



CORPUSA



Universidad del País Vasco Euskal Herriko Unibertsitatea

 **acceso**

ÍNDICE

PRENSA

2

RADIO

22

TELEVISIÓN

24

INTERNET

26



PRENSA

SELECCIÓN DE LAS NOTICIAS DE SU INTERÉS



Fecha	Titular/Medio	Pág.	Docs.
15/03/13	"EHUREN EGUNGO TESTUEN CORPUSA EUSKARAZKO BILDUMA HANDIENA DA, ETA MUNDUKO CORPUS HANDIENETAKOA " / BERRIA (HITZA BIZKAIKO)	4	1
12/03/13	EGUNGO TESTUEN CORPUSA (ETC) / DEIA	5	1
12/03/13	205 MILLONES DE TÉRMINOS EN EUSKERA / DEIA	6	1
12/03/13	NACE EL MAYOR CORPUS DE EUSKERA CON 205 MILLONES DE TÉRMINOS DE TEXTOS DE ESTE SIGLO / DIARIO NOTICIAS DE ALAVA	7	1
12/03/13	EGUNGO TESTUEN CORPUSA, EUSKARAK EGUNGO ARAZOEI AURRE EGITEKO TRESNA BERRIA / GARA (EUSKAL HERRIKO EGUNKARIA)	8	1
12/03/13	EN PRIMER PLANO / EL CORREO ESPAÑOL (ALAVA)	9	1
12/03/13	EN PRIMER PLANO / EL CORREO ESPAÑOL (BIZKAIA)	10	1
12/03/13	TRES EN RAYA / DIARIO NOTICIAS DE ALAVA	11	1
12/03/13	EUSKARAREN INOIZKO CORPUSIK HANDIENA SORTU ETA GARATU DU EHUKO EUSKARA INSTITUTUAK / BERRIA	12	1
12/03/13	EL MAYOR CORPUS DEL EUSKERA / NOTICIAS DE GUIPUZKOA	13	1
12/03/13	NACE EL MAYOR CORPUS DE EUSKERA CON 205 MILLONES DE TÉRMINOS DE TEXTOS DE ESTE SIGLO / NOTICIAS DE GUIPUZKOA	14	1
12/03/13	LA UPV ABRE UN SERVICIO ON-LINE DE INFORMACIÓN SOBRE EL EUSKERA, CON 204 MILLONES DE PALABRAS / EL CORREO ESPAÑOL (BIZKAIA)	15	1
12/03/13	NACE EL MAYOR CORPUS DE EUSKERA CON 205 MILLONES DE TÉRMINOS DE TEXTOS DE ESTE SIGLO / DEIA	16	2
12/03/13	EGUNGO TESTUEN CORPUS ERRALDOI ETA FIDAGARRIA / GARA (EUSKAL HERRIKO EGUNKARIA)	18	1
12/03/13	EN PRIMER PLANO / EL DIARIO VASCO	19	1
12/03/13	205 MILLONES DE PALABRAS PARA BUCEAR EN EL EUSKERA DEL SIGLO XXI / EL DIARIO VASCO	20	1
12/03/13	IBON SARASOLA, INVESTIGADOR DEL INSTITUTO DE EUSKERA DE LA UPV/EHU: "KOLDO MITXELENA ESTARÍA ENCANTADO CON ESTE CORPUS" / NOTICIAS DE GUIPUZKOA	21	1



“
EHUren Egungo
Testuen Corpusa
euskarazko bilduma
handiena da, eta
munduko corpus
handienetakoa»
IÑAKI GOIRIZELAIA
EHUko errektorea



Egungo Testuen Corpora (ETC)

HIZKUNTZEN azterketak, egun, corpusak erabiltzea eskatzen du. Orain dela urte batzuk, ordenagailuak iritsi ziren arte, fitxak erabiltzen genituen hizkuntza aztertzeko orduan aipamenak egin nahi genituenean, erabilera bitxi edo arruntak nonbait gorde nahi genituenean edo lana adibide egoki batzuekin hornitu nahi genuenean. Gero ordenagailuak etorri ziren, eta horrek sekulako iraultza ekarri zuen testuak egiteko eta antolatze orduan, zuzenketak egitean, edo, besterik gabe, gure oharrak modu sistematiko batean ordenatu eta gorde nahi genituenean. Berehala etorri ziren programa berriak, eta aukera eman ziguten gure idazleek historian zehar egin testuak euskarri elektronikoen artean gorde, kontsultatu eta lantzeko. Azkueren garaian, edo Mitxelenarenean, eta geroago ere, adibide egoki bat bilatu nahi genuelarik eskuz edo makinaz egindako apunte eta fitxetan begiratu behar genuen, edo apalategiko liburuetan, edo gure buruaren zoko-zokoan:

Iritzia**PELLO SALABURU (*)**

“Hau hola esaten du halakok”. Eta auctoritas horretaz baliatzen ginen gure esana eta iritzia indartzeko: zer ote da egokiago? “joan behar dugu” edo “joan behar gara”? Eta nork bere oroimen sakonetan, bere euskalkian edo hiztunen batek ematen zion informazioan oinarritzen zuen erantzuna. Mitxelena eta Ibon Sarasola Orotariko Euskal Hiztegia lantzen hasi zirenean gauzak zeharo aldatu ziren. Han bazen testu bilduma bat handia (corpus bat), aski fidagarria, ordenagailuaren bidez kontsulta zitekeena. Buruak askotan huts egiten du: pantailan dituzun testuek ez. Hiz-

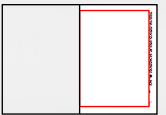
tegi hori egiteko 5 milioi hitzeko corpusa erabili zen orduan, eta han ematen ziren datu objektiboak aztertuz egin zen hiztegia. Aukeratu testuak tradiziokoak ziren, euskararen historiakoak, salbuespenak salbuespen. Hor biltzen da ezagutzen dugun euskara idatzia-ren historia.

Geroztik, nahiko arrunt bihurtu dira hizkuntzen corpusak: batzuk hizkuntzaren historiakoak, besteak hizkuntza modernoarenak; batzuk estatistikoak, beste batzuk terminologikoak; batzuek oinarritzko informazioa bakarrik ematen dutenak (hitza eta testuingurua), beste batzuk hitzak indexatuak ematen dituztenak (zer den hitz hori, non ager daitekeen, nolako funtzioa izan dezakeen, eta abar). Aski tresna baliagarriak dira horiek guztiak informazioa biltzeko, eta informazio hori erabilirik hizkuntzaren gainean esaten diren gauzak fundamentu handiagoarekin esateko.

Alde batera utzirik testu masa ikaragarriak dituzten corpus batzuk (Google-enak, esaterako), gure zereginetarako balio gehiegi ere ez dute-

nak, corpus erabilgarrienak milioi batzuen (badira 5 milio hitzekoak baino txikiagoak) eta 400 bat milioi hitzen tartekoak dira. Santiagoko Unibertsitatearen testu hispanikoen corpusak 3 milioi ditu: Euskaltzaindiaren Behatokiak 26,5 milioi; RAEk duen egungo espainolaren corpusak (CREA) 154 milioi, eta 1975etik aurrera argitaratutako testuak biltzen ditu. Elhuyarrek berriki aurkeztu du beste bat 125 milioi ingurukoa, Interneten dauden testuetan oinarritua. EHUko Euskara Institutuak (www.ei.ehu.es) aste honetan aurkeztu duen ETC corpusak 205 milio hitz inguru ditu, eta handienen artean kokatzen da, euskaraz dagoen handiena, duda gabe. Kontsultak egiteko interfazea oso erraza da erabiltzeko, eta aski intuitiboa, ez da gehiegi pentsatu beharrik. Eta informazio ugari ematen du egungo euskarari buruz, testu guztiak mende honetakoak baitira, urtez urte orekatuak eta hainbeste arlotakoak.

(*) Euskara Institutuaren zuzendaria



Precio: 1,30 Euros Núm. 12663

Gure lurraren deia

Martes 12 de marzo de 2013

2

Noticias de Bizkaia

Deia



205 MILLONES DE TÉRMINOS EN EUSKERA

El Instituto de Euskera de la UPV/EHU presenta y cuelga 'on line' el ETC, el mayor corpus de nuestra lengua jamás elaborado

PÁGINAS 14-15

El Parlamento navarro exige la dimisión de Yolanda Barcina

La Cámara le acusa de lucro irregular, como a Sanz, al exconsejero Miranda y al alcalde Maya

El Parlamento Foral de Navarra aprobó ayer una declaración en la que se exige la dimisión de la presidenta del Gobierno navarro, Yolanda Barcina, tras el escándalo de las dietas y sobresueldos percibidos en Caja Navarra y en empresas públicas por ella y otros miembros del Ejecutivo y de UPN. Un día después de que Barcina se viera obligada a anunciar que devolverá los más de

60.000 euros en dietas de la CAN y en vísperas del congreso en el que se juega su futuro y el de UPN, el Parlamento "reprueba la actuación del Gobierno de Navarra y de su presidenta, Yolanda Barcina, en la gestión de la CAN, declarando por ello la necesaria procedencia de su dimisión", según el texto aprobado por todas las fuerzas de la Cámara excepto UPN y PP. **P. 20-21** EDIT. EN **P. 3**

URKULLU PRESENTA LAS CUENTAS

 El lehendakari insiste en las 'líneas rojas' y resalta la prioridad educativa: mantener ratios y aulas

El invierno vuelve hoy a Euskadi y mañana traerá nieve a solo 200 metros

El convenio de 200.000 trabajadores vascos puede quedar extinguido en julio

Bilbao Basket, a 5 puntos de la semifinal europea: "Hoy no podemos fallar"

"Nos podían haber hecho cualquier cosa"

Drogan a cuatro jóvenes para robarles el dinero y el móvil



PÁGINA 8



nomina +  **k**

satisfaction

200 mini tablet*

zozketatuko ditugu

Zozketaren baldintzak bbk, vital, kutxa eta kutxabanken webguneetan.

*Samsung GALAXY P3110 16GB 7"

Gehien zoztuko zaituen nomina.

Kutxabank S.A., Kale Nagusia, 30-32, Bilbo, I.P. 495553077. Bizkaiko Merkataritza Erregistroan inbestratuta. S226. Tomoa 3. Lizentzia. BI-58727. Orria, 1. Folioa, 1. Inbestratza.

3



Presentación de Egungo Testuen Corpua, celebrada ayer en la sala Baroja del Paraninfo de la Universidad del País Vasco (UPV/EHU) situado en Abandoibarra. FOTO: PABLO VIÑAS

Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo

Elaborado por el Instituto de Euskera de la UPV/EHU está disponible on-line en www.ehu.es/etc

IDOIA ALONSO

BILBAO. La cultura vasca vivió ayer una jornada destacada. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó Egungo Testuen Corpua (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus on-line solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del *Diccionario General Vasco-Orotariko Euskal Hiztegia* de Euskaltzaindia, con 5 millones de entradas.

El nuevo corpus disecciona el euskera como nunca se había hecho hasta ahora, ofreciendo a los usuarios información completa de cada palabra, los textos y las frases en las que se inserta, en qué libro o artículo aparecen, de qué autor, permite comparar entre varios términos

el uso que se le da, la red semántica asociada o ver la evolución de su uso. Y todo sobre el euskera actual en base a textos de este siglo porque a partir del año 2000, fecha en que ve la luz *Hiztegi Batua*, el euskera se ha estabilizado.

La presentación de ETC contó con una nutrida representación de la vida política y académica de Euskadi encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron además el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriar-

te, el presidente de Euskaltzaindia así como distintos parlamentarios y docentes de la universidad.

El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar".

El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resultados son muy ricos", y entre sus múltiples posibilidades, ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos, lo cual "abre vías para analizar los motivos de esa evolución", apuntó Salaburu.

LENGUA, EL ALMA DEL SER HUMANO

Según explicaron, la consulta también se puede realizar teniendo en cuenta el origen del texto: si es original o una traducción. Lo mismo sucede respecto a la fuente. ETC informa dónde se utiliza más la palabra escogida: en prensa (y, en concreto, en qué medio), en literatura, ciencia, en las enciclopedias, en televisión... Así mismo, esta herramienta facilita la red semántica, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios,

que hoy en día se confeccionan utilizando información de este tipo", indicó Salaburu.

En su intervención, el lehendakari Urkullu destacó el valor del corpus y destacó la importancia del euskera en el orden de prioridades del Gobierno vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo". Para Urkullu el corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso".

Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de los mayores del mundo". Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".

"La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma, y la UPV ha creado un corpus del euskera"

"El euskera ha hecho un alto, se ha observado a sí mismo y se ve más fuerte y vigoroso"

IÑIGO URKULLU
Lehendakari

"Es la mayor colección que se ha hecho en euskera y ETC es uno de los mayores corpus del mundo"

"El corpus sistematiza el tremendo patrimonio lingüístico en euskera del presente siglo"

IÑAKI GOIRIZELAIA
Rector de la Universidad del País Vasco



Egungo Testuen Corpora, euskarak egungo arazoei aurre egiteko tresna berria

Euskara Institutuak garatutako Egungo Testuen Corpusak (ETC) 205 milioi hitz biltzen ditu, munduko handienetako bat eta euskaraz sekula egin den handiena delarik. Bere baitan

biltzen dituen XXI. mendeko testuetan agertzen diren hitzak interpretatzen ditu, bakoitzari buruz dagoen informazio osoa emateko. Tresna berri hau erabilerraza eta argia da.

Alvaro HILARIO | BILBO

Abandoibarran dagoen Bizkaia aretoan egin zuten atzo EHU/UPVko Euskara Institutuak azken urteotan garatu duen Egungo Testuen Corporaren (ETC) aurkezpena. 205 milioi hitz biltzen dituen corpus honek –munduko handienetako bat eta euskaraz sekula egin den handiena– bere baitan biltzen dituen XXI. mendeko testuetan agertzen diren hitz guztien gaineko informazioa ematen du. Egileetako bat da Ibon Sarasola, eta haren esanetan, ETCK «hizkuntza zehatz bat, une zehatz batean» ikasteko balio du.

Euskara Institutuko Pello Salaburu zuzendari eta Ibon Sarasola kidea ez ezik, Iñigo Urkullu lehendakariak eta EHUko errektore Iñaki Goirizelaia ere parte hartu zuten atzoko ekitaldian.

Euskaraz batuak dituen arazo berriei irtenbidea emateko jaiola da ETC: «Batuaren arazoak gero

eta sofistikatuagoak dira, eta beharrezkoak dira, beraz, hau bezalako tresna berriak», adierazi zuen Sarasolak.

Testuak eta hitzak

«Testuak dira. Corpusak interpretatzen ditu eta testuetan dauden hitzen inguruko informazioa ematen du. Ez dira liburu elektronikoak, bada beste ezer»; horrela definitu zuen ETCa Pello Salaburuk.

Segituan, eta ETC bera baliabidea, azaldu zituen Salaburuk tresna honen nondik norakoak: testu bilduma euskarri informatikoan dugu –«betiere, hizkerara mota baten adierazleak»– eta bertan diren hitzei buruzko informazioa ematen du; izan ere, hitz baten gaineko informazioa ematen du, hura hartzen duen esaldia kontuan hartuz.

Corpus arloan mota askotako corpusak daude eta «arrunten artean» kokatu zuen Euskara Institutuak eginiko hau, «erabi-

ERRONKA BERRIAK

Egungo Testuen Corpora euskarak dituen erronka berriei begira sortutako tresna da. Baliabide honen erabilerrari dagokionez, esan behar da hiztegiak, esaterako, corpus baten gainean direla eraikiak. Corpora www.ehu.es/etc helbidean kontsulta daiteke.

lerraza» baita eta «oso informazio interesgarria» zabaltzen duelako.

