

# El diccionario de la Academia de la Lengua Vasca - Euskaltzaindia

Madrid 15/03/2012

La Academia publicará en 2012 una nueva edición de su Diccionario Normativo.

Esta obra será un paso más en el camino hacia la normalización y la normativización de la lengua vasca.

La normativización del euskera, a pesar de ser el gran objetivo de la Academia desde su nacimiento en 1919, arranca definitivamente en 1968.

# El plan de Koldo Mitxelena

Establece un orden de prioridades:

1. Ortografía
2. Forma de las palabras antiguas de la lengua
3. Formación de palabras y adaptación de préstamos
4. Morfología del nombre y del verbo
5. Sintaxis

Propone una lista de cerca de 2.000 formas del acervo léxico

# El plan de Koldo Mitxelena

Ese mismo año de 1968 propuso los criterios que se debían tener en cuenta para elaborar un diccionario en el que las palabras vascas estuvieran descritas en euskera

1. El diccionario debía ser general, es decir debía contener palabras de todos los dialectos y de todas las épocas.
2. Debía ser una guía para hablantes, escritores y lectores.
3. Se debían fijar los límites en cuanto a qué variantes registrar y cuáles no, qué hacer con los arcaísmos, los préstamos y su adaptación, qué hacer con los neologismos.
4. El diccionario debía ser selectivo, para ayudar al usuario del mismo a distinguir lo bueno de lo no tan bueno.
5. Había que realizar una recopilación sistemática del léxico utilizado en la literatura vasca a lo largo de la historia.

# 1970-1980

## La Academia

1974 *Merkatalgo Izendegia*

1975 *Zortzi urte arteko ikastola hiztegia*

Pequeños vocabularios: *Filatelia, Arkitektura*

## Otros

1970 *Batasunaren Kutxa* de G. Aresti y X. Kintana

1972 *Ikastola Hiztegia*. Gordailu

1974 *Idazketa Hiztegia*. J. Fernandez Setien, A. Amondarain

1976 *Natur Zientziak. Hiztegia*. Jakin

1977 *Eskola Hiztegia*. X. Gereño

1977 *Hiztegi teknikoa*. L. M. Mujika

1977 *Euskal Hiztegi Modernoa*. X. Kintana y otros

1978 *Banka Hiztegia*. Varias Cajas y Bancos

1978 *Matematika. Hiztegia, hizkera, irakurbidea*. M. Zalbide

1979 *Fisika Hiztegia*. UZEI

1980 *Kimika Hiztegia*. UZEI

1980 *Politika Hiztegia*. UZEI

1980 *Psikologia Hiztegia*. UZEI

K. Mitxelena: "Se está poniendo el carro delante de los bueyes"

# 1981

La Academia, preocupada, celebra unas jornadas de reflexión para fijar la estrategia que va a desarrollar en cuanto al léxico

1. El Diccionario General y el moderno son parte de los quehaceres de la Academia.
2. La elaboración de diccionarios de especialidad no entra dentro de su ámbito de actuación.
3. Euskaltzaindia ve con buenos ojos que cualesquiera otros los elaboren, ya sea individualmente o de forma colectiva.
4. Euskaltzaindia ve la necesidad urgente de crear una Comisión de Lexicografía.

