hartuko bait ditu.

LEF programaren atal inportante bat, (hitz-elkarketaren eta eratorpenaren azterketa, epe luzerako planean LEF 2 zeritzana) 1992aren bukaeran amaitu nahi litzateke eta lexiko-berrikuntzarako semantikazko auke-irabideak lantzeko xedea duen LEF 3 oraindik bideratu gabe dago.

OEH dugu Hiztegitza-batzordearen lanen artean pisu eta alkanztzurik handienekoa. Bertan ari direnei dagokie, noski, gai honetaz hitz egitea, baina, batzordearen azken helburuez gogoeta egiteko behar-beharrezkoa den heinean, har dezagun hartaz ere hausnarketatxo bat egiteko lizentzia.

Lan eskerga da benetan OEH. Hasieran uste genuen baino handiagoa seguruenik. Hala dirudi, zeren eta epe luzerako planean esan genuen 1984-1990 epean lehentasuna izango zuela eta ondoren, hura menderatua izango genuelakoan, 1990-1996 epealdian beste programa batzuk aurrera eramateko egokiera izango genuela, baina gaur egun badakigu bukaera urrun dagoela oraindik. Hala ere, honek ez luke izan behar etsitzeko edo arrangura sortzeko arrazoi.

Azken helburuen eta tarteko bide-harrien arazora itzuliz, lasai esan daiteke OEHren kasuan tarteko helmuga garrantzitsu bat gauditua duela. Izan ere, corpusa informatizatua, corpus zerrendatua eta fitxategiak pres-tatuak edukitzeak zilegi egiten du erabat OEHren oinarritzko informazio-bilketa bukatua dagoela esatea.

Urrun dagoena, beraz, azken helburua da. Tartean beste bide-harririk ezartzen ez badugu, hainbat fruitu ematen aritu beharko lukeen altxor bikaina antzu edukiko dugu, fruitu horien premia larrian dagoen euskal-dungoa goseak dagoen bitartean. Gainera, azpiegiturako lanean ihardu-teak berez duen iluntasuna luzatu egingo dugu proiektuaren beraren kaltetan. Baliabideak jarri behar dituenak, nahiz emaitzen zain dagoenak, ikusmen estrategiko zorrotzik ez badu, —eta hori badakigu zentzumen urria dela gure artean— nekez ulertuko du zertan ari den lantalde bat hainbeste denbora “ezer gutxi” produzituz.

Gauza bat da OEHren helburu nagusia den paperezko edizioa eta beste bat OEHren corpus informatizatua. Honek sekulako jokoa eskain dezake planteamentu egokia egiten bada. Aurreraxeago zehaztuko ditugu aukeratako batzuk.

Beste printzipio garrantzitsu bat ere gogoan izan beharko genuke: Hiztegitzinta-Batzordearen lanaroko programa desberdinen arteko siner-gia bilatzea komeni zaigula.

1984eko epe luzerako planean 1996rako helburu bezala jartzen zen Hiztegi Hiritartu Arauemailea atzeratu beharra dago, zalantzarik gabe, obra erabat bukatu bezala planteatzen bada. Hala ere, azken helburu horretarako bidean hurrengo bosturtekoaren bukaerarako, tarteko helbu-ru batzuk ipin daitezke, gauden egoeratik abiatuta, aurrerago esango den bezala.

## 2.2. Nola ustia daiteke corpus informatizatu bat?

Corpus informatizatuen tratamendua heldutasunera iritsi den jakin-tzagai sendoa da gaur egun. Corpus-hizkuntzalaritza funtsean hizkuntz erabileraren eta hizkuntz barietateen azterketa deskribatzailea da. Corpus batek, testu-zati askoren bitartez, era desberdinetako hizkuntz erabilera-n azalpen adierazgarria egitea du xedetzat.