Neurriak, horiek zelan aukeratu, interfazea, informazioaren berdintasuna eta hitzei buruz zelako informazioa ematen den, item horien arabera sailkatzen dira corpus ezberdinak. Corpusen munduko mutur biak ikusarazteko, adibide bi erabili zituen Pello Salaburuk: batetik, Google Books, 500.000 milioi hitz biltzen dituen arren, haututako hitza azpimarratuta agertu baino egiten ez duena; eta bestetik, Kataluniako AnCora, milioi bat hitz dakartza eta horien guztien inguruko ezaugarriak eskaintzen ditu.

Azpimarratu zuen ETC «bilatze sistema argi eta erakargarria» dela, eta alde horretatik Josu Landak egindako lana eskertu zuen.

Aukeratutako testu guztiak XXI. mendekoak eta prosazkoak dira; egileen aburuz, hauex da

beste funtsezko ezaugarri bat: «Euskaraz asko egonkorta da 2000tik aurrera, urte horretan atera baitzuen Euskaltzaindiak 'Hiztegi batua', eta horrek eragin handia izan du», esan zuen Salaburuk. Ibon Sarasolak azaldu zuen hiztegiak corpus baten gainean daudela eraikita.

Testuak eta erabilerak

ETCn badira jatorriz euskaraz idatzitako testuak eta itzulpenak direnak. Corpusean bilaketak egitea erraza da, emaitzak «aberatsak» direlarik. «Etxea» hitza sartuz gero, lema (etxe) eta horren aldaera guztiak (etxeak, etxera...) agertzen dira. Bilaketa bukatu eta gero, automatikoki agertzen da «tarta» bat, lema eta aldaerak osatua, aldaerak zelan banatzen diren erakusten duena ehunekoetan. Aldaera bakoitzeko adibide guztiak ere kontsulta daitezke: hitza bere esaldian agertzen da, eta nor erabili duen, non, noiz eta abar ere ikus daitezke.

Ibon Sarasolak aipatu zuen hitzaren eboluzioa ere ikus daitekeela agertzen diren grafikoen: urte urte hitz bat gehiago edo gutxiago erabili den, urtekako ereduak ere azaltzen direlarik. Hitz bakoitzaren sare semantikoak ere eskaintzen ditu ETCk.

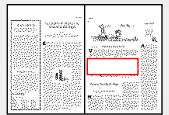
ETC eraikitzen ibili den Euskara Institutuko Ibon Sarasolak azpimarratu zuen jatorrizko testuetan eta itzulpenetan erabiltzen den euskaraz ezberdina dela. Bere ustez, kontu bi azaleratzen ditu honek: «euskaraz egiten dugunean arazoak saihesten» segitzen dugu; eta, bestalde, hizkuntza politika okerak egiten ari dira EAEn, zenbait itzulpen diruz lagunduz, esaterako.

Lagun Aro Fundazioaren diru-laguntza jaso du egitasmoak.

Ekzerretik hasita, Lagun Aroko zuzendaria, Pablo Mongelos; Iñaki Goirizelaia, EHUko errektorea; Iñigo Urkullu lehendakaria; Euskara Institutuko zuzendaria, Pello Salaburu; eta Ibon Sarasola, ETCren egileetako bat.

Mónica DEL VALLE | ARGAZKI PRESS





EN PRIMER PLANO

**JAVIER GÓMEZ
BERMÚDEZ**
JUEZ DE LA AUDIENCIA
NACIONAL



Afán de protagonismo. El magistrado admitió ayer a trámite la querrela presentada por IU por los 'papeles de Bárcenas' y ya ha citado a declarar a los 15 querellados. El magistrado pretende que la Fiscalía escuche a IU antes de sentar su criterio

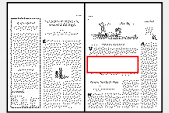
sobre cuál ha de ser el juzgado competente en el caso, ya que el titular del Juzgado Central número 5, Pablo Ruz, que instruía el 'caso Gürtel', ha acumulado el 'caso Bárcenas'. Una iniciativa que trasluce un afán de protagonismo algo inoportuno.

**PELLO
SALABURU**
DIRECTOR DEL INSTITUTO
DEL EUSKERA DE LA UPV



Nuevo corpus. Tanto los estudiosos del euskera como los propios hablantes interesados en profundizar en el conocimiento de la lengua vasca tienen una nueva herramienta a su disposición, el Egungo Testuen Corpusa (ETC) elaborado por el

Instituto de Euskera de la UPV/EHU. A diferencia de los diccionarios, los corpus recopilan cantidad de usos reales de cada término, permiten compararlos y trazar su evolución. Como instrumento del siglo XXI que es, el ETC es de libre consulta a través de Internet.



EN PRIMER PLANO

JAVIER GÓMEZ BERMÚDEZ

JUEZ DE LA AUDIENCIA NACIONAL



Afán de protagonismo. El magistrado admitió ayer a trámite la querrela presentada por IU por los 'papeles de Bárcenas' y ya ha citado a declarar a los 15 querellados. El magistrado pretende que la Fiscalía escuche a IU antes de sentar su criterio

sobre cuál ha de ser el juzgado competente en el caso, ya que el titular del Juzgado Central número 5, Pablo Ruz, que instruía el 'caso Gürtel', ha acumulado el 'caso Bárcenas'. Una iniciativa que trasluce un afán de protagonismo algo inoportuno.

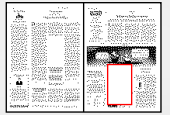
PELLO SALABURU

DIRECTOR DEL INSTITUTO DEL EUSKERA DE LA UPV



Nuevo corpus. Tanto los estudiosos del euskera como los propios hablantes interesados en profundizar en el conocimiento de la lengua vasca tienen una nueva herramienta a su disposición, el Egungo Testuen Corpusa (ETC) elaborado por el

Instituto de Euskera de la UPV/EHU. A diferencia de los diccionarios, los corpus recopilan cantidad de usos reales de cada término, permiten compararlos y trazar su evolución. Como instrumento del siglo XXI que es, el ETC es de libre consulta a través de Internet.



TRES EN RAYA



Ibon Sarasola

Escritor y lexicógrafo

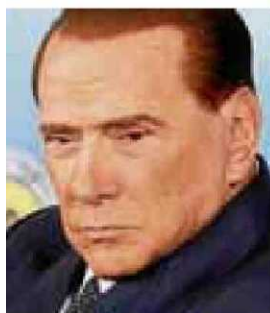
+ Presentó ayer 'Egungo Testuen Corpusa', un proyecto que culmina el ingente trabajo iniciado por Koldo Mitxelena hace tres décadas en la recopilación sistemática de 205 millones de términos en euskera y que encarrilla a la lengua vasca en el siglo XXI en una privilegiada posición.



Iñaki Goirizelaia

Rector de la UPV/EHU

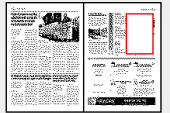
= En representación de toda la UPV, el rector vasco no dudó ayer en salir en defensa de Xabier Aierdi y Enrique Antolín, dos profesores de intachable trayectoria académica y democrática juzgados desde ayer por impartir clase a dos deportados de ETA en un absurdo proceso inquisitorial.



Silvio Berlusconi

Exprimer ministro italiano

- La petición del fiscal de Nápoles de dar inicio inmediato al juicio contra Berlusconi por la presunta compra de votos en el Senado para hacer caer el gobierno de Romano Prodi tiene esta vez una sólida base por la declaración de un senador que admite haber recibido dinero de 'Il Cavaliere'.



Egungo Testuen Corpusaren aurkezpena, atzo. MONIKA DEL VALLE / ARGAZKI PRESS

Euskararen inoizko corpusik handiena sortu eta garatu du EHUko Euskara Institutuak

Erredakzioa Bilbo

Orain arteko euskararen corpusik handiena garatu du EHU Euskal Herriko Unibertsitateko Euskara Institutuak. Egungo Testuen Corpusean bildu du. Interneten egin dute — www.ehu.es/etc kontsulta daiteke—, eta guztira 205 milioi hitz biltzen ditu. Euskara Institutuak garatutako aurreko corpusak, Ereduko Prosa Gaur izenekoak, 25 milioi zituen. Beraz, 40 aldiz handiagoa da. «Oso sinplea, argia eta intuitiboa da; erabilerraza da», nabarmendu du Pello Salaburuk, EHUko Euskara Institutuko zuzendariak.

Hitz bakoitzaren informazioa ematen du, eta baita haren aldaerak ere. Aldaera bakoitzaren adibide guztiak kontsultatzeko modua ere ematen du. Hitzaren eboluzioari buruzko grafiko bat kontsulta daiteke, eta baita haren sare semantikoa ere.



Año IX Núm. 2.643 1,30 euros

GUIPUZKOAR GUZTION EGUNKARIA

Martes, 12 de marzo de 2013

Noticias de

MONTANIER

Sigue a la espera de recibir una oferta del club para renovar



Gipuzkoa

PÁGINAS 44-45



NUEVO PAPA
COMIENZA LA CUENTA ATRÁS

PÁGINA 14-15

Más de 200.000 trabajadores vascos pueden perder sus convenios en julio

Casi 120 acuerdos sectoriales se extinguirán si no hay pacto entre patronal y sindicatos

DONOSTIA. Euskadi se enfrenta en los tres próximos meses a una encrucijada en la negociación colectiva. Más de 200.000 trabajadores pueden perder su convenio si patronal y sin-

dicatos no alcanzan un pacto antes del 7 de julio, fecha en la que finaliza la vigencia de la ultraactividad de los acuerdos tras la reforma laboral aprobada por el Gobierno del PP. Un

total de 117 convenios sectoriales siguen sin renovarse, lo que pone en peligro el mantenimiento de las condiciones laborales de los trabajadores. PÁGINAS 38-39



18. Korrikako antolatzaileak, atzo Andoainen. ARGAZKIA: R. PLAZA

KORRIKA BADATOR!

13. ORRIALDEA

EL MAYOR CORPUS DEL EUSKERA

El Instituto de Euskera de la UPV/EHU recoge más de 250 millones de términos de texto de este siglo

PÁGINA 12

Alcohol y estética, un peligroso cóctel en auge

CADA VEZ MÁS JÓVENES SE EMBORRACHAN SIN INGERIR ALIMENTOS PARA NO ENGORDAR

PÁGINAS 6-7

COSTAS AUTORIZA CREAR DUNAS EN ORIO, DEBA Y HONDARRIBIA

DONOSTIA. El Servicio Provincial de Costas de Gipuzkoa, dependiente del Gobierno central, ha autorizado a la Diputación la ocupación de bienes para llevar a cabo un proyecto de restauración de dunas en las playas de Orío, Hondarribia y Deba. El Ejecutivo foral dispone de tres meses, con el fin de que los trabajos concluyan antes del verano. PÁGINA 11

El Parlamento navarro exige la dimisión de Barcina

La Cámara le acusa de lucro irregular y afirma que "no está habilitada" para ser presidenta

PÁGINAS 16-17 EDITORIAL EN PÁG. 3

CUCHILLOS SANTOMU
cupón 9

ROBOT ASPIRADOR MAURUS
cupón 2



Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo

HA SIDO ELABORADO POR EL INSTITUTO DE EUSKERA DE LA UPV/EHU

La consulta se puede realizar según el origen del texto o la fuente

IDOIA ALONSO

BILBAO. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó ayer Egungo Testuen Corpusa (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus *online* solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del *Diccionario General Vasco-Orotariko Euskal Hiztegia* de Euskaltzaindia, con cinco millones de entradas.

El nuevo corpus disecciona el euskera como nunca se había hecho hasta ahora, ofreciendo a los usuarios información completa de cada palabra, los textos y las frases en las que se inserta, en qué libro o artículo aparecen, de qué autor, permite comparar entre varios términos el uso que se le da, la red semántica asociada o ver la evolución de su uso. Y todo sobre el euskera actual en la base a textos de este siglo porque a partir del año 2000, fecha en la que ve la luz *Hiztegi Batua*, el euskera se ha estabilizado.

La presentación de ETC contó con una nutrida representación de la vida política y académica de Euskadi que estuvo encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron, además, el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriarte, el presidente de Euskaltzaindia, así como distintos parlamentarios y docentes de la universidad.

El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrece sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar".

El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resultados son muy ricos". Entre las múltiples posibilidades ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando



Presentación de Egungo Testuen Corpusa, ayer, en Leioa. FOTO: PABLO VIÑAS

más o menos, lo cual "abre vías para analizar los motivos de esa evolución".

LENGUA, EL ALMA DEL SER HUMANO
 Según explicaron, la consulta también se puede realizar teniendo en cuenta el origen del texto: si es original o una traducción. Lo mismo

sucede respecto a la fuente. ETC informa dónde se utiliza más la palabra escogida: en prensa (y, en concreto, en qué medio), en literatura, ciencia, en las enciclopedias, en televisión... Asimismo, esta herramienta facilita la red semántica, la cadena formada por otras palabras que tienen las mismas

propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios, que hoy en día se confeccionan utilizando información de este tipo", indicó Salaburu.

En su intervención, el lehendakari Urkullu destacó el valor del corpus y la importancia del euske-

ra en el orden de prioridades del Gobierno Vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así nos sentimos hoy todos, reconfortados con este trabajo". Para Urkullu este corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso".

Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho en euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de los mayores del mundo".

Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".



La UPV abre un servicio on-line de información sobre el euskera, con 204 millones de palabras



MARTA
FDEZ. VALLEJO

mfernandez@elcorreo.com

Es el «corpus» más grande que «jamás se haya elaborado en lengua vasca», y ofrece información «exhaustiva de cada término»

BILBAO. La Universidad del País Vasco presentó ayer el corpus «más grande jamás elaborado en euskera y uno de los más extensos que hay en el mundo». El trabajo realizado por el Instituto de Euskera de la UPV/EHU, en formato on-line y abierto a cualquier usuario, cuenta con 204,9 millones de palabras. La ventaja de 'Egungo Testuen Corpusa' (ETC) es que ofrece una información exhaustiva de cada término y, además, las búsquedas en la página web resultan «sencillos y muy intuitivos», según destacó Pello Salaburu, director del Instituto de Euskera de la institución académica.

No se trata de un diccionario sino de un 'corpus'. Salaburu lo describió como «una masa de textos que dan información sobre el contenido de las palabras. No son libros electrónicos», aclaró. En la práctica, es una herramienta de consulta de la lengua. Ofrece datos de una palabra y de todas sus variantes. También se pueden hacer investigaciones más especializadas: introducir el término concreto inserto en una frase para determinar quién es el autor, cuándo y dónde aparece, buscar conjuntos de palabras, comparar entre varias el uso que se les da,



El lehendakari, Iñigo Urkullu, junto a Pello Salaburu, el rector Goirizelaia y el director general de Lagun Aro, Pablo Mongelos. :: F. GÓMEZ

recopilar las que empiezan de determinada manera y un largo etcétera de posibilidades.

Además permite observar la evolución del uso de un término mostrando año a año si se va utilizando más o menos, lo que facilita abrir vías para analizar los motivos. Todos los textos que incluye el corpus son en prosa y sacados de libros, prensa y de la Wikipedia; escogidos por su calidad pero actuales, del siglo XXI. «El euskera se ha estabilizado mucho a partir de 2000, pues ese año publicó

Euskaltzaindia 'Hiztegi Batua' y eso ha tenido gran influencia», añadió el catedrático.

«Futuro fuerte y vigoroso»

Salaburu definió el trabajo como el corpus «más grande elaborado nunca» en euskera. Recordó que otro realizado con anterioridad por la UPV, 'Ereduzko prosa Gaur', reúne 25 millones de palabras. Y, por ejemplo, para crear 'Orotariko Euskal Hiztegia' de Euskaltzaindia se utilizó uno de cinco millones de términos. Situó a ETC en-

tre «los más grandes» del mundo. «Y aunque los haya mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable», subrayó.