# 1983

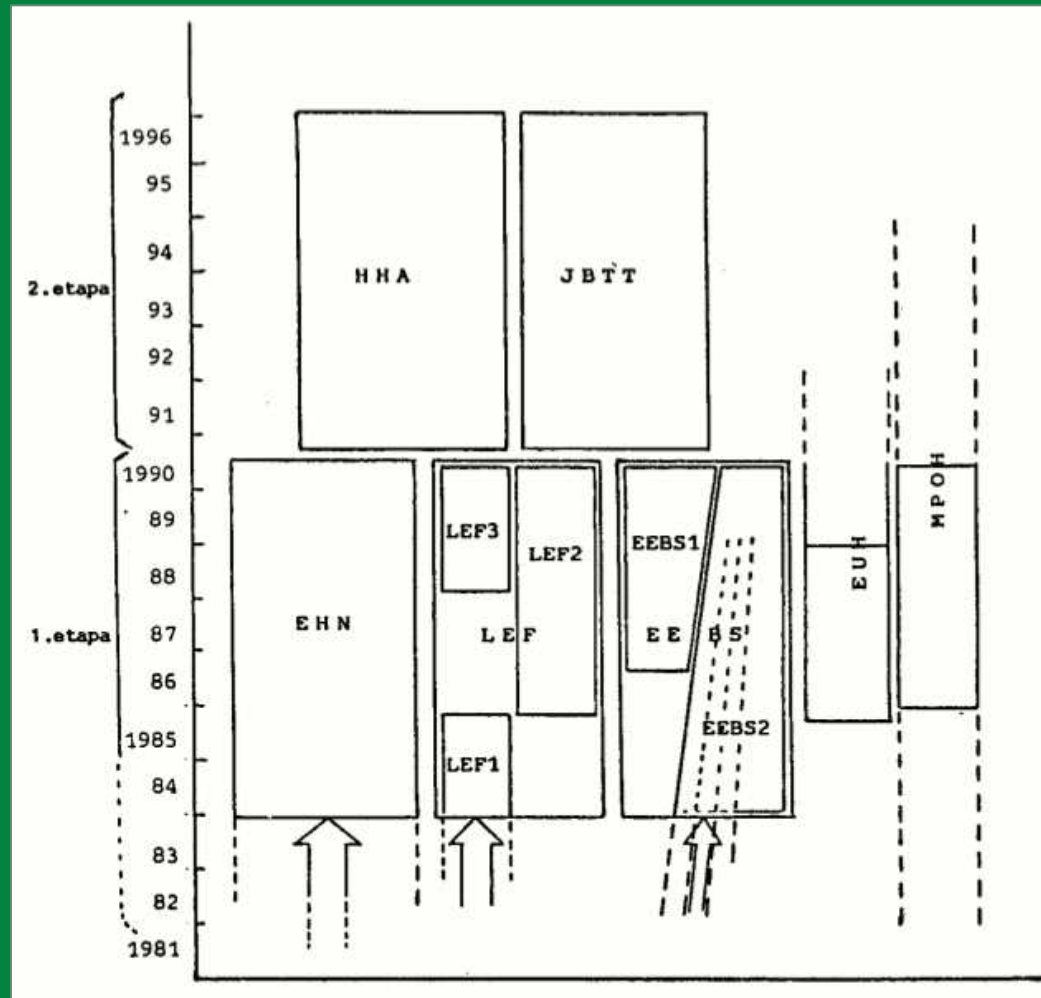
Euskaltzaindia crea una Comisión de Lexicografía con los siguientes cometidos:

1. Llevar a cabo el Diccionario General o Diccionario de Mitxelena.
2. Ocuparse de otros diccionarios de la Academia.
3. Fijar los criterios para adaptar los préstamos.
4. Investigar en el ámbito de la normativización del léxico

# 1984

Por fin se dispone de los medios para iniciar los trabajos de elaboración del Diccionario General, bajo la dirección de Koldo Mitxelena.

La Comisión de Lexicografía propone un plan de trabajo para el período 1984-1996:





# 1987

Fallece el profesor Mitxelena y poco después se publica el primer tomo del diccionario. Asume la dirección su ayudante Ibon Sarasola

# 1987-2005

Se publican los 16 tomos del Diccionario General Vasco.

14.000 páginas

Un corpus de 6 millones de palabras texto

# Características del *Diccionario General Vasco*

lema

Tradicición lexicográfica

**zahi** (V, G, AN, L, BN, Ae, S, R; Ht *VocGr* 425, Arch *VocGr*, *VocBM*, Gèze, Dv, H; zai Lcc, Lar, Añ, Izt 50r, Dv, H, Zam *Voc*), **zagi** (L, B, Sal; *VocB*), **sahi** (V-gip; SP; sai Mic 9r, Dv (G)), **sei** (V-arr). Ref.: Bon-Ond 151; A (zai, zagi); Lh (zagi); Lrq; Arin *AEF* 1960, 69; Satr *VocP*; Iz *Uiz*, *Als* 84, R 295; Etxba *Eib(saix)*; *CEEN* 1970, 349; Ibarra *Dima (seye)*; Elexp *Bérg*; Izeta *BHizt2* (zagi).

Acepciones

Y. Salvado. "Soma y salvado, *saia*" Mic 9r. "Afrecho, salvado" Lar. "Allegadores de la ceniza y derramadores de la harina, *zaietan zur ta iriñetan ero*" Ib. "*Zaietan ziogi, irunetan ero*, mezquino en salvado y pródigo en harina" Izt 50r. "*Ogi arinak zagi amitz* (B)" A. "*Zaietan urun, urunetan zaikoak gara gu* (V-ger)" Ib. "*Zaiaz uruna egin ezirik*, sin poder conseguir a fuerza de regalitos [...] (V-ple)" A *EYIII* 334. "El salvado mayor" Iz *Uiz*. Cf. *Zai-Zaku*, apodo en Urruz *Urz* 61. Gure *zaiok iriñ eske*. RIs 46. *Zaietan zur eta iriñetan ero*. Ib. 19. *Florea, irina, hoherena, gaztetasuna, eramaiten du etsai gaiotoak, eta gero nahi duzu kontenta dadin lainkoa zahiaz eta zahartzeko hondarrak*. Ax 182 (V 123). *Ez horak zahia jan, ez oilloer ützi*. Bela 22 (tbn., con alguna variación, en Saug 25 y O Pr 248). *Hobe da zahi hutsa ezi ez aho hutsa*. O Pr 246. *Lastasakuba baiminzan ala, / kendu jeuste zaiak*. Acto 295. *Zai apur bat galdu ez dedin bildurrez, ta urun ederra galduarren ardurarik ez*. Mg *CO* 20. *Eioten da, eralgi, alderatu zai*. Añ *LoraS* 55. *Irtan zaitte bada errota orretatik, bestela, zai, ta iriñ usuraz kargaturik ondatuko zera su eta garren lize izugarrian*. AA II 190. *Ogijari karia, zaija edo beste gauzank ezarten deutsenak*. Astar II 165. *Bein nagusi ena egin eztero, eta esan daroagun legez, lanari zai a kendu eztero, erraz izaten da zerbat geiago aziten*. Zav *Fab RIEV* 1907, 91. *Irina behar eta pagatzen du zahiz*. AstLas 74.