Corpus informatizatuen hainbat eta hainbat adibide aurki daitezke han eta hemen. Adibide gisa banaka batzuk aipatuko dira:

*Institut de estudis catalans*, Rafel irakaslearen zuzendaritzapean, pres-tatzen ari den katalan modernoaren corpusa, berrikitan martxan jarria den *Corpus de referència del español*, *Trésor de la langue française*, Ale-

maniako Mannheim-en *Institut für deutsche Sprache*-k sortu duen alemanaren datu-base lexikografikoa (LEDA), Holandako Nimegako Unibertsitatean dauden ingelesezko, arabierazko eta espainierazko corpus informatizatuak, Lancaster-Oslo/Bergen (LOB)ingeles britaniarrezko corpusa, etab.

Gure kasuan, hizkuntzaren batasuna eraikitzeko, behar beharrezkoa da, literatur tradizioaz gain, batezbesteko erabilera literario garaikidea ere ongi ezagutzea. Horiexek dira, hain zuzen ere, OEH eta EEBS proiektuen xedeak, hurrenez hurren: euskararen mendetako ondarearen eta XX. mendean zehar argitaraturiko idazkietan ageri den euskararen ezagutza zehatza lortzea. Euskararen ezagutza diogu eta ez euskal lexikoarena, hiztegegintza-arloaz haraindik ere interesa izan dezakeelako bertako informazioak: sintaxi-mailan, esamoldeen azterketan, dialektologian, onomastikan, estilistikan, etab.

Hiztegegintzan eta hiztegegintzatik landa corpus bat ikuspegi desberdinetatik ustiatu ahal izateko, ikertzaileen irispidean jarri behar da. Zazpi giltzaz itxitako kutxan gordeta baldin badago, seguru egongo da, baina baita alferrik ere. Datorren bosturtekoan zehar lortu beharreko helburu importante bat OEH eta EEBS proiektuetako corpus informatizatuak ustiatzeko moduan jartzea izango da.

Bi eratako arazoak gutxienez sortzen dira corpusa ikertzaileen eskuetan jarri nahi denean: informatikoak eta jabegoarekin zerikusia dutenak. Bai batzuek eta bai besteek konponbidea badute, hartara jarriz gero, beraz, erabakiak hartzean dago gakoa. Arazo hauen azterketak txosten honen mugak gainditzen ditu eta gainera teknikoan laguntza eta berariazko batzorde baten lana eskatzen. Alabaina, pentsatzen hasteko, zenbait oinarri ezarriko da ondoko lerroetan.

### 2.3. Datu-base lexikografiko baten beharra eta hau eratzeko zenbait oinarri

OEHren eta EEBSren corpus informatizatuetatik abiatuz, epe luzerako lanei euskarri egokia eman nahi bazaie, aldi berean pixkanaka euskararen batasun- eta normalizazio-prozesuetarako material erabilgarriak kaleratzeko bidea emanez, *Euskararen datu-base lexikografikoa* eraikitzea litzateke egokiena. Honelako datu-base batek asko errazten du hizkuntz materialaren bilketa, analisia, egituraketa, eta deskribapena.

Lehenik eta behin, lexikografoari etengabe sortzen zaion kezka larri bati ematen dio irtenbidea. Hiztegegintzan ari zaretenok ongi ezagutzen duzuen arazoa da eta besteok berehala ulertuko duzue. Askotan gertatzen zaigu, edozein eragiketa sinple burutzeko, ikerketatxo bat egin beharrean aurkitzen garela, hau egin ondoren, bilatzen genuen emaitza jaso-

tzen dugula, baina bidean lantzen joan garen hainbat informazio interesgarri, era egituratu batean jaso gabe geratzen zaigula. Hurrengo batean konturatzeko gara, egun batzuk edo hilabete batzuk lehenago aztertu genuen zenbait puntu berriro aztertzen ari garela. Arazo hau asko larritzen da lan berean lagun bat baino gehiago ari direnean.