Egungo Testuen Corpusa fue presentado ayer en Bilbao en un acto presidido por el lehendakari, en el Paraninfo de la UPV. Iñigo Urkullu destacó que este trabajo, que se ha prolongado durante cuatro años, «recoge el pasado y proyecta el futuro fuerte, vigoroso y completo del euskera». Reiteró, además, la apuesta de su equipo

por la lengua vasca «como una prioridad educativa». En el acto participaron también el rector de la Universidad, Iñaki Goirizelaia, y la consejera vasca de Educación, Política Lingüística y Cultura, Cristina Uriarte, así como la diputada foral vizcaína de Cultura, Josune Ariztondo, y el investigador del Instituto de Euskera de la UPV, Ibon Sarasola.

Web de Egungo Testuen Corpusa:
www.ehu.es/etc



PATRIMONIO LINGÜÍSTICO >



Presentación de Egungo Testuen Corpusa, celebrada ayer en la sala Baroja del Paraninfo de la Universidad del País Vasco (UPV/EHU) situado en Abandoibarra. FOTO: PABLO VIÑAS

Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo

Elaborado por el Instituto de Euskera de la UPV/EHU está disponible 'on line' en www.ehu.es/etc

IDOIA ALONSO

BILBAO. La cultura vasca vivió ayer una jornada destacada. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó Egungo Testuen Corpusa (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus *on line* solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del *Diccionario General Vasco-Orotariko Euskal Hiztegia* de Euskaltzaindia, con 5 millones de entradas.

El nuevo corpus disecciona el euskera como nunca se había hecho hasta ahora, ofreciendo a los usuarios información completa de cada palabra, los textos y las frases en las que se inserta, de qué autor, permite comparar entre varios términos el uso que se le da, la red semántica asociada o ver la evolución de su uso. Y todo sobre el euskera actual en base a textos de este siglo porque a partir del año 2000, fecha en que ve la luz *Hiztegi Batua*, el euskera se ha estabilizado.

La presentación de ETC contó con una nutrida representación de la vida política y académica de Euska-

di encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron además el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriarte, el presidente de Euskaltzaindia así como distintos parlamentarios y docentes de la universidad.

El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar".

El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resulta-

"La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma, y la UPV ha creado un corpus del euskera"

"El euskera ha hecho un alto, se ha observado a sí mismo y se ve más fuerte y vigoroso"

"Es la mayor colección que se ha hecho en euskera y ETC es uno de los mayores corpus del mundo"

"El corpus sistematiza el tremendo patrimonio lingüístico en euskera del presente siglo"

IÑAKI GOIRIZELAIA
Rector de la Universidad del País Vasco

dos son muy ricos". Entre sus múltiples posibilidades, ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos, lo cual "abre vías para analizar los motivos de esa evolución", apuntó Salaburu.

LENGUA, EL ALMA DEL SER HUMANO

Según explicaron, la consulta también se puede realizar teniendo en cuenta el origen del texto: si es original o una traducción. Lo mismo sucede respecto a la fuente. ETC informa dónde se utiliza más la palabra escogida: en prensa (y, en concreto, en qué medio), en literatura, ciencia, en las enciclopedias, en televisión... Así mismo, esta herramienta facilita la red semántica, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios, que hoy en día se confeccionan utilizando información de este tipo", indicó Salaburu.

En su intervención, el lehendakari Urkullu destacó el valor del corpus y también la importancia del euskera en el orden de prioridades del Gobierno vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha

creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo". Para Urkullu, el corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso".

Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho en euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de los mayores del mundo". Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".



PATRIMONIO LINGÜÍSTICO

IBON SARASOLA

INVESTIGADOR DEL INSTITUTO DE EUSKERA DE LA UPV/EHU

“Koldo Mitxelena estaría encantado con este corpus”

Lingüista, miembro de Euskaltzaindia e investigador, Ibon Sarasola presenta el fruto de varios años de trabajo, Egungo Testuen Corpusa, la continuación natural de 'Orotariko Euskal Hiztegia' pasado por el filtro de las TIC. El resultado es un corpus de 205 millones de términos

I. ALONSO

BILBAO. Sarasola cambió la Ingeniería Industrial por las palabras. Discípulo de Koldo Mitxelena, el Premio Euskadi de Investigación de 2011 introduce en este siglo la compilación del amplio patrimonio lexicográfico del euskera sobre textos producidos desde 2000 en un corpus al alcance de todos.

¿Qué hace un ingeniero de formación estudiando las palabras?

Porque creo que hoy en día para estudiar las palabras hay que tener formación de ingeniero.

¿En qué se diferencia un diccionario de este corpus que acaban de presentar?

Los diccionarios han cambiado mucho los últimos 50 años. Antes un diccionario daba el significado de una palabra. Hoy en día, en cambio, se tiende a dar el significado de la

palabra pero en el contexto en el que se usa. Por ejemplo, en español estallar y explotar no es lo mismo ya que las guerras estallan, no explotan, y para alguien que está aprendiendo español esa información es decisiva, más que la definición de guerra que seguramente ya conocerá. Un corpus te proporciona la palabra en su contexto, en la frase en la que vive. Entonces se recogen todas las apariciones de la palabra en distintos textos y haces un diccionario con los usos más interesantes. Hoy en día, detrás de todo diccionario que se precie hay un corpus del idioma.

Quizá a la gente de a pie su corpus de euskera le puede resultar algo lejano ¿Por qué es importante?

El corpus que acabamos de presentar nos permite conocer el euskera actual porque es un corpus de los últimos diez años. Es un corpus que recoge qué palabras en euskera se han usado en los textos y cuáles no, lo cual ayuda a ver la evolución que ha tenido una palabra, qué términos que se usaban antes ya no se usan, que nuevas palabras se han empezado a usar en la última década y todo ello en qué contextos, con qué otras palabras suelen relacionarse.

Usted sostiene que una palabra es más que semántica.

Es el conjunto de contextos en el que puede aparecer. El corpus compara el uso de las palabras en distintos tex-

tos, indicando tendencias, lo cual es muy importante en el caso de que una academia de la lengua quiera legislar sobre alguna palabra ya que con el corpus puede hacerlo en base a unas razones concretas. Dicho de otro modo, el corpus es como poner el euskera que se habla ahora bajo el microscopio. Si hubiera una guerra entre palabras ¿cuál elegir?. ¿La que tenga mayor tradición?, ¿la que empleen los mejores escritores? En todo esto ayuda un corpus ya que registra cada aparición de una palabra, en qué contexto, en qué libro, de qué autor, en qué página o artículo periodístico...

El conocimiento del euskera se ha disparado en 30 años gracias a la escuela. ¿Este proceso ha afectado al idioma?

Desde luego que sí, aunque no a su estructura. Yo creo que hoy

en día se escribe mejor que hace 20 años. Habrá un momento en el que la proporción de personas con el euskera como lengua materna y las que lo hayan aprendido en la escuela cambiará, y con un corpus podremos saber si los *euskaldunberris* escriben de una manera y los *euskaldunzaharras* de otra. Pero de momento, no se observa esa diferencia.

¿Qué pensaría Mitxelena de este corpus 'on line' tan moderno?

Estaría encantado porque su diccionario *Orotariko Euskal Hiztegia* se basa en el primer corpus que se hizo para la lengua vasca, y este es una continuación, es más de lo mismo, así que estaría encantado.



El profesor Sarasola en el Parainfante de la UPV/EHU. FOTO: PV.

Urkullu se compromete a mantener ratios y no cerrar aulas

El lehendakari insiste en que la educación “es una de las prioridades” del Gobierno

BILBAO. “Nos comprometemos a mantener el ratio de alumnos por profesor, a no cerrar ningún aula y a avanzar en el trilingüismo”. El lehendakari, Iñigo Urkullu, realizó estas manifestaciones apenas veinticuatro horas antes de que el Consejo de Gobierno apruebe hoy el anteproyecto de los Presupuestos de la CAV para el ejercicio 2013, que contarán con 1.200 millones de euros menos debido a la caída de ingresos. Tras reconocer que el Gobierno vasco dispone de “menos dinero” para la confección de las Cuentas, Urkullu reiteró que la educación es “una de las prioridades” del Gobierno vasco.

Hoy se pone negro sobre blanco el alcance de los ajustes en materia educativa que distintos partidos sitúan en el 7%, es decir, un recorte de 185 millones de euros. Con sus palabras, el lehendakari reafirmaba ayer en Bilbao lo que la consejera de Educación, Cristina Uriarte, ya ha avanzado en el Parlamento Vasco, que su departamento se adaptará al nuevo contexto económico sin dañar la calidad del sistema educativo manteniendo para ello “lo esencial”, entendiendo como tal, las condiciones del alumnado y el profesorado.

Urkullu también tuvo palabras para la reforma educativa que pretende aprobar el Gobierno español. Y advirtió de que su Gobierno “no está dispuesto a perder el tiempo en batallas estériles” con quienes quieren cambiar el modelo de educación y “condicionar” a Euskadi, donde hay “un pacto educativo sólido, que funciona y ha dado muy buenos resultados durante más de dos décadas”. No obstante, afirmó que el Gobierno vasco “siempre está dispuesto a mejorar y a recibir propuestas para avanzar”. >I.A.



El lehendakari Urkullu. FOTO: VIÑAS



Egungo Testuen Corpus erraldoi eta fidagarria

Euskarak, euskaldunek, beste tresna bat dute eskura, eta ez edonolakoa. EHUko Euskara Institutuak garatutako Egungo Testuen Corpora (ETC) tresna moderno eta erabilerraza da, hainbat arazori irtenbidea emateko modukoa. Eta zabala da, inoiz bildu den corpusik handiena baita. ETCk hainbat liburu eta hedabidetatik eta Wikipediatik hartutako prosazko testuak biltzen ditu, denak ere XXI. mendekoak eta kalitategatik hautatuak, baina baita Euskara Institutuak lehenago osatutako corpusetakoak ere, esate baterako Ereduzko Prosa Gaur izenekoan jasotako 25 milioi hitzak; hori ere ez da corpus ziztrina, hain justu, baina hala eman lezake ET-

Cren aldean, honek 205 milioi hitz ditu-eta. Eta haren tamainak garrantzi handia badu ere, ETCren beste ezaugarri batzuk ere nabarmendu beharra dago. Esate baterako, hitz bakoitzaren gainean ematen duen informazioa eta aukera ugariak: sartutako lema eta haren aldaerak agertuko dira, hala nola horien erabilera desberdinak ikusteko aukera, hitza edo hitzaren aldaera ez ezik, esaldi osoa ere agertzen du-eta. Ibon Sarasolak, Corpusaren egileetakoak, atzo ETCren aurkezpenean esan zuenez, euskarak gero eta arazo sofistikatuagoak ditu; beraz, arazo horiei irtenbidea emateko tresna sofistikatuak behar dira. ETC halakoa da eta, batez ere, erabilgarri eta «fidagarria».



EN PRIMER PLANO

**PELLO
SALABURU**
DIRECTOR DEL INSTITUTO
DEL EUSKERA DE LA UPV



Nuevo corpus. Tanto los estudiosos del euskera como los propios hablantes interesados en profundizar en el conocimiento de la lengua vasca tienen una nueva herramienta a su disposición, el Egungo Testuen Corpusa (ETC) elaborado por el

Instituto de Euskera de la UPV/EHU. A diferencia de los diccionarios, los corpus recopilan cantidad de usos reales de cada término, permiten compararlos y trazar su evolución. Como instrumento del siglo XXI que es, el ETC es de libre consulta a través de internet.

LUCIA LACARRA
BAILARINA



Una guipuzcoana en el Real. Vuelve al Teatro Real de Madrid trece años después. Y lo hace de la mano de la Compañía Nacional de Danza y con su marido, Marlon Dino, de pareja. La zumaiarra Lucia Lacarra, bailarina principal del prestigioso

Ballet de la Ópera de Munich, se siente feliz ante la idea pisar de nuevo el escenario madrileño donde ha bailado solo una vez. Considerada una de las principales figuras de la danza a nivel internacional, Lacarra confiesa que se encuentra en su mejor momento.



Pello Salaburu, director del Instituto de Euskera de la UPV/EHU, presentó ayer el Corpus en Bilbao en compañía de Pablo Mongelos (Lagun Aro), el rector Iñaki Goirizelaia, el lehendakari Iñigo Urkullu e Ibon Sarasola, profesor de la UPV/EHU e investigador del Instituto.

205 millones de palabras para bucear en el euskera del siglo XXI

'Egungo Testuen Corpusa', elaborado por el Instituto de Euskera de la UPV/EHU, es el mayor que se ha compilado en euskera y uno de los más amplios del mundo

■ N. AZURMENDI

SAN SEBASTIÁN. Un corpus es, en la acepción más sencilla y genérica del término, un conjunto ordenado de datos. 'Egungo Euskararen Corpusa (ETC)', el corpus del euskera actual que ha desarrollado el Instituto de Euskera de la UPV/EHU y ayer se presentó en Bilbao, es la mayor base de datos léxicos en euskera jamás elaborada, así como una de las más amplias del mundo.

De hecho, con 205 millones de palabras, este nuevo corpus del euskera puede mirar de tú a tú al Corpus de Referencia del Español Actual (CREA) de la Real Academia Española, que cuenta con algo más de 150 millones de formas procedentes de textos fechados a partir de 1975, y tampoco tiene muy lejos los cerca de 300 millones de formas desde los orígenes del idioma hasta 1974.

Si ETC se compara con los corpus anteriormente existentes en euskera las diferencias se disparan. Tal como recordaron ayer en la presentación, 'Ereduzko Prosa Gaur', un corpus desarrollado anteriormente por el Instituto de Euskera, abarcaba 25 millones de palabras y para elaborar 'Orotariko Euskal Hizte-

gia', de Euskaltzaindia, el trabajo más exhaustivo de análisis del euskera escrito que se ha llevado a cabo, se utilizó un corpus de 5 millones de palabras.

Una de las principales peculiaridades de este nuevo corpus, además de su extensión, es el hecho de que todos los textos que lo conforman son del siglo XXI. «El euskera se ha estabilizado mucho a partir de 2000, pues ese año publicó Euskaltzaindia 'Hiztegi Batua' y eso ha tenido gran influencia», indicó el director del Instituto de Euskera, Pello Salaburu.

En cuanto a las fuentes de las que proceden los textos, son libros, artículos de prensa y textos de Wikipedia, que proceden tanto de Iparalde como Hegoalde y han sido elegidos por su calidad, aunque los ejemplos no tienen carácter normativo. Los hay escritos originalmente en euskera, y también hay traducciones de otras lenguas, y a la hora de seleccionar los que se iban a incorporar al corpus «se han tomado en cuenta las fuentes y años de creación, para que tengan un peso similar».

Sencillo e intuitivo

Pero todo es cuestión de cantidad, porque incluso en el caso de compilaciones mayores, que las hay, la información que ofrecen sobre cada palabra es más básica y menos manejable que la que proporciona ETC. «Nuestra interfaz es muy simple, clara e intuitiva, fácil de usar», explicó Salaburu. A la hora de referir-

DATOS

► **Contenidos.** 205 millones de palabras, incluidas sus variantes, sobre las cuales se ofrece abundante información, ilustrada siempre con ejemplos reales.

► **Fuentes.** Los textos, tanto originales como traducidos y posteriores siempre al año 2000, proceden de libros, prensa y Wikipedia.

► **Accesibilidad.** Accesible en www.ehu.es/etc

Hay corpus mayores, pero la información que ofrecen sobre cada palabra es más básica y menos manejable

se a la sencillez de uso de ese recurso que se ha elaborado con la ayuda de la Fundación Lagun Aro destacó que «hemos analizado los sistemas que se utilizan en el mundo y creo que con el trabajo que ha realizado el técnico del Instituto de Euskera Josu Landa el nuestro ofrece muchas mejoras». La principal, conseguir que el corpus –disponible junto con otros muchos materiales muy útiles en la web del Instituto– se pueda consultar «de modo simple e intuitivo».