Ejemplos de uso extraídos del corpus

# La Comisión para la fijación de los criterios Lexicológicos (LEF) 1986-1994

Comité de gestión: 5 personas  
Plenario: 23 personas

1. Criterios para la adopción de préstamos.
2. Estudio de los procesos de formación de palabras

# Corpus estadístico del siglo XX 1987-2000

- Inventario exhaustivo de todo lo publicado
- Modelo estadístico
- Universo: > 200 millones de palabras texto
- Muestra inicial: 2 millones de palabras texto
- Muestra final: 4.655.300 palabras texto
- Muestreo estratificado: el período en el que se publicó el texto (3), el dialecto utilizado (7), el tipo de texto (14) y el tamaño del mismo (4)
- Digitalización + OCR + corrección manual
- Lematización manual de todo el corpus

[http://www.euskaracorpora.net/XXmendea/Konts arrunta fr.html](http://www.euskaracorpora.net/XXmendea/Konts_arrunta_fr.html)

# La página de consulta del corpus estadístico del siglo XX

The screenshot shows the website interface for EuskaraCorpus.net. At the top, there are logos for Euskaltzaindia and UZEI, and the URL www.EuskaraCorpusa.net. A navigation bar contains links: Ataria | XX. mendeko euskararen corpus estatistikoa | Nola erabili | Kontsulta arrunta | Kontsulta aurreratua.

The main content area is divided into several sections:

- Estatistikak:** Includes a search box with the instruction "Idatzi galdera, aukeratu bilaketa mota eta sakatu 'Bilatu' botoia". Below it are radio buttons for "Lemak" (selected) and "Testu-hitzak", and a "Trunkatzeko %, \_ karaktereak erabil ditzakezu" option with a "Bilatu" button and a "Laquntza" link.
- Epea (Períodos):** A dropdown menu with options: Guztiak, 1900-1939, 1940-1968, 1969-1990, 1991-1999.
- Euskalkia (Dialectos):** A dropdown menu with options: Guztiak, Bizkaiera, Gipuzkera, Zuberera, Lapurtera-Nafarrera.
- Testu-mota (Tipos de texto):** A dropdown menu with options: Guztiak, Saio-artikuluak, Administrazio-idazkiak, Ikaliburuak, Saio-liburuak.
- Kontsulta arrunta:** Includes the instruction "Idatzi testu-hitzak edo lemak koma bidez banatuak". Below it is a "Consulta" section with a dropdown menu set to "Lemak", radio buttons for "Hitz bat" (selected), "ETA", "EDO", "Hasieran", "Bukaeran", and "Tartean", a text input field, and a "hitz gehienez" dropdown set to "Ordenatua". There are "Bilatu" and "Garbitu" buttons, and a "Laquntza" link.

At the bottom right, it says "Egilea: UZEI || © Euskaltzaindia 2002".

# Ejemplo de consulta del corpus estadístico del siglo XX

Aplikazioak Lekuak Sistema

XX. mendeko Euskararen Corpusa. - Mozilla Firefox

Fitxategia Editatu Ikusi Historia Laster-markak Yahoo! Irresnak Hiztegiak Laguntza

http://www.euskaracorpora.net/XXmendea/Konts\_arrunta\_fr.html "se culmino"

Most Visited Getting Started Latest Headlines

Aukerak ES EU Euskalterm Elhuyar 3000 Hiztegi Batua Adorez sin. UZEI sin. Itzul Harluxet Wikipedia Mokoroa Intza Eurovoc Bergara

WEB SEARCH Bookmarks Mail Weather Forecast Stock Tracker Anti-Spy

Ubuntu Start Page (1 unread) Yahoo! Mail, ando... Euskalerria.org euskarazko l... Kargatzen...

