



# Corpusgintza gaur egun

MINTEGIA

# Corpus lingüísticos hoy en día

SEMINARIO

2010ko urtarrilaren 21a



# Corpus textuales y planificación lingüística

Xavier Gómez Guinovart



---

Universidade de Vigo

**TALG**

Grupo de investigación  
Tecnoloxías e Aplicacións  
da Lingua Galega

## Qué es un corpus

- un **corpus lingüístico** es una colección de **textos** escritos o hablados
- almacenados en **soporte informático**
- **representativos** de una variedad lingüística (o de un conjunto de variedades)
- recopilados para la **investigación** lingüística o filológica
- y/o con aplicación a diversos productos de las **industrias de la lengua**:
  - diccionarios, gramáticas, material didáctico, programas de ingeniería lingüística...



## Qué contiene un corpus

- según el **modo de producción** de los textos: textos escritos, textos orales o ambos
- según la **orientación cronológica** de la selección: textos de un período temporal o de varios períodos
- según el **tipo de lengua** de la selección: textos de lengua general o de variedades especializadas
- según el **número de lenguas**: textos de una lengua o de varias lenguas (y, en este caso, en traducción o no)
- según la **información añadida**: textos con información (lingüística/textual) adicional o sin ella
- según la **técnica de selección**: un conjunto de textos cerrado o abierto, ponderado o no, aleatorio o no



## Tipos de corpus

- según el **modo de producción**: corpus escritos, orales y mixtos
- según la **orientación cronológica**: corpus sincrónicos, diacrónicos, históricos y contemporáneos
- según el **tipo de lengua**: corpus generales, especializados, técnicos y dialectales
- según el **número de lenguas**: corpus monolingües y plurilingües, corpus paralelos y comparables
- según la **información añadida**: corpus etiquetados y sin etiquetar, corpus lematizados, corpus alineados
- según la **técnica de selección**: corpus abiertos y cerrados, corpus equilibrados, corpus de muestra o estadísticos



# La planificación lingüística

- la **planificación lingüística** es el intento de modificar la realidad sociolingüística de un territorio mediante la aplicación de un conjunto coherente de actuaciones de **política lingüística**
- en los países con lenguas minorizadas el objetivo fundamental de la planificación lingüística es la **normalización lingüística**



## La planificación del corpus

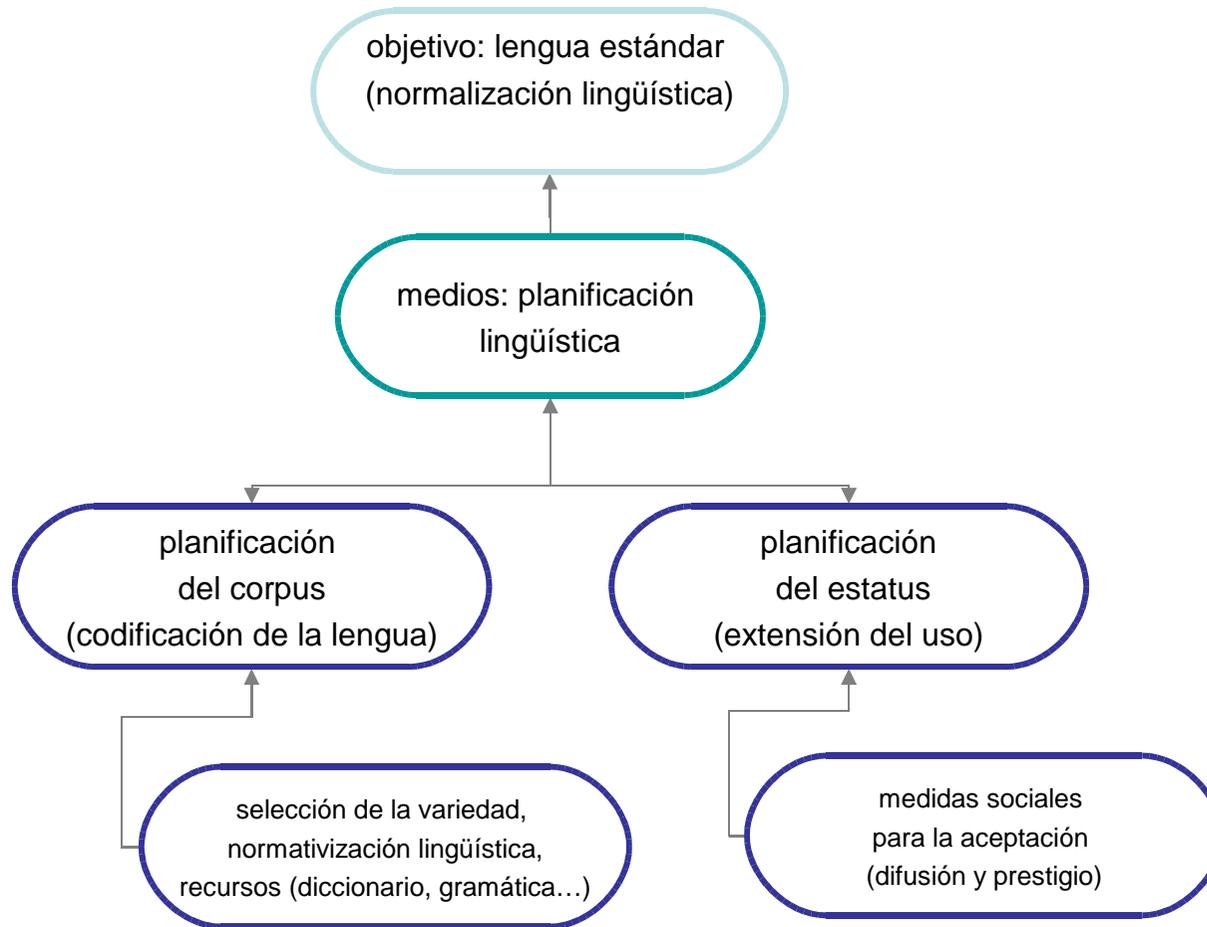
- la **planificación lingüística** se lleva a cabo en dos ámbitos:
  - el ámbito **lingüístico** (planificación del corpus)
  - el ámbito **social** (planificación del estatus)
- la **planificación del corpus** consiste en elegir una variedad de lengua (o un conjunto de formas de varias variedades) a fin de que sirva para todas las funciones sociales de la lengua
- suele estar ligada a la **estandarización** de una lengua, a la elaboración de una normativa ortográfica y gramatical, de un diccionario de referencia y de otros recursos básicos
- en este contexto, *corpus* = *variedad lingüística normativa*