Con respecto a la cantidad de in-

formación que ofrece, uno de los puntos fuertes del nuevo corpus en palabras de Pello Salaburu, es realmente amplia y muy completa.

Ofrece información sobre cada una de las 205 millones de palabras, «dando la posibilidad de distinguir entre lema (etxe) y todas sus variantes (etxea, etxeak, etxera, etxetik...)». Información, además, muy variada. Por ejemplo, una 'tarta' formada por el lema y sus variantes, que muestra en porcentaje cómo se reparten estas últimas. En el caso del ejemplo elegido (etxe), la variante con más apariciones (173.380, el 31,96% del total) es el indeterminado 'etxek'.

Pero ahí no acaban las opciones de explotación de la extensa base de datos. Además de facilitar información acerca de la frecuencia de aparición de las palabras cada año, se diferencian los textos originales de las traducciones, se clasifican por fuentes, se presentan casi incontables combinaciones con otros lemas...

Se pueden hacer también consultas sobre todos los ejemplos de cada variante: la palabra aparece en la frase en la que está inserta, y se indica quién es el autor, cuándo y dónde aparece... Asimismo, permite examinar las palabras que empiezan de determinada manera; comparar entre varios términos el uso que se les da, consultar conjuntos de palabras. En suma, una amplia gama de posibilidades para bucear en el euskera del siglo XXI por razones profesionales o académicas, por profundizar en el conocimiento de la lengua o por simple curiosidad.



IBON SARASOLA

INVESTIGADOR DEL INSTITUTO DE EUSKERA DE LA UPV/EHU

“Koldo Mitxelena estaría encantado con este corpus”

Lingüista, miembro de Euskaltzaindia e investigador, Ibon Sarasola presenta el fruto de varios años de trabajo, Egungo Testuen Corpusa, la continuación natural de 'Orotariko Euskal Hiztegia' pasado por el filtro de las TIC. El resultado es un corpus de 205 millones de términos

I.ALONSO

BILBAO. Sarasola cambió la Ingeniería Industrial por las palabras. Discípulo de Koldo Mitxelena, el Premio Euskadi de Investigación de 2011 introduce en el siglo XXI la compilación del amplio patrimonio lexicográfico del euskera en un corpus al alcance de todos.

¿Qué hace un ingeniero de formación estudiando las palabras?

Porque creo que hoy en día para estudiar las palabras hay que tener formación de ingeniero.

Hasta ahora ha participado en la elaboración de varios diccionarios de euskera. ¿En qué se diferencia un diccionario de este corpus que acaban de presentar?

Los diccionarios han cambiado mucho en su estructura los últimos 50 años. Antes un diccionario daba el significado de una palabra. Hoy

en día, en cambio, se tiende a dar, sí el significado de la palabra, pero en el contexto en el que se usa. Por ejemplo, en español estallar y explotar no es lo mismo, ya que las guerras estallan, no explotan, y para alguien que está aprendiendo español esa información es decisiva, más que la definición de guerra, que seguramente ya la conocerá. Un corpus te proporciona la palabra en su contexto, en la frase en la que vive. Entonces se recogen todas esas apariciones de la palabra y haces un diccionario recogiendo los usos más interesantes. Hoy en día, detrás de todo diccionario que se precie hay un corpus del idioma.

A la gente de a pie el corpus del euskera le puede sonar muy lejano o puede sentirse ajeno a él.

El corpus que acabamos de presentar nos permite conocer el euskera actual, porque es un corpus de los últimos diez años. Es un corpus que recoge qué palabras en euskera se han usado en los textos y cuáles no, lo cual ayuda a ver la evolución que ha tenido una palabra, qué términos que se usaban antes ya no se usan, qué nuevas palabras se han empezado a usar en la última década y todo ello en qué contextos, con qué otras palabras suelen relacionarse. El corpus compara el uso de las palabras en distintos textos, indicando tendencias, lo cual es muy importante

en el caso de que una academia de la lengua quiera legislar sobre alguna palabra, para que lo haga en base a unas razones concretas. Dicho de otro modo, el corpus es como poner la lengua que se habla ahora bajo el microscopio. Si hubiera una guerra entre palabras, ¿cuál elegir?, ¿la que tenga mayor tradición?, ¿la que la empleen los mejores escritores? En todo esto ayuda un corpus, ya que registra cada aparición de una palabra, en qué contexto, en qué libro, de qué autor, en qué página o artículo periodístico.

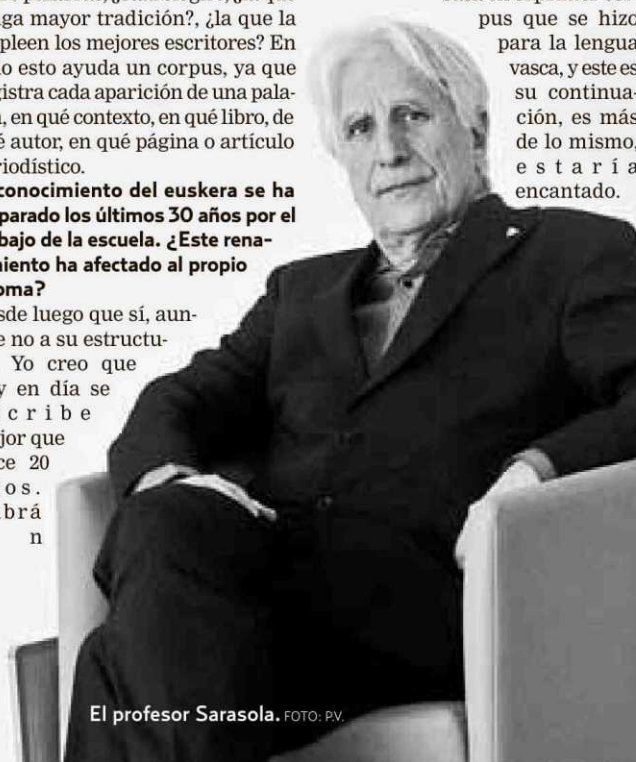
El conocimiento del euskera se ha disparado los últimos 30 años por el trabajo de la escuela. ¿Este renacimiento ha afectado al propio idioma?

Desde luego que sí, aunque no a su estructura. Yo creo que hoy en día se escribe mejor que hace 20 años. Habrá un

momento en el que la proporción de personas con el euskera como lengua materna y las que lo hayan aprendido en la escuela cambiará, y con un corpus podremos saber si los euskaldunberris escriben de una manera y los euskaldunzaharra de otra. Pero de momento, no parece que se haya producido esta diferencia.

¿Qué pensaría Koldo Mitxelena de este corpus del euskera 'online' tan moderno?

Estaría encantado porque su diccionario *Orotariko Euskal Hiztegia* se basa en el primer corpus que se hizo para la lengua vasca, y este es su continuación, es más de lo mismo, estaría encantado.



El profesor Sarasola. FOTO: PV.



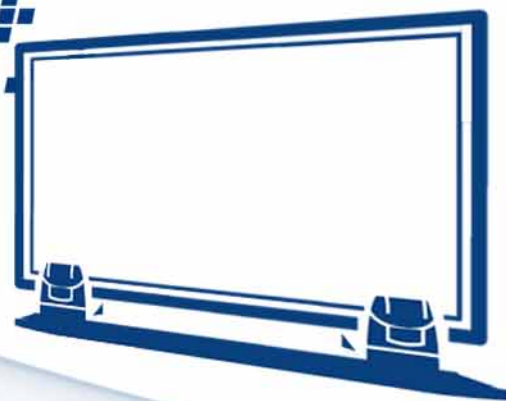
RADIO

SELECCIÓN DE LAS NOTICIAS DE SU INTERÉS



Universidad
del País Vasco Euskal Herriko
Unibertsitatea

Fecha	Titular/Medio
12/03/13	ONDA VASCA - EUSKADI HOY - 07:49h - 00:01:00 *EUSKADI. LA UNIVERSIDAD DEL PAIS VASCO HA PRESENTADO EL 'CORPUS DE EUSKERA' MAS GRANDE JAMAS ELABORADO CON TEXTOS DE ESTE SIGLO. DECLARACIONES DE PELLO SALABURU, DIRECTOR DEL INSTITUO DE EUSKERA.
11/03/13	RADIO EUSKADI - BOULEVARD - 10:09h - 00:00:28 LA UNIVERSIDAD DEL PAIS VASCO PRESENTA EL CORPUS ACTUALIZADO DE LOS TEXTOS EN EUSKERA.



TELEVISIÓN

SELECCIÓN DE LAS NOTICIAS DE SU INTERÉS



Universidad
del País Vasco Euskal Herriko
Unibertsitatea

Fecha	Titular/Medio
11/03/13	EUSKAL TELEBISTA 2 - TELEBERRI 2 - 21:41h - 00:01:03 EL INSTITUTO DE EUSKERA DE LA UPV/EHU HA ELABORADO, BAJO EL TITULO EGUNGO TESTUEN CORPUSA, LA MAYOR COLECCION DE TEXTOS EN EUSKERA JAMAS REALIZADA. DECLARACIONES DE PELLO SALABURU, DIRECTOR DEL INSTITUTO DE EUSKERA.
11/03/13	EUSKAL TELEBISTA 1 - GAUR EGUN 2 - 20:32h - 00:01:01 EL INSTITUTO DEL EUSKERA DE LA UNIVERSIDAD DEL PAIS VASCO HA ELABORADO EL MAYOR CORPUS DEL EUSKERA CON 205 MILLONES DE PALABRAS. DECLARACIONES DE PELLO SALABURU, DIRECTOR DEL INSTITUTO DE EUSKERA DE LA UPV/EHU. IMAGENES DE IÑIGO URKULLU, LEHENDAKARI; IÑAKI GOIRIZELAIA, RECTOR DE LA UPV/EHU.
11/03/13	EUSKAL TELEBISTA 2 - TELEBERRI 1 - 15:45h - 00:01:11 EL INSTITUTO DE EUSKERA DE LA UPV/EHU HA ELABORADO EL MAYOR CORPUS DEL EUSKERA, LA MAYOR COLECCION DE TEXTOS EN EL IDIOMA VASCO JAMAS REALIZADA, QUE HA SIDO PRESENTADO HOY EN UN ACTO EN BILBAO AL QUE HA ASISTIDO EL LEHENDAKARI VASCO. DECLARACIONES DE PELLO SALABURU, DIRECTOR INSTITUTO DE EUSKERA.
11/03/13	EUSKAL TELEBISTA 1 - GAUR EGUN 1 - 14:45h - 00:01:09 EL INSTITUTO DEL EUSKERA DE LA UNIVERSIDAD DEL PAIS VASCO HA ELABORADO EL MAYOR CORPUS DEL EUSKERA CON 205 MILLONES DE PALABRAS. DECLARACIONES DE PELLO SALABURU, DIRECTOR DEL INSTITUTO DE EUSKERA DE LA UPV/EHU. IMAGENES DE IÑIGO URKULLU, LEHENDAKARI; IÑAKI GOIRIZELAIA, RECTOR DE LA UPV/EHU.



INTERNET

SELECCIÓN DE LAS NOTICIAS DE SU INTERÉS



Universidad
del País Vasco Euskal Herriko
Unibertsitatea

Fecha	Titular/Medio	Pág.	Docs.
13/03/13	La Universidad del País Vasco ha publicado en su web el mayor corpus de palabras en euskera, con 205 millones de términos / El Confidencial Autónomo	28	1
12/03/13	Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo / Noticias de Alava	29	2
12/03/13	Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo / Noticias de Navarra	31	2
12/03/13	Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo / Noticias de Gipuzkoa	33	2
12/03/13	Nace el mayor corpus de euskera de este siglo / Deia.com	35	2
12/03/13	Ibon Sarasola: "Koldo Mitxelena estaría encantado con este corpus" / Deia.com	37	2
12/03/13	Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo / Deia.com	39	2
12/03/13	"Koldo Mitxelena estaría encantado con este corpus" / Deia.com	41	2
11/03/13	Nace el mayor corpus del euskera, con 205 millones de palabras / EITB	43	1
11/03/13	Urkullu señala que no perderán tiempo en batallas estériles sobre la Ley Wert / El Mundo	44	1
11/03/13	Activada la alerta por nieve en cotas bajas para el miércoles / Vasco Press	45	1
11/03/13	Urkullu dice que no perderán tiempo en batallas estériles sobre la Ley Wert / El Mundo	46	1
11/03/13	Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" / El Economista.es	47	3
11/03/13	Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" / Europa Press	50	3
11/03/13	Urkullu:no perderemos el tiempo en batallas estériles sobre reforma educativa / ABC.es	53	1
11/03/13	El PNV lleva al Parlamento Vasco la moción de CIU a favor de la consulta en Cataluña / Vasco Press	54	1
11/03/13	Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" / lainformacion.com	55	3
11/03/13	Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" / TeInteresa.es	58	3
11/03/13	Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por ..." / Gente Digital	61	3

Miércoles, 13 de marzo de 2013 04:15

Audiencia: 12.570

Ranking: 4

VPE: 33

Página: 1

Tipología: online

La Universidad del País Vasco ha publicado en su web el mayor corpus de palabras en euskera, con 205 millones de términos

Miércoles, 13 de marzo de 2013

Facebook Twitter Buzz Negociame Wikio Meneame Delicious Comentarios Escribe su comentario Nombre * Correo electrónico * Escribe estos caracteres para enviar el comentario * * Campos obligatorios. Su dirección de correo electrónico no será publicada. Normas de uso: Esta es la opinión de los lectores, no de ECD. Su comentario no debe exceder de los 500 caracteres. No está permitido escribir comentarios contrarios a las leyes españolas o injuriantes.

Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo

Martes, 12 de marzo de 2013

Iñigo Urkullu Lehendakari Iñaki Goirizelaia Rector de la Universidad del País Vasco Bilbao. La cultura vasca vivió ayer una jornada destacada. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó Egungo Testuen Corpusa (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus on-line solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del Diccionario General Vasco-Orotariko Euskal Hiztegia de Euskaltzaindia, con 5 millones de entradas. El nuevo corpus disecciona el euskera como nunca se había hecho hasta ahora, ofreciendo a los usuarios información completa de cada palabra, los textos y las frases en las que se inserta, en qué libro o artículo aparecen, de qué autor, permite comparar entre varios términos el uso que se le da, la red semántica asociada o ver la evolución de su uso. Y todo sobre el euskera actual en base a textos de este siglo porque a partir del año 2000, fecha en que ve la luz Hiztegi Batua, el euskera se ha estabilizado. La presentación de ETC contó con una nutrida representación de la vida política y académica de Euskadi encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron además el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriarte, el presidente de Euskaltzaindia así como distintos parlamentarios y docentes de la universidad. El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar". El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resultados son muy ricos". Entre sus múltiples posibilidades, ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos, lo cual "abre vías para analizar los motivos de esa evolución", apuntó Salaburu. lengua, el alma del ser humano Según explicaron, la consulta también se puede realizar teniendo en cuenta el origen del texto: si es original o una traducción. Lo mismo sucede respecto a la fuente. ETC informa dónde se utiliza más la palabra escogida: en prensa (y, en concreto, en qué medio), en literatura, ciencia, en las enciclopedias, en televisión Así mismo, esta herramienta facilita la red semántica, la cadena formada por otras palabras que tienen las mismas

Audiencia: 24.026

Ranking: 5

VPE: 130

Página: 2

Tipología: online

propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios, que hoy en día se confeccionan utilizando información de este tipo", indicó Salaburu. En su intervención, el lehendakari Urkullu destacó el valor del corpus y destacó la importancia del euskera en el orden de prioridades del Gobierno vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo". Para Urkullu el corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de los mayores del mundo". Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".

Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo

Martes, 12 de marzo de 2013

Elaborado por el Instituto de Euskera de la UPV/EHU está disponible on-line en www.ehu.es/etc La cultura vasca vivió ayer una jornada destacada. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó Egungo Testuen Corpusa (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus on-line solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del Diccionario General Vasco-Orotariko Euskal Hiztegia de Euskaltzaindia, con 5 millones de entradas. Bilbao. La presentación de ETC contó con una nutrida representación de la vida política y académica de Euskadi encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron además el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriarte, el presidente de Euskaltzaindia así como distintos parlamentarios y docentes de la universidad. El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar". El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resultados son muy ricos". Entre sus múltiples posibilidades, ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos, lo cual "abre vías para analizar los motivos de esa evolución", apuntó Salaburu. lengua, el alma del ser humano En su intervención, el lehendakari Urkullu destacó el valor del corpus y destacó la importancia del euskera en el orden de prioridades del Gobierno vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo". Para Urkullu el corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de

Audiencia: 102.915

Ranking: 5

VPE: 689

Página: 2

Tipología: online

los mayores del mundo". Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".

Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo

Martes, 12 de marzo de 2013

ha sido elaborado por el instituto de euskera de la UPV/EHU La consulta se puede realizar según el origen del texto o la fuente votos comentarios (?) Presentación de Egungo Testuen CorpUSA, celebrada ayer en la sala Baroja del Paraninfo de la Universidad del País Vasco (UPV/EHU) situado en Abandoibarra. (Foto: Pablo Viñas) Vista: Más texto Más visual Galería Noticia Bilbao. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó ayer Egungo Testuen CorpUSA (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus online solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del Diccionario General Vasco-Orotariko Euskal Hiztegia de Euskaltzaindia, con cinco millones de entradas. El nuevo corpus disecciona el euskera como nunca se había hecho hasta ahora, ofreciendo a los usuarios información completa de cada palabra, los textos y las frases en las que se inserta, en qué libro o artículo aparecen, de qué autor, permite comparar entre varios términos el uso que se le da, la red semántica asociada o ver la evolución de su uso. Y todo sobre el euskera actual en la base a textos de este siglo porque a partir del año 2000, fecha en la que ve la luz Hiztegi Batua, el euskera se ha estabilizado. La presentación de ETC contó con una nutrida representación de la vida política y académica de Euskadi que estuvo encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron, además, el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriarte, el presidente de Euskaltzaindia, así como distintos parlamentarios y docentes de la universidad. El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrece sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra infertaz es muy simple, clara e intuitiva, fácil de usar". El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resultados son muy ricos". Entre las múltiples posibilidades ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos, lo cual "abre vías para analizar los motivos de esa evolución". lengua, el alma del ser humano Según explicaron, la consulta también se puede realizar teniendo en cuenta el origen del texto: si es original o una traducción. Lo mismo sucede respecto a la fuente. ETC informa dónde

se utiliza más la palabra escogida: en prensa (y, en concreto, en qué medio), en literatura, ciencia, en las enciclopedias, en televisión Asimismo, esta herramienta facilita la red semántica, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios, que hoy en día se confeccionan utilizando información de este tipo", indicó Salaburu. En su intervención, el lehendakari Urkullu destacó el valor del corpus y la importancia del euskera en el orden de prioridades del Gobierno Vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así nos sentimos hoy todos, reconfortados con este trabajo". Para Urkullu este corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho en euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de los mayores del mundo". Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".

Nace el mayor corpus de euskera de este siglo

Martes, 12 de marzo de 2013

205 millones de términos en euskera Elaborado por el Instituto de Euskera de la UPV/EHU está disponible 'on line' en www.ehu.es/etc Bilbao. La cultura vasca vivió ayer una jornada destacada. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó Egungo Testuen Corpusa (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus on line solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del Diccionario General Vasco-Orotariko Euskal Hiztegia de Euskaltzaindia, con 5 millones de entradas. El nuevo corpus disecciona el euskera como nunca se había hecho hasta ahora, ofreciendo a los usuarios información completa de cada palabra, los textos y las frases en las que se inserta, en qué libro o artículo aparecen, de qué autor, permite comparar entre varios términos el uso que se le da, la red semántica asociada o ver la evolución de su uso. Y todo sobre el euskera actual en base a textos de este siglo porque a partir del año 2000, fecha en que ve la luz Hiztegi Batua, el euskera se ha estabilizado. La presentación de ETC contó con una nutrida representación de la vida política y académica de Euskadi encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron además el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriarte, el presidente de Euskaltzaindia así como distintos parlamentarios y docentes de la universidad. El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar". El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resultados son muy ricos". Entre sus múltiples posibilidades, ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos, lo cual "abre vías para analizar los motivos de esa evolución", apuntó Salaburu. lengua, el alma del ser humano Según explicaron, la consulta también se puede realizar teniendo en cuenta el origen del texto: si es original o una traducción. Lo mismo sucede respecto a la fuente. ETC informa dónde se utiliza más la palabra escogida: en prensa (y, en concreto, en qué medio), en literatura, ciencia, en las enciclopedias, en televisión Así mismo, esta herramienta facilita la red semántica, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar

Audiencia: 80.897**Ranking:** 6**VPE:** 529**Página:** 2**Tipología:** online

diccionarios, que hoy en día se confeccionan utilizando información de este tipo", indicó Salaburu. En su intervención, el lehendakari Urkullu destacó el valor del corpus y también la importancia del euskera en el orden de prioridades del Gobierno vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo". Para Urkullu, el corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho en euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de los mayores del mundo". Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".

Ibon Sarasola: "Koldo Mitxelena estaría encantado con este corpus"

Martes, 12 de marzo de 2013

Ibon Sarasola Investigador del Instituto de euskera de la UPV/EHU Lingüista, miembro de Euskaltzaindia e investigador, Ibon Sarasola presenta el fruto de varios años de trabajo, Egungo Testuen Corpora, la continuación natural de 'Orotariko Euskal Hiztegia' pasado por el filtro de las TIC. El resultado es un corpus de 205 millones de términos Bilbao. Sarasola cambió la Ingeniería Industrial por las palabras. Discípulo de Koldo Mitxelena, el Premio Euskadi de Investigación de 2011 introduce en este siglo la compilación del amplio patrimonio lexicográfico del euskera sobre textos producidos desde 2000 en un corpus al alcance de todos. ¿Qué hace un ingeniero de formación estudiando las palabras? Porque creo que hoy en día para estudiar las palabras hay que tener formación de ingeniero. ¿En qué se diferencia un diccionario de este corpus que acaban de presentar? Los diccionarios han cambiado mucho los últimos 50 años. Antes un diccionario daba el significado de una palabra. Hoy en día, en cambio, se tiende a dar el significado de la palabra pero en el contexto en el que se usa. Por ejemplo, en español estallar y explotar no es lo mismo ya que las guerras estallan, no explotan, y para alguien que está aprendiendo español esa información es decisiva, más que la definición de guerra que seguramente ya conocerá. Un corpus te proporciona la palabra en su contexto, en la frase en la que vive. Entonces se recogen todas las apariciones de la palabra en distintos textos y haces un diccionario con los usos más interesantes. Hoy en día, detrás de todo diccionario que se precie hay un corpus del idioma. Quizá a la gente de a pie su corpus de euskera le puede resultar algo lejano ¿Por qué es importante? El corpus que acabamos de presentar nos permite conocer el euskera actual porque es un corpus de los últimos diez años. Es un corpus que recoge qué palabras en euskera se han usado en los textos y cuáles no, lo cual ayuda a ver la evolución que ha tenido una palabra, qué términos que se usaban antes ya no se usan, que nuevas palabras se han empezado a usar en la última década y todo ello en qué contextos, con qué otras palabras suelen relacionarse. Usted sostiene que una palabra es más que semántica. Es el conjunto de contextos en el que puede aparecer. El corpus compara el uso de las palabras en distintos textos, indicando tendencias, lo cual es muy importante en el caso de que una academia de la lengua quiera legislar sobre alguna palabra ya que con el corpus puede hacerlo en base a unas razones concretas. Dicho de otro modo, el corpus es como poner el euskera que se habla ahora bajo el microscopio. Si hubiera una guerra entre palabras ¿cuál elegir?, ¿la que tenga mayor tradición?, ¿la que empleen los mejores escritores? En todo esto ayuda un corpus ya que registra cada aparición de una palabra, en qué contexto, en qué libro, de qué autor, en qué página o artículo periodístico... El conocimiento del euskera se ha disparado en 30 años gracias a la escuela. ¿Este proceso ha afectado al idioma? Desde luego que sí, aunque no a su estructura. Yo creo que hoy en día se escribe mejor que hace 20 años. Habrá un momento en el que la proporción de personas con el euskera como lengua materna y las que lo hayan aprendido en la escuela cambiará, y con

Audiencia: 80.897**Ranking:** 6**VPE:** 529**Página:** 2**Tipología:** online

un corpus podremos saber si los euskaldunberris escriben de una manera y los euskaldunzaharras de otra. Pero de momento, no se observa esa diferencia. ¿Qué pensaría Mitxelena de este corpus 'on line' tan moderno? Estaría encantado porque su diccionario Orotariko Euskal Hiztegia se basa en el primer corpus que se hizo para la lengua vasca, y este es una continuación, es más de lo mismo, así que estaría encantado.

Nace el mayor corpus de euskera con 205 millones de términos de textos de este siglo

Martes, 12 de marzo de 2013

patrimonio lingüístico Elaborado por el Instituto de Euskera de la UPV/EHU está disponible 'on line' en www.ehu.es/etc Bilbao. La cultura vasca vivió ayer una jornada destacada. El Instituto de Euskera de la Universidad del País Vasco (UPV/EHU) presentó Egungo Testuen Corpusa (ETC), el corpus de euskera más grande jamás elaborado con 205 millones de palabras. La creación de este corpus on line solo es comparable con la magna tarea de compilación lexicográfica que comenzó en 1984 Koldo Mitxelena y culminó 21 años después con la publicación del último volumen del Diccionario General Vasco-Orotariko Euskal Hiztegia de Euskaltzaindia, con 5 millones de entradas. El nuevo corpus disecciona el euskera como nunca se había hecho hasta ahora, ofreciendo a los usuarios información completa de cada palabra, los textos y las frases en las que se inserta, en qué libro o artículo aparecen, de qué autor, permite comparar entre varios términos el uso que se le da, la red semántica asociada o ver la evolución de su uso. Y todo sobre el euskera actual en base a textos de este siglo porque a partir del año 2000, fecha en que ve la luz Hiztegi Batua, el euskera se ha estabilizado. La presentación de ETC contó con una nutrida representación de la vida política y académica de Euskadi encabezada por el lehendakari, Iñigo Urkullu. En el acto intervinieron además el rector de la UPV/EHU, Iñaki Goirizelaia; el director del Instituto de Euskera, Pello Salaburu; Pablo Mongelos, director general de Lagun Aro, y el Premio Euskadi de Investigación 2011, el profesor Ibon Sarasola, principal responsable del proyecto junto a Josu Landa en la parte informática. En la sala Baroja del Paraninfo de la UPV/EHU también estuvieron siguiendo el acto la consejera de Educación, Política Lingüística y Cultura, Cristina Uriarte, el presidente de Euskaltzaindia así como distintos parlamentarios y docentes de la universidad. El corpus ya se puede consultar en www.ehu.es/etc. Pese a no ser uno de los corpus más grandes si se compara con los 150.000 millones de términos que tiene el de Google, sí pasa por ser uno de los más completos de un idioma. ETC "es uno de los corpus más amplios del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", afirmó ayer Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar". El director del Instituto de Euskera comentó que realizar búsquedas "resulta muy sencillo, y los resultados son muy ricos". Entre sus múltiples posibilidades, ofrece un gráfico que permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos, lo cual "abre vías para analizar los motivos de esa evolución", apuntó Salaburu. lengua, el alma del ser humano Según explicaron, la consulta también se puede realizar teniendo en cuenta el origen del texto: si es original o una traducción. Lo mismo sucede respecto a la fuente. ETC informa dónde se utiliza más la palabra escogida: en prensa (y, en concreto, en qué medio), en literatura, ciencia, en las enciclopedias, en televisión Así mismo, esta herramienta facilita la red semántica, la cadena formada por otras palabras que

Audiencia: 80.897**Ranking:** 6**VPE:** 529**Página:** 2**Tipología:** online

tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios, que hoy en día se confeccionan utilizando información de este tipo", indicó Salaburu. En su intervención, el lehendakari Urkullu destacó el valor del corpus y también la importancia del euskera en el orden de prioridades del Gobierno vasco. "La lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", dijo Urkullu, y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". Añadió que "hoy (por ayer) el euskera ha hecho un alto en el camino y se ha contemplado a sí mismo y se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo". Para Urkullu, el corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". Por su parte, el rector Goirizelaia afirmó que el trabajo del Instituto de Euskera es "la mayor colección que se ha hecho en euskera y, teniendo en cuenta la amplia información de cada palabra, es uno de los mayores del mundo". Dijo que la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo" a través de un centro que ha trabajado "sin parar, realizando estudios, jornadas y ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia manifestó que ETC "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se mostró convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro".

"Koldo Mitxelena estaría encantado con este corpus"

Martes, 12 de marzo de 2013

Ibon Sarasola Investigador del Instituto de euskera de la UPV/EHU Lingüista, miembro de Euskaltzaindia e investigador, Ibon Sarasola presenta el fruto de varios años de trabajo, Egungo Testuen Corpora, la continuación natural de 'Orotariko Euskal Hiztegia' pasado por el filtro de las TIC. El resultado es un corpus de 205 millones de términos Bilbao. Sarasola cambió la Ingeniería Industrial por las palabras. Discípulo de Koldo Mitxelena, el Premio Euskadi de Investigación de 2011 introduce en este siglo la compilación del amplio patrimonio lexicográfico del euskera sobre textos producidos desde 2000 en un corpus al alcance de todos. ¿Qué hace un ingeniero de formación estudiando las palabras? Porque creo que hoy en día para estudiar las palabras hay que tener formación de ingeniero. ¿En qué se diferencia un diccionario de este corpus que acaban de presentar? Los diccionarios han cambiado mucho los últimos 50 años. Antes un diccionario daba el significado de una palabra. Hoy en día, en cambio, se tiende a dar el significado de la palabra pero en el contexto en el que se usa. Por ejemplo, en español estallar y explotar no es lo mismo ya que las guerras estallan, no explotan, y para alguien que está aprendiendo español esa información es decisiva, más que la definición de guerra que seguramente ya conocerá. Un corpus te proporciona la palabra en su contexto, en la frase en la que vive. Entonces se recogen todas las apariciones de la palabra en distintos textos y haces un diccionario con los usos más interesantes. Hoy en día, detrás de todo diccionario que se precie hay un corpus del idioma. Quizá a la gente de a pie su corpus de euskera le puede resultar algo lejano ¿Por qué es importante? El corpus que acabamos de presentar nos permite conocer el euskera actual porque es un corpus de los últimos diez años. Es un corpus que recoge qué palabras en euskera se han usado en los textos y cuáles no, lo cual ayuda a ver la evolución que ha tenido una palabra, qué términos que se usaban antes ya no se usan, que nuevas palabras se han empezado a usar en la última década y todo ello en qué contextos, con qué otras palabras suelen relacionarse. Usted sostiene que una palabra es más que semántica. Es el conjunto de contextos en el que puede aparecer. El corpus compara el uso de las palabras en distintos textos, indicando tendencias, lo cual es muy importante en el caso de que una academia de la lengua quiera legislar sobre alguna palabra ya que con el corpus puede hacerlo en base a unas razones concretas. Dicho de otro modo, el corpus es como poner el euskera que se habla ahora bajo el microscopio. Si hubiera una guerra entre palabras ¿cuál elegir?, ¿la que tenga mayor tradición?, ¿la que empleen los mejores escritores? En todo esto ayuda un corpus ya que registra cada aparición de una palabra, en qué contexto, en qué libro, de qué autor, en qué página o artículo periodístico... El conocimiento del euskera se ha disparado en 30 años gracias a la escuela. ¿Este proceso ha afectado al idioma? Desde luego que sí, aunque no a su estructura. Yo creo que hoy en día se escribe mejor que hace 20 años. Habrá un momento en el que la proporción de personas con el euskera como lengua materna y las que lo hayan aprendido en la escuela cambiará, y con

Audiencia: 80.897**Ranking:** 6**VPE:** 529**Página:** 2**Tipología:** online

un corpus podremos saber si los euskaldunberris escriben de una manera y los euskaldunzaharras de otra. Pero de momento, no se observa esa diferencia. ¿Qué pensaría Mitxelena de este corpus 'on line' tan moderno? Estaría encantado porque su diccionario Orotariko Euskal Hiztegia se basa en el primer corpus que se hizo para la lengua vasca, y este es una continuación, es más de lo mismo, así que estaría encantado.