EUSKALTZAINDIA  
EUSKARA BATERA  
www.EuskaraCorpusa.net  
UZEI

Ataria | **XX. mendeko euskararen corpus estatistikoa**  
Nola erabili | Kontsulta arrunta | Kontsulta aurreratua

### Estatistikak

Idatzi galdera, aukeratu bilaketa mota eta sakatu "Bilatu" botoia

Lemak:  Testu-hitzak:

Trunkatzeko %, \_ karaktereak erabili ditzakezu

Bilatu  
Laguntza

203. 1969-1990 Euskara Batua Saio-artikuluak [A. Eguzkitza\\_0221](#)  
Higikortasun horren froga aipatu dugun azken problema dugu: modal bezala erabiltzen direneko euskalkietan ere, aditz modalen **era** perfektua aditz semimodalen forma bera da.

204. 1969-1990 Euskara Batua Saio-artikuluak [A. Eguzkitza\\_0216](#)  
Azkenez lthurri gramatikariak dakarren etsenplu hau ere aipa dezakegu: <mention></mention> nik dakidanez euskalki gehienetan **era** honetara itzul litekeena: <mention></mention>.

205. 1969-1990 Euskara Batua Saio-artikuluak [A. Eguzkitza\\_0216](#)  
Beste **era** batera esatearren, honako kasuan <mention> *behar izan* </mention>-ek aspektu-atzizkiak hartzen ditu eta egoera hauxe dugu: <mention></mention>.

206. 1969-1990 Euskara Batua Saio-artikuluak [A. Eguzkitza\\_0218](#)  
Jakina, perpaus biak beste **era** desberdinez ere adieraz daitezke, eman dezagun: <mention></mention>.

207. 1969-1990 Euskara Batua Saio-artikuluak [J.M. Larrarte\\_0222](#)  
Sintagma mugagabe hori, lehen esan dudanez, bi **eratako** sintagma izan daiteke: atributo sintagma edota izen sintagma.

208. 1969-1990 Euskara Batua Saio-artikuluak [J.M. Larrarte\\_0223](#)  
**Era** mugatuan, berriz, hogeiren bat bakarrik azaltzen dira.

209. 1969-1990 Euskara Batua Saio-artikuluak [A. Eguzkitza\\_0220](#)  
Lehenago aipatu dudan lthurriren adibidean (lthurri 1979, 40): <mention></mention> zera esateko erabili dugu: <mention></mention>, hau da, **era** perfektua adierazteko.

210. 1969-1990 Euskara Batua Saio-artikuluak [X. Amuriza\\_0073](#)  
Gehiegitasuna adierazteko aditza bera **era** superlatiboan jarzea ere ondo legoke, jakina: <mention>Kargatuegi = lar-kargatu</mention>.

211. 1969-1990 Euskara Batua Saio-artikuluak [Zabal 1976\\_0012](#)  
- 28 banku nazionalizatuak izan dira, baina lehengo **eran** segitzen dute lanean.

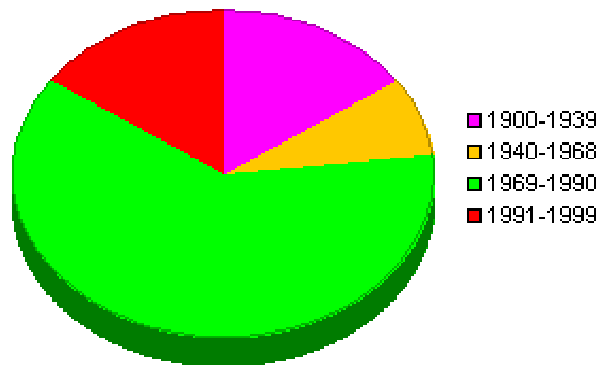
www.euskaracorpora.net --> Datuak transferitzen...

Eguneraketa-Kudeatz... XX. mendeko Euskar... txostena.doc - Open... aurkezpena.ppt - Ope...

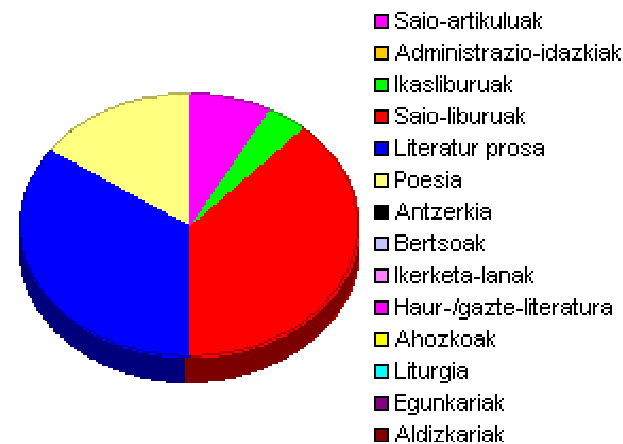
# Ejemplo de estadísticas de ocurrencias de una voz en función del período, del dialecto y del tipo de texto