Adibidez, testu-hitz bati legokiokeen lema edo hiztegiko sarrera-burua zein litzatekeen erabakitzen ari garenean, era askotako galderak sortzen zaizkigu: zein izango da hitz honentzat aukeratuko dugun grafia? inguruko beste testu-hitz batekin ez ote luke osatu behar lexia bat? ba al dagokio sarrera-bururik ala azpisarrera batean joan beharko luke? testu-hitzean ageri den forma euskalki batean duena da ala idaztoker bat dago? hitzaren ezagutzen ez genuen adiera baten aurrean gaude, gaia ezagutzen ez dugulako testua gaizki interpretatzen ari gara ala autoreak desegokiro erabili du hitza testuinguru horretan? etab, etab. Analisi morfologikoa, fonologikoa, semantikoa, sintaktikoa eta ez dakit zenbat eratakoa egin dugu, baina, azkenean, beste batean eroso kontsulta daitekeen informazio-bilketa eta -egituraketarik ez dugu egin. Hori denbora- eta diru-xahutze ikaragarri eta onartezina da. Denbora eta dirua harrika botatzea da, eta ez bata, ez bestea dira horretarako. Hobe da denbora eta dirua zerbait iraunkorra eta erabilgarria eraikitzeke erabiltzea, hitz batean, inbertitzea eta ez xahutzea. Datu-base bat eraikiz hauxe da lortuko dena, hain zuzen ere, ikerketa lexikografikoaren emaitzak etorkizunean prozesatu ahal izateko moduan gordetzeko egitura bat sortzea. Besterik da, horri lehenbailehen etekina ateratzeko zerbait egin behar dela. Hori ere aintzakotzat hartu da eta ondoren esango da zehatzago zerbait horretaz.

Datu-base honek, jakina, sistema informatiko batean egon behar du eta erraztasunak eman behar dizkio hizkuntzalariari, ordenadorearekin modu interaktiboan lan egiteko. Sistema dagoen lekutik kanpoko ikertzailen irispidean jartzeko ere azter daitezke soluzio desberdinak: kopia eguneratuen aldizkako banaketa eta kontsulta telematikoa, esate baterako.

Datu-base baten ezaugarri nagusiak informazio-dukina, datu-bildumaren egituraketa, erabiltzaileak informazioa eskatzeko dituen erak eta datuak gehitzeko, kentzeko eta aldatzeko moduen antolaketa dira. Azken hiruak ezaugarri informatikoak eta estudio tekniko batek ebatzi beharrekoak dira. Bidenabar esan daiteke, ulertzen duenarentzat, *Oracle* moduko 4. belaunaldiko sistema bat SQL galdeketa-lengoaia erabiltzen duena litzatekeela egokiena. Informazio-dukinari dagokionez gauza gehiago zehaztuko dira hemen.

Datu-base honek izango lituzkeen xedeak asko lirateke, lehentxoago esan bezala, baina hiztegitza-arlokoak hartuko dira hemen kontutan:

a) Testu-hitzak eta lemak haiek dauden testuekin eta testuoi dagoen fitxategi bibliografikoarekin erlazionatu behar dira, beraz, fitxategi

bibliografikoek eta testu-fitxategiek bertan izan behar dute. Pentsatu beharra dago, bestalde, fitxategi bibliografikoek bere badutela berezko interesa.

b) Lexikografoak eroso lan egingo badu, eskueran (begien irispidan, pantailan bertan) izan behar ditu lehendik argitaraturik dauden hiztegi interesgarriak, besteak beste, ezin pentsa daitekeelako corpusean azaltzen ez den baina hiztegietan azaltzen den zenbait hitz hiztegi berrietan sartuko ez denik. Gure kasuan *OEH*, *Hauta-lanerako euskal hiztegia* (prestatuak dauden neurrian), *Euskararako hiztegia*, *Hiztegia 2000*, etab. izango lirateke. Argitaraturiko hiztegiekin batera argitara eman gabeko beste fitxategi batzuk ipin daitezke: Duden fitxategia, Hizpal fitxategia, Berran fitxategia... Urtetan zehar hiztegegintzarako, testuen zuzenketarako, eta antzeko lanetan eginiko ikerketa puntualen bilduma informatizatuak eta egituratuak dira hauek.