## La planificación del estatus

- la **planificación del estatus** se centra en cambiar el estatus social de una lengua o variedad
- se trata de lograr la **aceptación** por parte de la sociedad del corpus elegido y codificado, y de lograr su **utilización** para todas las funciones de la lengua
- una lengua así **normalizada** actúa como lengua común (facilita la comunicación entre los miembros de la sociedad), como elemento unificador de la nación y como símbolo de independencia en relación con otras naciones (función simbólica)





Universidade de Vigo

# Corpus y planificación lingüística

- usos en la **planificación del corpus**
  - la lingüística de corpus como base teórica y metodológica para la elaboración de **recursos lingüísticos** (gramáticas, diccionarios...)
  - los corpus como **base empírica** de las decisiones normativas
- usos en la **planificación del estatus**
  - aplicaciones de los corpus en las **tecnologías de la lengua**, fundamentales para el uso y prestigio de la lengua
  - corpus para el **aprendizaje** e investigación lingüística
  - la **función simbólica** de los corpus nacionales





# Corpus lingüísticos y planificación del corpus

- **recursos lingüísticos básicos** mejorables con corpus:  
diccionarios y gramáticas
  - diccionarios de frecuencias
  - diccionarios de la lengua, diccionarios de uso
  - diccionarios históricos
  - diccionarios especializados, bancos de datos terminológicos
  - diccionarios de idiomas
  - glosarios y BDT plurilingües
  - gramáticas de la lengua, gramáticas de uso



# Corpus lingüísticos y planificación del corpus

- usos de los corpus en la elaboración de **diccionarios y gramáticas** (1/2)
  - para tratar de determinar qué significados tiene una palabra a partir de sus contextos de uso
  - para ver con qué otras palabras se suele combinar una palabra en los textos
  - para ver en qué construcciones gramaticales se usa una palabra en los textos
  - para descubrir relaciones léxicas y definiciones codificadas en los textos



# Corpus lingüísticos y planificación del corpus

- usos de los corpus en la elaboración de **diccionarios y gramáticas** (2/2)
  - para identificar las posibles traducciones documentadas de una palabra en su contexto
  - para determinar la frecuencia de uso de palabras, construcciones, acepciones y traducciones
  - para poder fundamentar las decisiones sobre normativa de la lengua también en el uso





# Corpus lingüísticos y planificación del corpus

- la **consulta de corpus** como recurso lingüístico básico
  - los corpus **monolingües** como herramienta de consulta
    - para ver el uso real de una palabra en los textos, sus sentidos y sus construcciones gramaticales
    - para palabras que no están en el diccionario
  - los corpus **plurilingües** como herramienta de consulta
    - para ver las traducciones contextualizadas
    - para ver las palabras y las traducciones que no están en el diccionario
    - para ver la posible traducción cuando no hay diccionario bilingüe