*Lema: zabaldi*

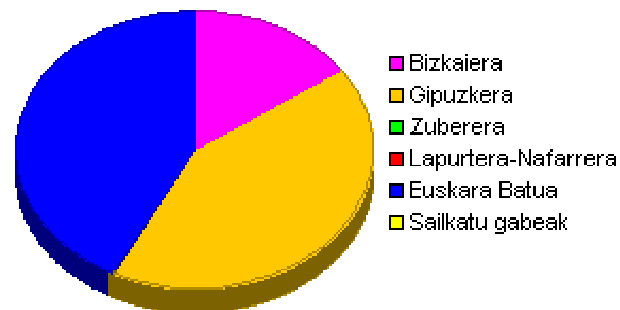
**Epea**



**Testu-mota**



**Euskalkia**



# 1991

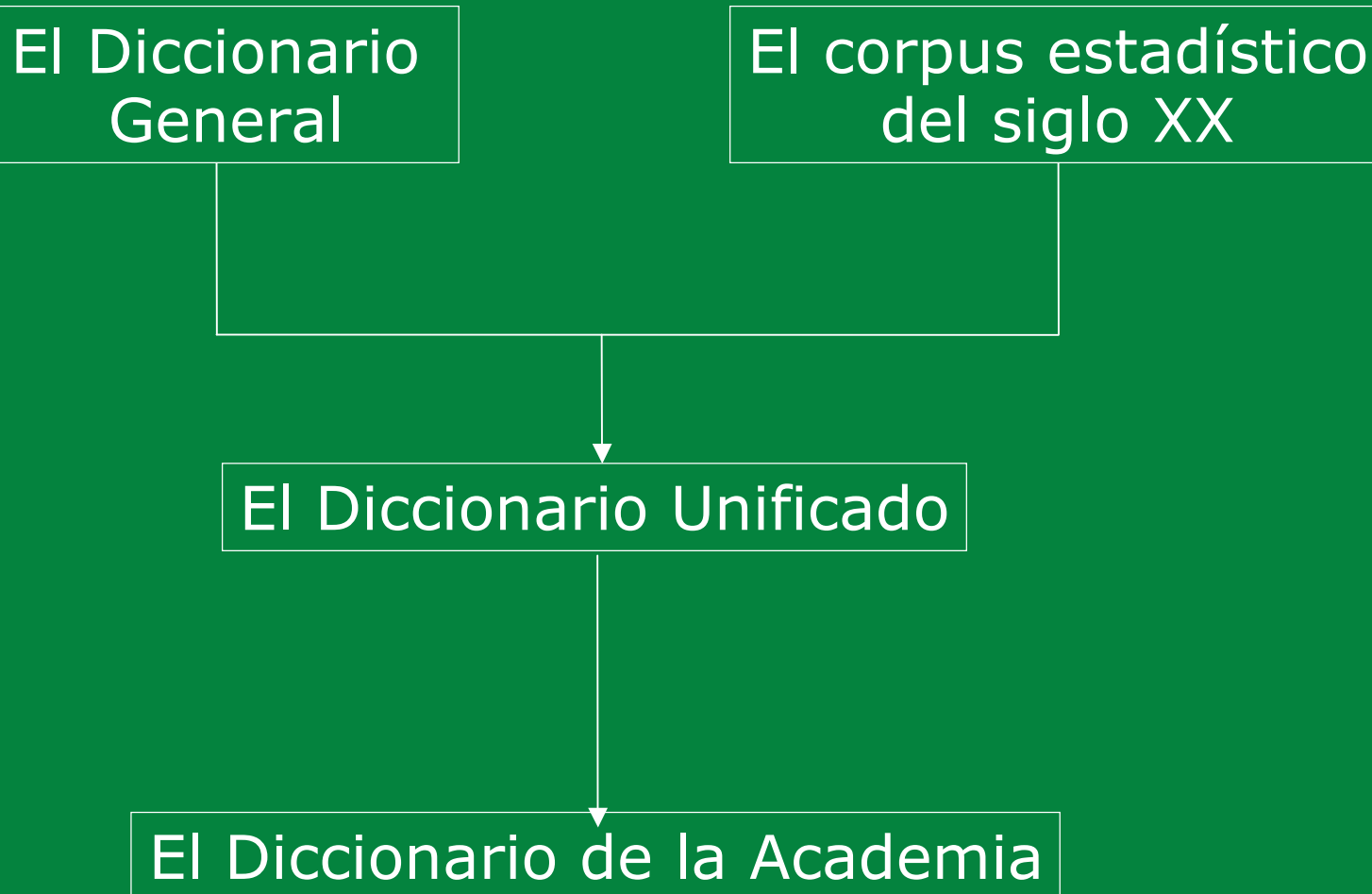
La Comisión de Lexicografía propone al pleno de la Academia un plan de acción para el siguiente lustro.