c) Lexikografoek lantzen duten informazioa era egituratu batean gorde beharko da. Gero, denboran zehar, hiztegi desberdinak prestatzea pentsatu ahala, bakoitzean artikuluek izango duten egitura formala erabaki beharko da. Ez da berdin izango definizioak dituen haurrentzako hiztegi bat prestatzea, hiztegi ortografiko bat egitea, hiztegi entziklopediko bat lantzea, etab. Bakoitzak bere egitura, metahizkuntza erabiltzeko bere era, etab. izango ditu, baina hiztegegileak iturburu beretik aterako du informazioa. Honek esan nahi du bi lan-mota daudela etengabe.

- Datu-base lexikografoaren enborra osatzen duen oinarrizko informazioa aberastea, zuzentzea, eguneratzea, eta
- Oinarrizko informaziotik abiatuz produktu lexikografiko konkrituak prestatzea.

Bi eginkizun hauek ez dago arras banantzerik, txostenaren hasieran azken helburuez eta tarteko bide-harriez proposatu diren irizpideak urratu gabe. Ezin gara ibili oinarrizko informazioa aurrena ongi landu eta gero aplikazioak edo produktu konkrituak egingo ditugula pentsatzen. Erantzun zain dauden premia askoren soluzioak “ad calendas graecas” uztea litzateke. Baina, luzarorako informazio-enbor sendo eta iraunkor bat egitea ahaztu eta aldian aldiko produktuak prestatzen ihardutea alferlan asko egitea litzateke.

#### 2.4. Datu-base lexikografikoaren erabilera hiztegegintzatik kanpora

Lexikografiatik kanpora honelako datu-base batek duen interesa azaltzen hasteko, egokia izango da hurbil-hurbilekoa den lexikologia aipatzea. Azken urteotan LEF batzordean egin diren lanak gogoan izatea aski da. Hitz-elkarketa eta eratorpena aztertzeko, berariaz aukeratu den corpus bat izan dugu eskuartean eta bertatik hitz jakin batzuk beren testuinguru



adierazgarriekin hartu ditugu. Ongi da. Horri esker, adibidetegekiak prestatu eta txosten asko egin ahal izan dira. Pentsa ezazue, ordea, zeinen gauza ederra litzatekeen atzizki (edo bukaera) bera duten lemak, hauei dagozkien testu-hitzak eta testuinguruak aukeratu ahal izatea. Nahi izanez gero denak eta nahi izanez gero autore batenak, epealdi edo euskalki batekoak edo beste irizpideren baten arabekoak. Gauza hauek, maila apalean behintzat, landu izan ditugu eta baita emaitza interesgarriak erdietsi ere. Hau aukera posible bat besterik ez da. Beste aukera batzuk, galdeketa-sistema egoki batek eskaini behar litzuzkeenak izan litezke: hitz jakin bat bilatzea, grafema-katea batez hasten edo bukatzen diren hitzak edo katea hori hitz-barnean duten hitzak bilatzea, esaldi berean edozein hitz-konbinazio edo grafema-kateen konbinazio agertzen diren kasuak bilatzea, hitz-formen maiztasun absolutu eta erlatiboak ezagutzea, etab.

Oinarrizko irizpideetan aipatzen zen sinergiaz baliatzea izango litzateke OEHko eta EEBSko bildumei LEFeko sailkapen-irizpideen markak ezartzea, hitz elkartu eta eratorrietan edo maileguzko hitzetan ageri diren egokitzapenak seinatzea...