Nace el mayor corpus del euskera, con 205 millones de palabras

Lunes, 11 de marzo de 2013

Elaborado por el Instituto del Euskera de la UPV/EHU, está disponible online a través de ehu.es/etc. Web de 'Egungo Testuen Corpusa'. Foto: EITB La comunidad educativa vasca cuenta con el corpus de euskera más grande jamás elaborado con un total de 204,9 millones de palabras. Elaborado por el Instituto del Euskera de la UPV/EHU, está disponible online a través de ehu.es/etc. Egungo Testuen Corpusa (ETC) ha sido presentado en Bilbao en un acto en el que han intervenido el lehendakari Urkullu, el rector de la UPV/EHU Iñaki Goirizelaia, el director del Instituto del Euskera, Pello Salaburu y el investigador Ibon Sarasola, así como el director general de Lagun Aro, Pablo Mongelos. ETC es el corpus más grande jamás elaborado en euskera que ha necesitado cuatro años para llevarlo a cabo. En formato online y abierto a cualquiera, cuenta con 204,9 millones de palabras en vasco y ofrece información de cada una de ellas, donde la posibilidad de distinguir entre lema y todas sus variantes. También se pueden hacer consultas sobre todos los ejemplos de cada variante, comparar entre varios términos el uso que se les a o consultar conjuntos de palabras. Además permite observar la evolución del uso de un termino mostrando año a año si se va utilizando más o menos, lo que permite abrir vías para analizar los motivos de esa evolución. Todos los textos en prosa sacados de libros, prensa y textos de la Wikipedia escogidos por su calidad pero del siglo XXI. "El euskera se ha estabilizado mucho a partir del 2000, pues ese año publicó Euskaltzaindia Hiztegi Batua' y eso ha tenido gran influencia", ha indicado Salaburu. ETC se complementa con otros corpus elaborados en la anterioridad por el Instituto del Euskera de la UPV/EHU, como son 'Ereduzko prosa Gaur' con 25 millones de palabras. Para elaborar 'Orotariko Euskal Hiztegia' de Euskaltzaindia utilizó un corpus de cinco millones de palabras.

Urkullu señala que no perderán tiempo en batallas estériles sobre la Ley Wert

Lunes, 11 de marzo de 2013

El lehendakari ha afirmado hoy que el Gobierno vasco "no perderá el tiempo en batallas estériles" sobre reformas del modelo educativo, en referencia implícita al proyecto de Ley Orgánica de Mejora de la Calidad Educativa (LOMCE) elaborado por el Ministerio de Educación y conocida como la Ley Wert. "Estamos dispuestos a mejorar y recibir propuestas para avanzar, pero en Euskadi tenemos un sistema educativo sólido, concertado con la comunidad educativa, que funciona y que durante más de dos décadas nos ha dado muy buenos resultados. Esta es nuestra apuesta y para ello adoptaremos las decisiones pertinentes", ha dicho. Durante su intervención en el acto de presentación de un nuevo corpus en euskera, Urkullu ha criticado que "quien cambia de modelo cada cinco años, vuelve a querer cambiarlo una vez mas y quiere condicionarnos", en alusión al Gobierno español. Ha insistido en que el modelo educativo vasco "funciona" y ha avanzado que en el proyecto de Presupuestos para 2013 que aprobará mañana el Ejecutivo autónomo se mantendrá la educación como prioridad "aunque se dispone de menos dinero". "Mantendremos el ratio alumnos-profesor; no cerraremos ninguna aula y avanzaremos en el trilingüismo: este es nuestro compromiso con la comunidad educativa", ha subrayado. Sobre el corpus presentado, el mayor jamás elaborado de euskera, el lehendakari ha apuntado que "recoge el pasado y proyecta el futuro fuerte, vigoroso y completo del euskera". El corpus, al que se accede a través de internet, reúne 205 millones de palabras y una de sus características consiste en que los textos que lo conforman son del siglo XXI. Esta herramienta ofrece información sobre cada palabra, compara el uso entre varios términos y también proporciona información sobre construcciones. En el acto de presentación han participado también el rector de la Universidad del País Vasco (UPV-EHU), Iñaki Goirizelaia, y el director del Instituto del Euskera de esta universidad, Pello Salaburu. También ha contado con la presencia de la consejera vasca de Educación, Política Lingüística y Cultura, Cristina Uriarte, así como de la diputada foral vizcaína de Cultura, Josune Ariztondo.

Activada la alerta por nieve en cotas bajas para el miércoles

Lunes, 11 de marzo de 2013

Bilbao, 11(vasco Press).- El Departamento de Seguridad del Gobierno Vasco ha activado la alerta naranja por nieve para el próximo miércoles a consecuencia de la entrada de una nueva masa de aire polar que dejará precipitaciones en forma de nieve en cotas relativamente bajas. Los empleados municipales de Bilbao no quieren participar en desahucios Bilbao, 11 (Vasco Press).- Los trabajadores del Ayuntamiento de Bilbao han solicitado, a través de sus sindicatos, no participar en los desalojos de viviendas por desahucios. 11.3.2013 - 17:23 Detenido cuando golpeaba a su pareja en la calle en Getxo Bilbao, 11 (Vasco Press).- La Policía Local de Getxo detuvo ayer domingo de madrugada a un varón de 30 años de edad, domiciliado en Bilbao, cuando estaba golpeando a su pareja en la calle. 11.3.2013 - 15:18 El PNV lleva al Parlamento Vasco la moción de CIU a favor de la consulta en Cataluña Vitoria, 11 (Vasco Press).- El PNV defenderá en el Parlamento Vasco la propuesta de resolución que presentó sin éxito CiU en el Debate del Estado de la Nación para pedir la apertura de un proceso de diálogo entre la Generalidad de Cataluña y el Gobierno de España que haga posible la celebración de una consulta a la ciudadanía catalana. 11.3.2013 - 15:17 Nace el mayor corpus del euskera que reúne 205 millones de palabras Bilbao, 11 (Vasco Press).- Desde hoy la comunidad educativa vasca cuenta con el corpus de euskera más grande jamás elaborado con un total de 204,9 millones de palabras. Elaborado por el Instituto del Euskera de la UPV/EHU está disponible on-line a través de www.ehu.es/etc

Urkullu dice que no perderán tiempo en batallas estériles sobre la Ley Wert

Lunes, 11 de marzo de 2013

El lehendakari ha afirmado hoy que el Gobierno vasco "no perderá el tiempo en batallas estériles" sobre reformas del modelo educativo, en referencia implícita al proyecto de Ley Orgánica de Mejora de la Calidad Educativa (LOMCE) elaborado por el Ministerio de Educación y conocida como la Ley Wert. "Estamos dispuestos a mejorar y recibir propuestas para avanzar, pero en Euskadi tenemos un sistema educativo sólido, concertado con la comunidad educativa, que funciona y que durante más de dos décadas nos ha dado muy buenos resultados. Esta es nuestra apuesta y para ello adoptaremos las decisiones pertinentes", ha dicho. Durante su intervención en el acto de presentación de un nuevo corpus en euskera, Urkullu ha criticado que "quien cambia de modelo cada cinco años, vuelve a querer cambiarlo una vez mas y quiere condicionarnos", en alusión al Gobierno español. Ha insistido en que el modelo educativo vasco "funciona" y ha avanzado que en el proyecto de Presupuestos para 2013 que aprobará mañana el Ejecutivo autónomo se mantendrá la educación como prioridad "aunque se dispone de menos dinero". "Mantendremos el ratio alumnos-profesor; no cerraremos ninguna aula y avanzaremos en el trilingüismo: este es nuestro compromiso con la comunidad educativa", ha subrayado. Sobre el corpus presentado, el mayor jamás elaborado de euskera, el lehendakari ha apuntado que "recoge el pasado y proyecta el futuro fuerte, vigoroso y completo del euskera". El corpus, al que se accede a través de internet, reúne 205 millones de palabras y una de sus características consiste en que los textos que lo conforman son del siglo XXI. Esta herramienta ofrece información sobre cada palabra, compara el uso entre varios términos y también proporciona información sobre construcciones. En el acto de presentación han participado también el rector de la Universidad del País Vasco (UPV-EHU), Iñaki Goirizelaia, y el director del Instituto del Euskera de esta universidad, Pello Salaburu. También ha contado con la presencia de la consejera vasca de Educación, Política Lingüística y Cultura, Cristina Uriarte, así como de la diputada foral vizcaína de Cultura, Josune Ariztondo.

Urkuillu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor"

Lunes, 11 de marzo de 2013

Dice que no perderá "tiempo en batallas estériles" con quien quiere cambiar el modelo educativo y "condicionar" a Euskadi BILBAO, 11 (EUROPA PRESS) El lehendakari, Iñigo Urkuillu, ha reconocido que el Gobierno vasco dispone de "menos dinero" para los Presupuestos de 2013, pero se ha comprometido a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" porque la Educación es su "prioridad". Además, ha advertido de que no está dispuesto a "perder el tiempo en batallas estériles" con quienes quieren cambiar el modelo de educación y "condicionar" a Euskadi, donde hay "un pacto educativo sólido, que funciona y ha dado muy buenos resultados durante más de dos décadas". Urkuillu ha asistido este lunes en Bilbao al acto de presentación del proyecto 'Egungo Testuen Corpusa', un colección on-line que compila el mayor corpus del euskera "jamás elaborado", acompañado por el rector de la UPV/EHU, Iñaki Goirizelaia, el director del Instituto de Euskera de la UPV/EHU, Pello Salaburu, el profesor de la UPV/EHU e investigador del Instituto del Euskera Ibon Sarasola, y del director general de Lagun Aro, Pablo Mongelos, fundación que ha apoyado este proyecto. En su intervención, el lehendakari ha afirmado que "la lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". El presidente del Gobierno vasco ha añadido que, "hoy, el euskera, ha hecho un alto en el camino y se ha completado a sí mismo". "Se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo, con el resultado y con el ejemplo de la dedicación, con el trabajo que habéis hecho con tesón y esfuerzo", ha agregado, para destacar que este Corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". El lehendakari ha destacado que "hoy se unen aquí la lengua y la educación". "Nuestro pueblo ha unido lengua y educación. Durante tres décadas hemos recuperado el euskera en el sistema educativo, recuperado y garantizado, y este corpus es la base más sólida para la lengua y la educación", ha aseverado. Tras insistir en que "hoy euskera y educación avanzan juntos en Euskadi", ha afirmado que Egungo Testuen Corpusa es "la base sobre la que confirmamos nuestra apuesta por el euskera y confirmamos también nuestra prioridad educativa". En ese sentido, ha destacado que en Euskadi hay "un pacto educativo sólido que funciona, un pacto que, durante más de dos décadas, nos ha dado muy buenos resultados" y ha criticado que, "ahora, quien ha cambiado su modelo cada cinco años, vuelve a querer cambiarlo una vez más y quiere condicionarnos". Según ha asegurado, su Gobierno está "dispuesto a mejorar, siempre los hemos estado, y a recibir propuestas para avanzar, siempre lo hemos estado, pero no estamos dispuestos a perder el tiempo en batallas estériles". **MODELO SÓLIDO** "Nuestro modelo está concertado con la comunidad educativa, es sólido y funciona, hemos demostrado

capacidad para incorporar el euskera al sistema educativo, y somos la comunidad con mayor conocimiento de una tercera lengua. es nuestra apuesta y nuestro camino", ha remarcado, para asegurar que su determinación es "avanzar, seguir avanzando y, para ello, adoptaremos las decisiones pertinentes". En ese sentido, ha recordado que este martes el Gobierno presentará su proyecto de Presupuestos para Euskadi en 2013, tras aprobarlo en Consejo de Gobierno". En esta línea, ha reconocido que disponen de menos dinero, y ha reiterado que la educación es "nuestra prioridad y lo cumpliremos, incluso afrontando las limitaciones de este inicio de legislatura". De esta forma, ha asegurado que el Gobierno vasco va a "mantener el ratio de alumnos por profesor, no vamos a cerrar ningún aula, y vamos a avanzar en el trilinguismo. Éste es el compromiso del Gobierno con la comunidad educativa".

EGUNGO TESTUEN CORPUSA Egungo Testuen CorpUSA, el último proyecto desarrollado por el Instituto del Euskera, es el corpus más grande jamás elaborado en euskera. Este corpus on-line reúne 205 millones de palabras, mientras que Ereduzko Prosa Gaur, un corpus desarrollado anteriormente por el Instituto de Euskera, abarcaba 25 millones de palabras, y para elaborar Orotariko Euskal Hiztegia, de Euskaltzaindia, se utilizó un corpus de 5 millones de palabras. Egungo Testuen CorpUSA es uno de los corpus "más amplio del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", ha dicho Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar". Otras característica fundamental de este corpus es que todos los textos que lo conforman son del siglo XXI. "El euskera se ha estabilizado mucho a partir de 200, pues ese año publicó Euskaltzaindia Hiztegi Batua y eso ha tenido gran influencia", ha indicado Salaburu. Entre los textos, "los hay de Hegoalde e Iparralde, todos son en prosa y han sido escogidos por su calidad", ha explicado, para señalar que algunos han sido escritos originalmente en euskera y otros son traducciones de otras lenguas. Asimismo, se han tomado en cuenta las fuentes y años de creación, para que tengan un peso similar. Realizar búsquedas en Egungo Testuen CorpUSA resulta "muy sencillo" y todos los resultados son "muy ricos", ya que ofrece información de cada palabra, dando la posibilidad de distinguir entre lema (etxe) y todas sus variantes (etxea, etxeak, etxera, etxetik). Al hacer la búsqueda, aparece automáticamente una "tarta" formada por el lema y sus variantes y muestra un porcentaje de cómo se reparten estas últimas. También se pueden hacer consultas sobre todos los ejemplos de cada variante y permite examinar las palabras que empiezan de determinada manera, comparar entre varios términos el uso que se les da y consultar conjuntos de palabras. Además, ofrece amplia información y posibilidades de búsqueda sobre las palabras escogidas. Así, un gráfico permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos y los ejemplos se pueden ver clasificados año a año. La consulta también se puede realizar teniendo en cuenta el origen del texto: si es original, traducido o nuestro, y da información si se utiliza más cuando ha sido creado o traducido. Lo mismo sucede respecto a la fuente, ya que informa sobre dónde se utiliza más la palabra escogida. Asimismo, proporciona la red semántica, es decir, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios y también permite combinar un lema con otro, para poder así comprobar qué palabras o qué categoría de palabras aparecen delante y detrás del término escogido", ha precisado Salaburu.