El Pleno de la Academia adoptó los siguientes acuerdos:

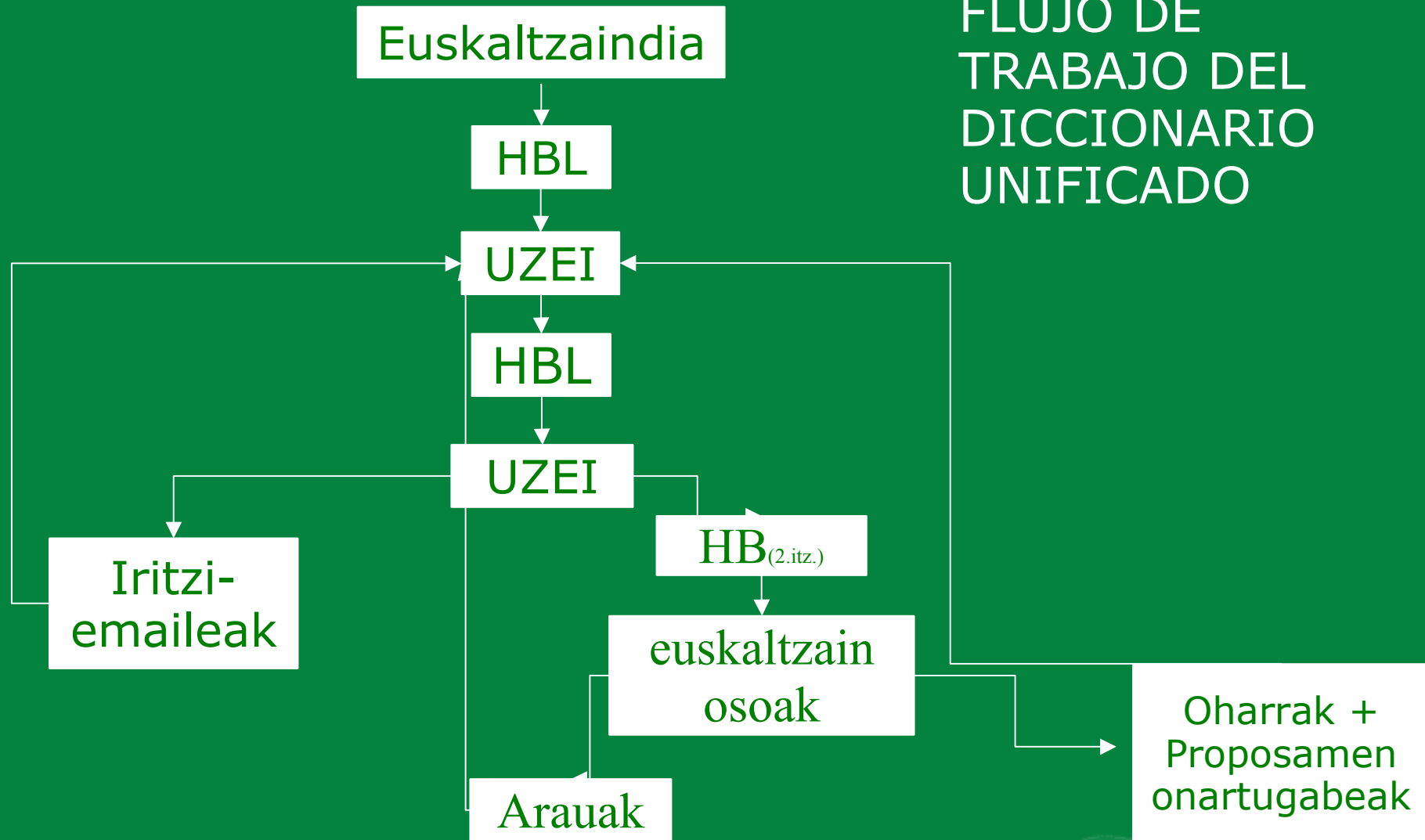
1. Crear la comisión del diccionario unificado (Hiztegi Batua).
2. Crear la base de datos lexicográfica del euskera, aunando el corpus del Diccionario General y el del siglo XX.
3. Poner en marcha la elaboración del diccionario unificado, para tenerlo a punto al cabo de cinco años.