Urrunxeago dauden beste arlo batzuekiko lotura oso eskueran daukagu gaur egun, Gasteizko Filologi Fakultateak, Deustuko Unibertsitateak, Donostiako Informatika-Fakultateak eta UZEIk elkarrekin dituzten harremani esker. Esate baterako, dagoeneko lortua dagoen analizatzaile morfologikoak izan lezake aplikaziorik honelako datu-base baten barnean eta, alderantziz, lengoia naturalaren prozesamenduaren arloan hain beharrezkoa den lexiko-mailako informazioa datu-basetik aterako litzateke. Honetarako, jakina, datu-basean makinak (eta ez gizakiak) erabili beharreko datuak ere sartzeari komeniko litzateke. Honela, gaur egun garatua dagoen analizatzaile morfologikoak, bi mailako morfologiaren formalismoa deritzana, erabiltzen du, eta, horretarako, lehen ondoan hauen jarraitze-klaseei buruzko informazioa duen hiztegi bat du edo analizatzaile morfologiko horretan oinarritzen den Xuxen zuzentzaile ortografikoak, idaztokerrak errazki hautemateko erabiltzen dituen estrategiaren artean maiztasun-hiztegi batez baliatzea baldin badago, hiztegi horren iturririk egokiena datu-basea litzateke noski.

Euskararentzako analizatzaile sintaktikoei buruzko ikerketan ezer gutxi egin den arren, syntaxiaz arduratzen direnak balia daitezke datu-baseaz eta honen tresnez gauza askotarako.

### **3. Ondoko bost urtetarako lan-programaren proposamena**

#### **3.1. Helburuak**

Datozen bost urtetarako lan-programan bi eratako helburuak jarriko ditugu:

1 Hiztegi Hiritartu Arauemailea hemendik hamar-hamabost urtetarako azken helburu bezala harturik, hartarako informazioa bilduko duen *datu-base lexikografiko erlazional bat eraikitzea*. EEBS eta OEHko corpusak hartuko lirateke oinarritzat, eta bestelako informazio osagarriak gehituko lirateke, tradizio lexikografikoa eta zeinnahi argibide lexikologiko bertan jasotzeko modua eginez.

Dagoeneko prestatua dugu UZEIn honelako datu-base baten prototipoa EEBSko 200.000 testu-hitzeko corpus-zati lematizatu batean oinarritua.

2 Euskararen batasun-prozesuan beharrezkoak diren *hiztegi-arloko material praktikoak prestatzea* Euskaltzaindiarentzat eta, honek erabakitakoaren arabera, baita argitalpenerako ere.

Hiztegi Arauemaile osatu batera iritsi aurretik Euskaltzaindia joan daiteke hainbat orientazio ematen hiztegi aldetik. Gaur egun horrelako zerbaiten premia handia dago euskal gizartean. Duela hogeitabost urtetik hona Euskaltzaindiak ortografiatz, deklinabideaz, aditzaz, maileguzko hitzen idazkeraz, etab. emandako arauen bidez batasun-bidea urratuz joan da. Bitartean euskara eskolan, komunikabideetan, administrazioan eta beste esparru batzuetan sartu da eta beharrian berriak agertu dira. Denbora honetan, 1968an Arantzazuko Biltzarraren ondorioz argitara eman zuen *Batasunaren Hiztegia* eta 1975ean kaleraturiko *Zortzi urte arteko ikastola hiztegia* ditu Euskaltzaindiak batasunerako eman dituen erabaki-mailako oinarritzko hiztegiak. Bistan da ez dagoela proportziorik euskararen urte hauetako garapenaren eta hiztegi-arloko araugintzak denbora berean, alderdi kuantitatibotik behintzat, izan duen pisuaren artean. Hiztegin-tzaren atzerapen horren arrazoiak, gainditzen ari garen egiturazko ahulezietan eta gai horretan lankide izan behar genuenon harremanak bideratzeko foro egokien faltan bilatu behar dira. Egitura-arazoak asko konpondu dira giza nahiz diru-baliabideak, antolakuntza, gestio ekonomikoa, kontrol akademikoa eta euskarri teknikoa jarri diren neurrian. Bestalde, azken urteotan Euskaltzaindiaren Batzorde eta Azpibatzerdeak topagune aproposak eta dinamika berri baten sortzaile ari dira gertatzen. Aldeko egoera honi eta egina den azpiegiturako lanari probetxua ateratzeko egokiera dugu orain.