Por su parte, el rector de la UPV/EHU, Iñaki Goirizelaia, ha afirmado que este corpus es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra que se ofrece, uno de los mayores del mundo". Según ha asegurado, la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo, a través del Instituto de Euskera dirigido por Pello Salaburu, creado en 1996 y que, desde entonces, ha trabajado sin parar, realizando estudios, jornadas y trabajos, además de ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia ha destacado que el corpus presentado este lunes "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se ha mostrado convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro y servirá para facilitar su desarrollo"

Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor"

Lunes, 11 de marzo de 2013

Dice que no perderá "tiempo en batallas estériles" con quien quiere cambiar el modelo educativo y "condicionar" a Euskadi BILBAO, 11 Mar. (EUROPA PRESS) - El lehendakari, Iñigo Urkullu, ha reconocido que el Gobierno vasco dispone de "menos dinero" para los Presupuestos de 2013, pero se ha comprometido a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" porque la Educación es su "prioridad". Además, ha advertido de que no está dispuesto a "perder el tiempo en batallas estériles" con quienes quieren cambiar el modelo de educación y "condicionar" a Euskadi, donde hay "un pacto educativo sólido, que funciona y ha dado muy buenos resultados durante más de dos décadas". Urkullu ha asistido este lunes en Bilbao al acto de presentación del proyecto 'Egungo Testuen Corpusa', un colección on-line que compila el mayor corpus del euskera "jamás elaborado", acompañado por el rector de la UPV/EHU, Iñaki Goirizelaia, el director del Instituto de Euskera de la UPV/EHU, Pello Salaburu, el profesor de la UPV/EHU e investigador del Instituto del Euskera Ibon Sarasola, y del director general de Lagun Aro, Pablo Mongelos, fundación que ha apoyado este proyecto. En su intervención, el lehendakari ha afirmado que "la lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". El presidente del Gobierno vasco ha añadido que, "hoy, el euskera, ha hecho un alto en el camino y se ha completado a sí mismo". "Se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo, con el resultado y con el ejemplo de la dedicación, con el trabajo que habéis hecho con tesón y esfuerzo", ha agregado, para destacar que este Corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". El lehendakari ha destacado que "hoy se unen aquí la lengua y la educación". "Nuestro pueblo ha unido lengua y educación. Durante tres décadas hemos recuperado el euskera en el sistema educativo, recuperado y garantizado, y este corpus es la base mas sólida para la lengua y la educación", ha aseverado. Tras insistir en que "hoy euskera y educación avanzan juntos en Euskadi", ha afirmado que Egungo Testuen Corpusa es "la base sobre la que confirmamos nuestra apuesta por el euskera y confirmamos también nuestra prioridad educativa". En ese sentido, ha destacado que en Euskadi hay "un pacto educativo sólido que funciona, un pacto que, durante más de dos décadas, nos ha dado muy buenos resultados" y ha criticado que, "ahora, quien ha cambiado su modelo cada cinco años, vuelve a querer cambiarlo una vez más y quiere condicionarnos". Según ha asegurado, su Gobierno está "dispuesto a mejorar, siempre los hemos estado, y a recibir propuestas para avanzar, siempre lo hemos estado, pero no estamos dispuestos a perder el tiempo en batallas estériles". **MODELO SÓLIDO** "Nuestro modelo está concertado con la comunidad educativa, es sólido y funciona, hemos demostrado

capacidad para incorporar el euskera al sistema educativo, y somos la comunidad con mayor conocimiento de una tercera lengua. es nuestra apuesta y nuestro camino", ha remarcado, para asegurar que su determinación es "avanzar, seguir avanzando y, para ello, adoptaremos las decisiones pertinentes". En ese sentido, ha recordado que este martes el Gobierno presentará su proyecto de Presupuestos para Euskadi en 2013, tras aprobarlo en Consejo de Gobierno". En esta línea, ha reconocido que disponen de menos dinero, y ha reiterado que la educación es "nuestra prioridad y lo cumpliremos, incluso afrontando las limitaciones de este inicio de legislatura". De esta forma, ha asegurado que el Gobierno vasco va a "mantener el ratio de alumnos por profesor, no vamos a cerrar ningún aula, y vamos a avanzar en el trilinguismo. Éste es el compromiso del Gobierno con la comunidad educativa".

EGUNGO TESTUEN CORPUSA Egungo Testuen CorpUSA, el último proyecto desarrollado por el Instituto del Euskera, es el corpus más grande jamás elaborado en euskera. Este corpus on-line reúne 205 millones de palabras, mientras que Ereduzko Prosa Gaur, un corpus desarrollado anteriormente por el Instituto de Euskera, abarcaba 25 millones de palabras, y para elaborar Orotariko Euskal Hiztegia, de Euskaltzaindia, se utilizó un corpus de 5 millones de palabras. Egungo Testuen CorpUSA es uno de los corpus "más amplio del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", ha dicho Pello Salaburu, para añadir que "nuestra infertaz es muy simple, clara e intuitiva, fácil de usar". Otras característica fundamental de este corpus es que todos los textos que lo conforman son del siglo XXI. "El euskera se ha estabilizado mucho a partir de 200, pues ese año publicó Euskaltzaindia Hiztegi Batua y eso ha tenido gran influencia", ha indicado Salaburu. Entre los textos, "los hay de Hegoalde e Iparralde, todos son en prosa y han sido escogidos por su calidad", ha explicado, para señalar que algunos han sido escritos originalmente en euskera y otros son traducciones de otras lenguas. Asimismo, se han tomado en cuenta las fuentes y años de creación, para que tengan un peso similar. Realizar búsquedas en Egungo Testuen CorpUSA resulta "muy sencillo" y todos los resultados son "muy ricos", ya que ofrece información de cada palabra, dando la posibilidad de distinguir entre lema (etxe) y todas sus variantes (etxea, etxeak, etxera, etxetik). Al hacer la búsqueda, aparece automáticamente una "tarta" formada por el lema y sus variantes y muestra un porcentaje de cómo se reparten estas últimas. También se pueden hacer consultas sobre todos los ejemplos de cara variante y permite examinar las palabras que empiezan de determinada manera, comparar entre varios términos el uso que se les da y consultar conjuntos de palabra. Además, ofrece amplia información y posibilidades de búsqueda sobre las palabras escogidas. Así, un gráfico permite observar al evolución del uso de un término, mostrando año a año si se va utilizando más o menos y los ejemplos se pueden ver clasificados año a año. La consulta también se puede realizar teniendo en cuenta el origen del texto: si es original, traducido o nuestro, y da información si se utiliza más cuando ha sido creado o traducido. Lo mismo sucede respecto a la fuente, ya que informa sobre dónde se utiliza más la palabra escogida. Asimismo, proporciona la red semántica, es decir, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios y también permite combinar un lema con otro, para poder así comprobar qué palabras o qué categoría de palabras aparecen delante y detrás del término escogido", ha precisado Salaburu.

Audiencia: 1.231.727

Ranking: 7

VPE: 10.088

Página: 3

Tipología: online

Por su parte, el rector de la UPV/EHU, Iñaki Goirizelaia, ha afirmado que este corpus es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra que se ofrece, uno de los mayores del mundo". Según ha asegurado, la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo, a través del Instituto de Euskera dirigido por Pello Salaburu, creado en 1996 y que, desde entonces, ha trabajado sin parar, realizando estudios, jornadas y trabajos, además de ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia ha destacado que el corpus presentado este lunes "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se ha mostrado convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro y servirá para facilitar su desarrollo"

Urkullu: no perderemos el tiempo en batallas estériles sobre reforma educativa

Lunes, 11 de marzo de 2013

El lehendakari Iñigo Urkullu ha afirmado hoy que el Gobierno vasco "no perderá el tiempo en batallas estériles" sobre reformas del modelo educativo, en referencia implícita al proyecto de Ley Orgánica de Mejora de la Calidad Educativa (Lomce) elaborado por el Ministerio de Educación. "Estamos dispuestos a mejorar y recibir propuestas para avanzar, pero en Euskadi tenemos un sistema educativo sólido, concertado con la comunidad educativa, que funciona y que durante más de dos décadas nos ha dado muy buenos resultados. Esta es nuestra apuesta y para ello adoptaremos las decisiones pertinentes", ha dicho. Durante su intervención en el acto de presentación de un nuevo corpus en euskera, Urkullu ha criticado que "quien cambia de modelo cada cinco años, vuelve a querer cambiarlo una vez más y quiere condicionarnos", en alusión al Gobierno español. Ha insistido en que el modelo educativo vasco "funciona" y ha avanzado que en el proyecto de presupuestos para 2013 que aprobará mañana el Ejecutivo autónomo se mantendrá la educación como prioridad "aunque se dispone de menos dinero". "Mantendremos el ratio alumnos-profesor; no cerraremos ninguna aula y avanzaremos en el trilingüismo: este es nuestro compromiso con la comunidad educativa", ha subrayado. Sobre el corpus presentado, el mayor jamás elaborado de euskera, el lehendakari ha apuntado que "recoge el pasado y proyecta el futuro fuerte, vigoroso y completo del euskera". El corpus, al que se accede a través de internet, reúne 205 millones de palabras y una de sus características consiste en que los textos que lo conforman son del siglo XXI. Esta herramienta ofrece información sobre cada palabra, compara el uso entre varios términos y también proporciona información sobre construcciones. En el acto de presentación han participado también el rector de la Universidad del País Vasco (UPV-EHU), Iñaki Goirizelaia, y el director del Instituto del Euskera de esta universidad, Pello Salaburu. También ha contado con la presencia de la consejera vasca de Educación, Política Lingüística y Cultura, Cristina Uriarte, así como de la diputada foral vizcaína de Cultura, Josune Ariztondo.

El PNV lleva al Parlamento Vasco la moción de CIU a favor de la consulta en Cataluña

Lunes, 11 de marzo de 2013

Vitoria, 11 (Vasco Press).- El PNV defenderá en el Parlamento Vasco la propuesta de resolución que presentó sin éxito CiU en el Debate del Estado de la Nación para pedir la apertura de un proceso de diálogo entre la Generalidad de Cataluña y el Gobierno de España que haga posible la celebración de una consulta a la ciudadanía catalana. Nace el mayor corpus del euskera que reúne 205 millones de palabras Bilbao, 11 (Vasco Press).- Desde hoy la comunidad educativa vasca cuenta con el corpus de euskera más grande jamás elaborado con un total de 204,9 millones de palabras. Elaborado por el Instituto del Euskera de la UPV/EHU está disponible on-line a través de www.ehu.es/etc

Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor"

Lunes, 11 de marzo de 2013

Dice que no perderá "tiempo en batallas estériles" con quien quiere cambiar el modelo educativo y "condicionar" a Euskadi El lehendakari, Iñigo Urkullu, ha reconocido que el Gobierno vasco dispone de "menos dinero" para los Presupuestos de 2013, pero se ha comprometido a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" porque la Educación es su "prioridad". Además, ha advertido de que no está dispuesto a "perder el tiempo en batallas estériles" con quienes quieren cambiar el modelo de educación y "condicionar" a Euskadi, donde hay "un pacto educativo sólido, que funciona y ha dado muy buenos resultados durante más de dos décadas". Urkullu ha asistido este lunes en Bilbao al acto de presentación del proyecto 'Egungo Testuen Corpusa', un colección on-line que compila el mayor corpus del euskera "jamás elaborado", acompañado por el rector de la UPV/EHU, Iñaki Goirizelaia, el director del Instituto de Euskera de la UPV/EHU, Pello Salaburu, el profesor de la UPV/EHU e investigador del Instituto del Euskera Ibon Sarasola, y del director general de Lagun Aro, Pablo Mongelos, fundación que ha apoyado este proyecto. En su intervención, el lehendakari ha afirmado que "la lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". El presidente del Gobierno vasco ha añadido que, "hoy, el euskera, ha hecho un alto en el camino y se ha completado a sí mismo". "Se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo, con el resultado y con el ejemplo de la dedicación, con el trabajo que habéis hecho con tesón y esfuerzo", ha agregado, para destacar que este Corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". El lehendakari ha destacado que "hoy se unen aquí la lengua y la educación". "Nuestro pueblo ha unido lengua y educación. Durante tres décadas hemos recuperado el euskera en el sistema educativo, recuperado y garantizado, y este corpus es la base mas sólida para la lengua y la educación", ha aseverado. Tras insistir en que "hoy euskera y educación avanzan juntos en Euskadi", ha afirmado que Egungo Testuen Corpusa es "la base sobre la que confirmamos nuestra apuesta por el euskera y confirmamos también nuestra prioridad educativa". En ese sentido, ha destacado que en Euskadi hay "un pacto educativo sólido que funciona, un pacto que, durante más de dos décadas, nos ha dado muy buenos resultados" y ha criticado que, "ahora, quien ha cambiado su modelo cada cinco años, vuelve a querer cambiarlo una vez más y quiere condicionarnos". Según ha asegurado, su Gobierno está "dispuesto a mejorar, siempre los hemos estado, y a recibir propuestas para avanzar, siempre lo hemos estado, pero no estamos dispuestos a perder el tiempo en batallas estériles". "Nuestro modelo está concertado con la comunidad educativa, es sólido y funciona, hemos demostrado capacidad para incorporar el euskera al sistema

educativo, y somos la comunidad con mayor conocimiento de una tercera lengua. es nuestra apuesta y nuestro camino", ha remarcado, para asegurar que su determinación es "avanzar, seguir avanzando y, para ello, adoptaremos las decisiones pertinentes". En ese sentido, ha recordado que este martes el Gobierno presentará su proyecto de Presupuestos para Euskadi en 2013, tras aprobarlo en ". En esta línea, ha reconocido que disponen de menos dinero, y ha reiterado que la educación es "nuestra prioridad y lo cumpliremos, incluso afrontando las limitaciones de este inicio de legislatura". De esta forma, ha asegurado que el Gobierno vasco va a "mantener el ratio de alumnos por profesor, no vamos a cerrar ningún aula, y vamos a avanzar en el trilinguismo. Éste es el compromiso del Gobierno con la comunidad educativa". Egungo Testuen Corpusa, el último proyecto desarrollado por el Instituto del Euskera, es el corpus más grande jamás elaborado en euskera. Este corpus on-line reúne 205 millones de palabras, mientras que Ereduzko Prosa Gaur, un corpus desarrollado anteriormente por el Instituto de Euskera, abarcaba 25 millones de palabras, y para elaborar Orotariko Euskal Hiztegia, de Euskaltzaindia, se utilizó un corpus de 5 millones de palabras. Egungo Testuen Corpusa es uno de los corpus "más amplio del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", ha dicho Pello Salaburu, para añadir que "nuestra infertaz es muy simple, clara e intuitiva, fácil de usar". Otras característica fundamental de este corpus es que todos los textos que lo conforman son del siglo XXI. "El euskera se ha estabilizado mucho a partir de 200, pues ese año publicó Euskaltzaindia Hiztegi Batua y eso ha tenido gran influencia", ha indicado Salaburu. Entre los textos, "los hay de Hegoalde e Iparralde, todos son en prosa y han sido escogidos por su calidad", ha explicado, para señalar que algunos han sido escritos originalmente en euskera y otros son traducciones de otras lenguas. Asimismo, se han tomado en cuenta las fuentes y años de creación, para que tengan un peso similar. Realizar búsquedas en Egungo Testuen Corpusa resulta "muy sencillo" y todos los resultados son "muy ricos", ya que ofrece información de cada palabra, dando la posibilidad de distinguir entre lema (etxe) y todas sus variantes (etxea, etxeak, etxera, etxetik). Al hacer la búsqueda, aparece automáticamente una "tarta" formada por el lema y sus variantes y muestra un porcentaje de cómo se reparten estas últimas. También se pueden hacer consultas sobre todos los ejemplos de cara variante y permite examinar las palabras que empiezan de determinada manera, comparar entre varios términos el uso que se les da y consultar conjuntos de palabra. Además, ofrece amplia información y posibilidades de búsqueda sobre las palabras escogidas. Así, un gráfico permite observar al evolución del uso de un término, mostrando año a año si se va utilizando más o menos y los ejemplos se pueden ver clasificados año a año. La consulta también se puede realizar teniendo en cuenta el origen del texto: si es original, traducido o nuestro, y da información si se utiliza más cuando ha sido creado o traducido. Lo mismo sucede respecto a la fuente, ya que informa sobre dónde se utiliza más la palabra escogida. Asimismo, proporciona la red semántica, es decir, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios y también permite combinar un lema con otro, para poder así comprobar qué palabras o qué categoría de palabras aparecen delante y detrás del término escogido", ha precisado Salaburu. Por su parte, el rector de la UPV/