# La secuencia de los proyectos



# El flujo de trabajo del Diccionario Unificado



# El Diccionario Unificado

	<b>tobera</b>	entrada	subentrada	ejemplo de uso
	<b>tobera-mustra</b>			
	<b>togoar</b> (herritarra)		taxon	
homógrafos	<b>toil 1</b> <i>Mustelus mustelus</i>			
	<b>toil 2</b> <i>batez ere Ipar.</i> izond. 'pisua, motela': <i>gorputza toil delarik, loa uxatzea zail da.</i>			
	<b>toildu</b> , toil(du), toiltzen. <i>da/du</i> ad.		régimen verbal	
	<b>tok pisin</b> (hizkuntza)			
	<b>toka 1</b> adlag. 'gizonezko bati hika'		categoría gramatical	
	<b>toka 2</b> iz. 'jokoa'			
	<b>tokamen</b> <i>Heg.:</i> <i>irabazietan bakoitzari dagokion tokamena.</i>			
	<b>tokatu</b> , toka(tu), tokatzen. <i>Heg. da/zaio</i> ad. <i>Herr.</i>			
	<b>toketa</b> 'gizonezkoarekiko hitanoa'			
	<b>tokharia</b> (herritarra)			marca de registro
	<b>toki</b>			
	<b>-en tokian</b>			
	<b>tokian tokiko</b>			
	<b>tokitan:</b> <i>tokitan dabil!</i>			
	<b>tokitara:</b> <i>tokitara joan da!</i>			
	<b>tokialdaketa</b>			
	<b>tokialdatu</b> , tokialda, tokialdatzen. <i>da/du</i> ad.			
	<b>tokika</b> adlag. 'han eta hemen'			
	<b>tokikotasun</b>			
	<b>tokikotu</b> , tokiko(tu), tokikotzen. <i>da</i> ad. 'bertakotu'			
	<b>tokilabilaso</b> <i>Ipar.</i> 'aitonaren aitona, tokilasoa'			
	<b>tokilaso</b> <i>Ipar.</i>			
	<b>tokilo</b> <i>Ipar.</i> 'moldakaitza, jendetasunik gabea'			

La 1ª edición en papel se publicó en 2000 con 20.000 formas

# Consulta del Diccionario Unificado en Internet

## Hiztegi Batua

A - B - C - D - E - F - G - H - I - J - K - L - M - N - Ñ - O - P - Q - R - S - T - U - V - W - X - Y - Z

Forma:

Bilatu

1 / 3051

>> Hurrengoak

**a 1** 1 iz. 'euskal alfabetoaren lehen letra' 2 iz. "'a' letraren izena'

**a bildu** (@ ikurraren izena)

**a 2** interj.: *a!*

**aaiundar** (herritarra)

aba\* e. **aita** (elizgizonaren titulua)

**ababor** iz. [batez ere leku-denborazko kasuetan erabilia]: *ababorrean ur-zulo bat dago.*

**abade** 1 iz. 'monasterioko burua' 2 iz. *Bizk.* 'apaiza'

**abade egin** *Bizk.* 'apaiz egin'

**abade-etxe** edo **abade etxe** *Bizk.* 'apaiz etxea'

<http://www.euskaltzaindia.net/hiztegibatua>

# 2005

Finalizada la edición del Diccionario General se inicia el proyecto del Diccionario de Euskaltzaindia:

*Euskaltzaindiaren Hiztegia. Adierak eta adibideak*

20.000 formas de la 1ª edición del Diccionario Unificado enriquecidas con definiciones y ejemplos de uso procedentes de la base de datos lexicográfica.

Se publicará en papel en otoño de 2012.

<http://www.euskaltzaindia.net/hiztegibatua>

# 2009

El Pleno faculta a una comisión para agilizar la ampliación del Diccionario Unificado con otras 20.000 voces. El trabajo quedará culminado en 2012 tras haberlo sometido a la aprobación por parte del Pleno.

Faculta a una segunda comisión para supervisar los trabajos de elaboración del Diccionario de la Academia. Estos finalizaron en diciembre de 2012.

<http://www.euskaltzaindia.net/hiztegibatua>

# El Diccionario de Euskaltzaindia

entrada	categoría gramatical	definición	Ejemplo de uso
<b>tobera</b>	iz. pl.	Andregaiari ezkontza bezperan, gorabeherak dituzten senar-emazteei edo berriro ezkontzen den gizon alargunari egiten zaien musika edo zintzarri joaldia. <i>Toberak jo ditzagun.</i>	
<b>tobera-mustra</b>		Ezkontza hausteren bat-edo gertatzen denean, herriko plazan antolatzen den antzerkia. Ik. <b>xaribari</b> . <i>Tobera-mustra batzuk muntatu.</i>	
			referencia cruzada
<b>togoar 1</b>	izlag.	Togokoa, Togori dagokiona. <i>Togoar agintariak.</i>	
<b>2</b>	iz.	Togoko herritarra.	
<b>toil1</b>	iz.	Marrazoaren eta arraiaren ordenako itsas arraina, handi samarra eta gorputz-luzea, azala latza eta kolore-grisa duena ( <i>Mustelus mustelus</i> ).	
		marca dialectal	taxon
<b>toil2</b>	izond.	<i>Batez ere Ipar. Pisua, motela. Gorputza toil delarik, loa uxatzea zail da. Pilota toila. Adimendu toila. Euskara hain beratz eta toil baldin badago, apaizek daukatela enbalditurik.</i>	
--			régimen verbal
<b>toildu</b>	toil/toildu, toiltzen	<i>da/du ad. Toil edo toilago bihurtu, moteldu. Bihotza toildu.</i>	
--			
<b>toka1</b>	adlag.	Gizonezko bati hika. Ik. <b>noka</b> . <i>Eta hasi zitzaion erregeri toka. Hurrei ez zaie toka eta noka hitz egin behar.</i>	
--			
<b>toka2</b>	iz.	Kaleko jokoa, distantzia jakin batera dagoen metalezko xafla estu baterantz metalezko diskoak jaurtitzean datzana, eta xafla jotzea helburu duena. <i>Sagardoa galanki edango dugu, tokan jokatzen dugun bitartean. Tokan errege eta maisu.</i>	