### 3.2. Lan-prozedura

Hiztegi-alorreko erabaki bat edo gomendio bat kaleratu baino lehen lan handia dago:

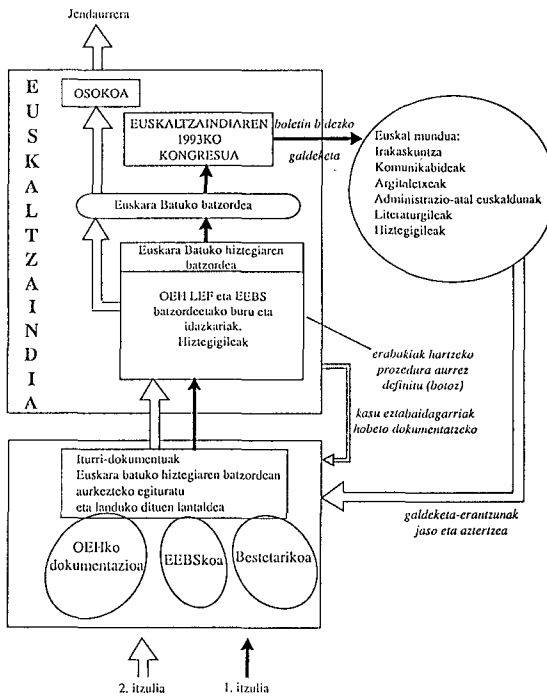
- a) Materialaren prestakuntzan: zerrendak, txostenak, etab.
- a1) Formari dagokionez aparteko arazorik ez duten hitzen zerrendak presta litezke lehenbailehen.

a2) Formaz edo/eta adieraz argitasun-beharretan diren hitzen inguruko txostenak landu behar dira, hartu nahi den erabakia behar bezala oinarritua eta dokumentatua gerta dadin.

b) Euskaltzaindiaren barneko eztabaida bideratzen. Euskaltzaindiak ihardupide eroso eta argia formulatu beharko du, jasoko duen informazioa zein izango den eta nola moldatua nahi duen zehatz finkatuz eta barne-eztabaidak eta hauen ondorioak modu ahalik eta sistematikoenean egituratzeko.

c) Euskaltzaindia da, noski, bere erabaki eta gomendioen erantzule bakarra. Dagokion erantzukizunaz gain, onarpen soziala ere begiratu nahi badu, ordea, erabiltzaileen iritzia eta kezak jasotzen saiatu beharko du (Ikus honetaz *LEF batzordearen proposamena Euskaltzaindiari* izenekoan egiten diren kontsideroak).

Zerrendak eta txostenak Euskaltzaindian aztertu beharko lirateke onspena emateko, informazio osagarria eskatzeko eta erabiltzaile kualifikatuen iritzia biltzeko. Prozesu hau aurrera eramateko era bat baino gehiago asma daiteke. Guk geure egiten dugu, hemen aurkezteko, Mikel Zalbidek asmatu eta proposatu ziguna:



### 3.3. Lan-programa

- EEBSren 1. fasean bildutako corpora 1987tik aurrerako testuetatik ateratako lagin batez eguneratu. Mendearen hasieratik 1987. urtea arteko corpus-zatia biltzeko eta hustutzeko irizpide eta prozedura berberak erabiliko dira, corpus osoa koherentea izan dadin.