Lun, 11 de mar de 2013 15:08**Audiencia:** 481.018**Ranking:** 6**VPE:** 2.266**Página:** 3**Tipología:** online

EHU, Iñaki Goirizelaia, ha afirmado que este corpus es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra que se ofrece, uno de los mayores del mundo". Según ha asegurado, la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo, a través del Instituto de Euskera dirigido por Pello Salaburu, creado en 1996 y que, desde entonces, ha trabajado sin parar, realizando estudios, jornadas y trabajos, además de ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia ha destacado que el corpus presentado este lunes "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se ha mostrado convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro y servirá para facilitar su desarrollo"

Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor"

Lunes, 11 de marzo de 2013

Dice que no perderá "tiempo en batallas estériles" con quien quiere cambiar el modelo educativo y "condicionar" a Euskadi El lehendakari, Iñigo Urkullu, ha reconocido que el Gobierno vasco dispone de "menos dinero" para los Presupuestos de 2013, pero se ha comprometido a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" porque la Educación es su "prioridad". Además, ha advertido de que no está dispuesto a "perder el tiempo en batallas estériles" con quienes quieren cambiar el modelo de educación y "condicionar" a Euskadi, donde hay "un pacto educativo sólido, que funciona y ha dado muy buenos resultados durante más de dos décadas". Urkullu ha asistido este lunes en Bilbao al acto de presentación del proyecto 'Egungo Testuen Corpusa', un colección on-line que compila el mayor corpus del euskera "jamás elaborado", acompañado por el rector de la UPV/EHU, Iñaki Goirizelaia, el director del Instituto de Euskera de la UPV/EHU, Pello Salaburu, el profesor de la UPV/EHU e investigador del Instituto del Euskera Ibon Sarasola, y del director general de Lagun Aro, Pablo Mongelos, fundación que ha apoyado este proyecto. En su intervención, el lehendakari ha afirmado que "la lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". El presidente del Gobierno vasco ha añadido que, "hoy, el euskera, ha hecho un alto en el camino y se ha completado a sí mismo". "Se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo, con el resultado y con el ejemplo de la dedicación, con el trabajo que habéis hecho con tesón y esfuerzo", ha agregado, para destacar que este Corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". El lehendakari ha destacado que "hoy se unen aquí la lengua y la educación". "Nuestro pueblo ha unido lengua y educación. Durante tres décadas hemos recuperado el euskera en el sistema educativo, recuperado y garantizado, y este corpus es la base mas sólida para la lengua y la educación", ha aseverado. Tras insistir en que "hoy euskera y educación avanzan juntos en Euskadi", ha afirmado que Egungo Testuen Corpusa es "la base sobre la que confirmamos nuestra apuesta por el euskera y confirmamos también nuestra prioridad educativa". En ese sentido, ha destacado que en Euskadi hay "un pacto educativo sólido que funciona, un pacto que, durante más de dos décadas, nos ha dado muy buenos resultados" y ha criticado que, "ahora, quien ha cambiado su modelo cada cinco años, vuelve a querer cambiarlo una vez más y quiere condicionarnos". Según ha asegurado, su Gobierno está "dispuesto a mejorar, siempre los hemos estado, y a recibir propuestas para avanzar, siempre lo hemos estado, pero no estamos dispuestos a perder el tiempo en batallas estériles". MODELO SÓLIDO "Nuestro modelo está concertado con la comunidad educativa, es sólido y funciona, hemos demostrado capacidad para incorporar

el euskera al sistema educativo, y somos la comunidad con mayor conocimiento de una tercera lengua. es nuestra apuesta y nuestro camino", ha remarcado, para asegurar que su determinación es "avanzar, seguir avanzando y, para ello, adoptaremos las decisiones pertinentes". En ese sentido, ha recordado que este martes el Gobierno presentará su proyecto de Presupuestos para Euskadi en 2013, tras aprobarlo en Consejo de Gobierno". En esta línea, ha reconocido que disponen de menos dinero, y ha reiterado que la educación es "nuestra prioridad y lo cumpliremos, incluso afrontando las limitaciones de este inicio de legislatura". De esta forma, ha asegurado que el Gobierno vasco va a "mantener el ratio de alumnos por profesor, no vamos a cerrar ningún aula, y vamos a avanzar en el trilinguismo. Éste es el compromiso del Gobierno con la comunidad educativa".

EGUNGO TESTUEN CORPUSA Egungo Testuen Corpusa, el último proyecto desarrollado por el Instituto del Euskera, es el corpus más grande jamás elaborado en euskera. Este corpus on-line reúne 205 millones de palabras, mientras que Ereduzko Prosa Gaur, un corpus desarrollado anteriormente por el Instituto de Euskera, abarcaba 25 millones de palabras, y para elaborar Orotariko Euskal Hiztegia, de Euskaltzaindia, se utilizó un corpus de 5 millones de palabras. Egungo Testuen Corpusa es uno de los corpus "más amplio del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", ha dicho Pello Salaburu, para añadir que "nuestra infertaz es muy simple, clara e intuitiva, fácil de usar".

Otras característica fundamental de este corpus es que todos los textos que lo conforman son del siglo XXI. "El euskera se ha estabilizado mucho a partir de 200, pues ese año publicó Euskaltzaindia Hiztegi Batua y eso ha tenido gran influencia", ha indicado Salaburu. Entre los textos, "los hay de Hegoalde e Iparralde, todos son en prosa y han sido escogidos por su calidad", ha explicado, para señalar que algunos han sido escritos originalmente en euskera y otros son traducciones de otras lenguas. Asimismo, se han tomado en cuenta las fuentes y años de creación, para que tengan un peso similar. Realizar búsquedas en Egungo Testuen Corpusa resulta "muy sencillo" y todos los resultados son "muy ricos", ya que ofrece información de cada palabra, dando la posibilidad de distinguir entre lema (etxe) y todas sus variantes (etxea, etxeak, etxera, etxetik). Al hacer las búsqueda, aparece automáticamente una "tarta" formada por el lema y sus variantes y muestra un porcentaje de cómo se reparten estas últimas. También se pueden hacer consultas sobre todos los ejemplos de cara variante y permite examinar las palabras que empiezan de determinada manera, comparar entre varios términos el uso que se les da y consultar conjuntos de palabra. Además, ofrece amplia información y posibilidades de búsqueda sobre las palabras escogidas. Así, un gráfico permite observar al evolución del uso de un término, mostrando año a año si se va utilizando más o menos y los ejemplos se pueden ver clasificados año a año. La consulta también se puede realizar teniendo en cuenta el origen del texto: si es original, traducido o nuestro, y da información si se utiliza más cuando ha sido creado o traducido. Lo mismo sucede respecto a la fuente, ya que informa sobre dónde se utiliza más la palabra escogida. Asimismo, proporciona la red semántica, es decir, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios y también permite combinar un lema con otro, para poder así comprobar qué palabras o qué categoría de palabras aparecen delante y detrás del término escogido", ha precisado Salaburu. Por su parte, el rector de la UPV/

EHU, Iñaki Goirizelaia, ha afirmado que este corpus es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra que se ofrece, uno de los mayores del mundo". Según ha asegurado, la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo, a través del Instituto de Euskera dirigido por Pello Salaburu, creado en 1996 y que, desde entonces, ha trabajado sin parar, realizando estudios, jornadas y trabajos, además de ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia ha destacado que el corpus presentado este lunes "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se ha mostrado convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro y servirá para facilitar su desarrollo"

Urkullu se compromete a no cerrar ningún aula y a mantener "el ratio de alumnos por ...

Lunes, 11 de marzo de 2013

Dice que no perderá "tiempo en batallas estériles" con quien quiere cambiar el modelo educativo y "condicionar" a Euskadi BILBAO, 11 (EUROPA PRESS) El lehendakari, Iñigo Urkullu, ha reconocido que el Gobierno vasco dispone de "menos dinero" para los Presupuestos de 2013, pero se ha comprometido a no cerrar ningún aula y a mantener "el ratio de alumnos por profesor" porque la Educación es su "prioridad". Además, ha advertido de que no está dispuesto a "perder el tiempo en batallas estériles" con quienes quieren cambiar el modelo de educación y "condicionar" a Euskadi, donde hay "un pacto educativo sólido, que funciona y ha dado muy buenos resultados durante más de dos décadas". Urkullu ha asistido este lunes en Bilbao al acto de presentación del proyecto 'Egungo Testuen Corpusa', un colección on-line que compila el mayor corpus del euskera "jamás elaborado", acompañado por el rector de la UPV/EHU, Iñaki Goirizelaia, el director del Instituto de Euskera de la UPV/EHU, Pello Salaburu, el profesor de la UPV/EHU e investigador del Instituto del Euskera Ibon Sarasola, y del director general de Lagun Aro, Pablo Mongelos, fundación que ha apoyado este proyecto. En su intervención, el lehendakari ha afirmado que "la lengua es el alma del ser humano y el euskera es el cuerpo de nuestro alma", y la UPV/EHU "ha creado el corpus del euskera, un trabajo magnífico, excelente, completo, una base magnífica para el euskera, ahora a disposición de todos". El presidente del Gobierno vasco ha añadido que, "hoy, el euskera, ha hecho un alto en el camino y se ha completado a sí mismo". "Se ve más fuerte, más completo y vigoroso, y así no sentimos hoy todos, reconfortados con este trabajo, con el resultado y con el ejemplo de la dedicación, con el trabajo que habéis hecho con tesón y esfuerzo", ha agregado, para destacar que este Corpus "recoge el pasado y proyecta el futuro del euskera, un futuro más fuerte, más completo y vigoroso". El lehendakari ha destacado que "hoy se unen aquí la lengua y la educación". "Nuestro pueblo ha unido lengua y educación. Durante tres décadas hemos recuperado el euskera en el sistema educativo, recuperado y garantizado, y este corpus es la base mas sólida para la lengua y la educación", ha aseverado. Tras insistir en que "hoy euskera y educación avanzan juntos en Euskadi", ha afirmado que Egungo Testuen Corpusa es "la base sobre la que confirmamos nuestra apuesta por el euskera y confirmamos también nuestra prioridad educativa". En ese sentido, ha destacado que en Euskadi hay "un pacto educativo sólido que funciona, un pacto que, durante más de dos décadas, nos ha dado muy buenos resultados" y ha criticado que, "ahora, quien ha cambiado su modelo cada cinco años, vuelve a querer cambiarlo una vez más y quiere condicionarnos". Según ha asegurado, su Gobierno está "dispuesto a mejorar, siempre los hemos estado, y a recibir propuestas para avanzar, siempre lo hemos estado, pero no estamos dispuestos a perder el tiempo en batallas estériles". **MODELO SÓLIDO** "Nuestro modelo está concertado con la comunidad educativa, es sólido y funciona, hemos demostrado

capacidad para incorporar el euskera al sistema educativo, y somos la comunidad con mayor conocimiento de una tercera lengua. es nuestra apuesta y nuestro camino", ha remarcado, para asegurar que su determinación es "avanzar, seguir avanzando y, para ello, adoptaremos las decisiones pertinentes". En ese sentido, ha recordado que este martes el Gobierno presentará su proyecto de Presupuestos para Euskadi en 2013, tras aprobarlo en Consejo de Gobierno". En esta línea, ha reconocido que disponen de menos dinero, y ha reiterado que la educación es "nuestra prioridad y lo cumpliremos, incluso afrontando las limitaciones de este inicio de legislatura". De esta forma, ha asegurado que el Gobierno vasco va a "mantener el ratio de alumnos por profesor, no vamos a cerrar ningún aula, y vamos a avanzar en el trilinguismo. Éste es el compromiso del Gobierno con la comunidad educativa".

EGUNGO TESTUEN CORPUSA Egungo Testuen CorpUSA, el último proyecto desarrollado por el Instituto del Euskera, es el corpus más grande jamás elaborado en euskera. Este corpus on-line reúne 205 millones de palabras, mientras que Ereduzko Prosa Gaur, un corpus desarrollado anteriormente por el Instituto de Euskera, abarcaba 25 millones de palabras, y para elaborar Orotariko Euskal Hiztegia, de Euskaltzaindia, se utilizó un corpus de 5 millones de palabras. Egungo Testuen CorpUSA es uno de los corpus "más amplio del mundo, pues aunque los hay mayores, la información que ofrecen sobre cada palabra es más básica y menos manejable", ha dicho Pello Salaburu, para añadir que "nuestra interfaz es muy simple, clara e intuitiva, fácil de usar".

Otras característica fundamental de este corpus es que todos los textos que lo conforman son del siglo XXI. "El euskera se ha estabilizado mucho a partir de 200, pues ese año publicó Euskaltzaindia Hiztegi Batua y eso ha tenido gran influencia", ha indicado Salaburu. Entre los textos, "los hay de Hegoalde e Iparralde, todos son en prosa y han sido escogidos por su calidad", ha explicado, para señalar que algunos han sido escritos originalmente en euskera y otros son traducciones de otras lenguas. Asimismo, se han tomado en cuenta las fuentes y años de creación, para que tengan un peso similar. Realizar búsquedas en Egungo Testuen CorpUSA resulta "muy sencillo" y todos los resultados son "muy ricos", ya que ofrece información de cada palabra, dando la posibilidad de distinguir entre lema (etxe) y todas sus variantes (etxea, etxeak, etxera, etxetik). Al hacer la búsqueda, aparece automáticamente una "tarta" formada por el lema y sus variantes y muestra un porcentaje de cómo se reparten estas últimas. También se pueden hacer consultas sobre todos los ejemplos de cada variante y permite examinar las palabras que empiezan de determinada manera, comparar entre varios términos el uso que se les da y consultar conjuntos de palabras. Además, ofrece amplia información y posibilidades de búsqueda sobre las palabras escogidas. Así, un gráfico permite observar la evolución del uso de un término, mostrando año a año si se va utilizando más o menos y los ejemplos se pueden ver clasificados año a año. La consulta también se puede realizar teniendo en cuenta el origen del texto: si es original, traducido o nuestro, y da información si se utiliza más cuando ha sido creado o traducido. Lo mismo sucede respecto a la fuente, ya que informa sobre dónde se utiliza más la palabra escogida. Asimismo, proporciona la red semántica, es decir, la cadena formada por otras palabras que tienen las mismas propiedades que el término buscado. "Estas redes son imprescindibles para elaborar diccionarios y también permite combinar un lema con otro, para poder así comprobar qué palabras o qué categoría de palabras aparecen delante y detrás del término escogido", ha precisado Salaburu.

Audiencia: 29.330

Ranking: 5

VPE: 114

Página: 3

Tipología: online

Por su parte, el rector de la UPV/EHU, Iñaki Goirizelaia, ha afirmado que este corpus es "la mayor colección que se ha hecho euskera y, teniendo en cuenta la amplia información de cada palabra que se ofrece, uno de los mayores del mundo". Según ha asegurado, la UPV/EHU está "orgullosa por haber tenido la oportunidad de impulsar este trabajo, a través del Instituto de Euskera dirigido por Pello Salaburu, creado en 1996 y que, desde entonces, ha trabajado sin parar, realizando estudios, jornadas y trabajos, además de ofreciendo instrumentos para utilizar mejor el euskera, como gramáticas, corpus de textos y diccionarios". Goirizelaia ha destacado que el corpus presentado este lunes "sistematiza el tremendo patrimonio lingüístico surgido en el presente siglo" y se ha mostrado convencido de "este trabajo para clasificar, archivar y reunir se convertirá en un fructífero patrimonio de nuestra lengua en el futuro y servirá para facilitar su desarrollo"