# El Observatorio del Léxico

- Es un corpus oportunista
- Contiene textos publicados a partir de 2000, básicamente en medios de comunicación
- Se alimenta constantemente
- 17 millones de palabras texto
- Etiquetado y lingüísticamente anotado



# El Observatorio del Léxico

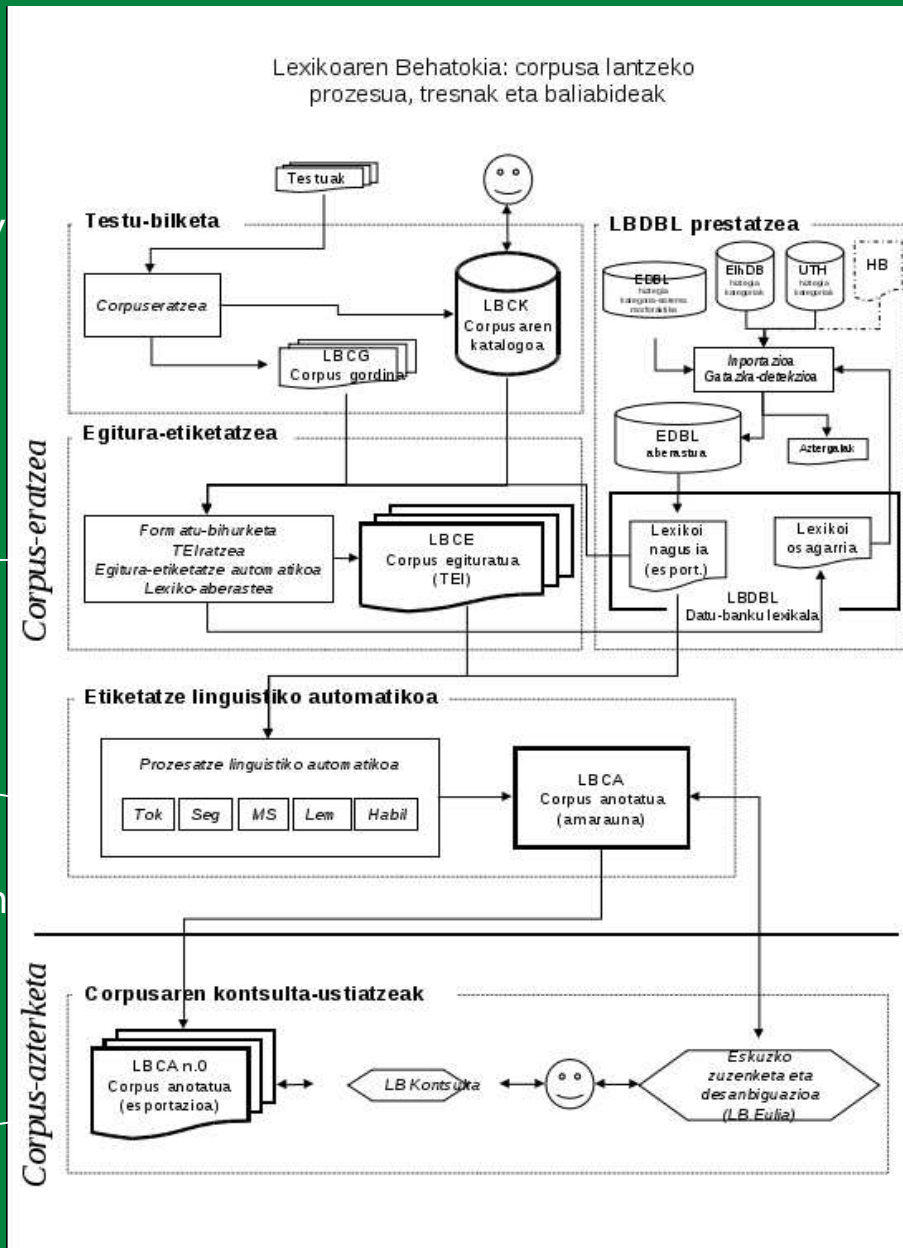
## El flujo de trabajo

Recopilación y catalogación

Etiquetado de la estructura (TEI)

Tokenización  
Segmentación morfológica  
Desambiguación automática

Desambiguación manual  
Corrección



Las herramientas y recursos de Procesamiento de Lenguaje Natural son la clave para construir y mantener el Observatorio del Léxico con costes y plazos razonables.