- Bost urtetan zehar 50.000 sarrera-bururen formak finkatzeko behar den informazioa lantzeko konpromisoa hartuko genuke. Kopuru hau baino txikiagoa ez litzateke aski izango gure gizarteak dituen beharrei erantzuteko. Laneko erritmoa ondokoa izango litzateke urtez urte:

Urtea	Sarrera-kopurua
1992	5.000
1993	15.000
1994	15.000
1995	10.000
1996	5.000

Lehen urtean ezin izango litzateke material gehiago prestatu, lehen-bizi EEBSren 1. fasea errematatu eta 2. faserako prestakuntza egin beharko litzatekeelako. Handik aurrera bi urtetan erritmoa mantenduko litzateke baina azken bi urteetan moteldu egingo litzateke, errazetik zailera joango ginatekeelako.

Landuko litzatekeen lehen zerrenda arazo berezirik gabeko hitzena izango litzateke hain zuzen.

- Ikuspegi lexikologikotik landutako zerrendak:
  - \* LEF batzordeak dagoeneko landuak dituen sailkapen-irizpideen arabera hitz elkartu eta eratorriei dagozkien lemak markatuz gero, adibidetei aberatsak presta daitezke; hitz elkartuak eta eratorriak beren testuinguruekin sailez sail eta maileguzko hitzak egokitzapen-arazoen arabera zerrenda litezke.
  - \* LEF batzordeak oraindik landu gabe dauzkan beste arloetako adibidetei egokiak ere para litezke: zabalkuntza semantikozko lexiko-sorkuntzari eta hedabide sintagmatiko desberdinei dagozkienak, alegia.

- Ikuspegi morfosintaktikotik landutako zerrendak: juntagailuen erabilera, aditz-erregimenen erabilera, osagaien segidak, etab.
- Ikuspegi pragmatikotik landutako zerrendak: esapideak, lokuzioak,...

### 3.4. Txostenetan gogoan izan beharreko irizpideak

Zerrenden lanketarako ezarri den erritmoari erantzun ahal izateko, hilean mila hitzeko zerrenda bat aztertu beharko luke batezbeste Euskara batuko hiztegiaren batzordeak, Euskara batuko batzordeari aurkezteko. Lan hau ezinezkoa litzateke arazoak dituzten hitzei buruzko txostenek egitura simple eta zehatza ez balute.

Derrigorrez metodologia sistematikoa diseinatu beharko da, ahal dela literatura askoko txostenen ordez parametro edo tasun batzuen deskribapen laburra bilatuz. Ez da hau, noski, diseinu hori eztabaidatzeko leku egokia, baina parametro horien lerro nagusiak zein izan litezkeen aipa genezake:

- \* Formaren tradizioa corpusean eta lexikografian, dialektoen (Iparralde-ekialde vs hegoalde-mendebalde, mailen (jasoa/zaindua/zainduga-bea), berbaldi-moten (ahozkoa/idatzizkoa; tekniko/arrunta) araberrako banaketa.
- \* Formaren bizitasuna vs arkaikotasuna.
- \* Forma lehiakideen maiztasuna.
- \* Forma bakoitzaren emankortasuna, lexikogenesiaren ikuspegitik: eratorriak eta hitz elkartuak eratzeko edo zabalkuntza semantikoz adiera berriak sortzeko erabiltzen al den, alegia.
- \* Forma bakoitzak duen erregularitasuna, hau da analogiaz paradigma batean txertatzeko egokiera.

Azken hitz gisa, esan beharra dago txosten hau ez dela lan bukatua, Hiztegitzinta-Batzordeak lanean segitzeko abiapuntua baizik, alabaina, hona hemen laburbildurik Euskaltzaindiari proposa dakizkiokeen erabakigaiak:

1. Euskara batuko hiztegiaren batzordea sortzea.
2. OEH eta EEBSren corpusekin Euskararen datu-base lexikografikoa sortzeari onespena ematea.
3. Datozen bost urtetan 50.000 sarrerako euskara baturako hitz-zerrenda moldatzeko lan-programa onartzea.