**ASCO. m.** La alteración y movimiento fuerte del estómago, causado de la repugnancia y aversión que se tiene à alguna cosa, que provoca à vómito, ò causa notable defazon ò disgusto. Covarr. *discurre* que puede venir de la voz Griega *Zy-bos*, que vale cola lúcia, porque naturalmente mueve à asco; pero mas verisimil parece que se haya formado por la figura Onomatopéya del sonido violento que hace la garganta al tiempo de semejante descomposición y movimiento. Lat. *Nausea*. C. LUCAN. cap. 3. Cuidando que lo facian por *asco* que de él tomaban, comenzó à llorar, y quejarse de gran pesar y quebranto del *asco* que de él ovieron. ALFAR. fol. 211. Haciéndola fuerza la obligó à comerlo contra toda su voluntad y con *asco*.

Es un *asco*. Modo de hablar con que se dá à entender que alguna cosa es lúcia y provocativa à náusea: y tambien se usa de esta phrase quando está tan mal hecha, y tan toscamente executada, que causa fastidio y disgusto el verla. Lat. *Res facta est: Nauseam movet*. ANT. AGUST. Dial. fol. 141. Hai otras piedras tan mal dibujadas, que *es un asco* el verlas cabe las otras.

Hacer *ascos* de alguna cosa. Es despreciarla y reputarla por ruin, y de poco ò ningun aprécio, como por lo contrario: No hai que hacer ascos, es decir no hai que desestimarla, ni hacer desprécio. Lat. *Tadio, nausea affixi. Nauseare. Fastidire*. ALFAR. fol. 213. Si dixéres que *hago ascos* de mi próprio trato.... ò que *hago melindres*, pesaráme de que lo juzgues à tal. QUEV. Muf. 6, Rom. 78.

*Quando ella mirlada hacia ascos, torciendo el hocico.*

## Avance de la vigésima tercera edición

asco.

(De *asqueroso*).

1. m. Alteración del estómago causada por la repugnancia que se tiene a algo que incita a vómito.
2. m. Impresion desagradable causada por algo que repugna.
3. m. Esta misma cosa.

estar hecho un ~.

1. loc. verb. coloq. Estar muy sucio.

hacer ~s.

1. loc. verb. coloq. Hacer afectadamente desprecio poco justificado de algo.

no hacer ~s a algo.

1. loc. verb. irón. coloq. Aceptarlo de buena gana.

ser un ~ algo.

1. loc. verb. coloq. Ser muy indecoroso y despreciable.
2. loc. verb. coloq. Ser muy malo o imperfecto, no valer nada.

## Avance de la vigésima tercera edición

### asco.

(De *asqueroso*).

1. m. Alteración del estómago causada por la repugnancia que se tiene a algo que incita a vómito.
2. m. Impresión desagradable causada por algo que repugna.
3. m. Esta misma cosa.

### estar hecho un ~.

1. loc. verb. coloq. Estar muy sucio.

### hacer ~s.

1. loc. verb. coloq. Hacer afectadamente desprecio poco justificado de algo.

### no hacer ~s a algo.

1. loc. verb. irón. coloq. Aceptarlo de buena gana.

### ser un ~ algo.

1. loc. verb. coloq. Ser muy indecoroso y despreciable.
2. loc. verb. coloq. Ser muy malo o imperfecto, no valer nada.

Maldonado, Concepción (dir.) (1996):  
*CLAVE. Diccionario de uso del español actual*. Madrid: SM.

### asco

#### s.m.

- 1 Impresión desagradable causada por algo que provoca aversión: *Las babosas me dan asco.*
  - 2 Lo que produce esta impresión: *Esta comida es un asco y no hay quien se la coma.*
  - 3 Alteración del estómago causada por algo que resulta muy desagradable, y que generalmente provoca náuseas o vómitos: *Durante su embarazo, los huevos fritos le daban asco.* SINÓNIMO: *repugnancia*
  - 4 col. Lo que se considera mal hecho, de poca calidad o de poco valor: *Este dibujo es un asco y tendrás que repetirlo.*
  - 5 **hacer ascos a algo col.** · Despreciarlo o rechazarlo porque no gusta: *Necesito trabajar y no le haré ascos a ninguna oferta.*
- hecho un asco col.** · En muy malas condiciones físicas o psíquicas: *Esta enfermedad me ha dejado hecha un asco.*
- ETIMOLOGÍA: Del antiguo *usgo* (repugnancia).

**ASCO com.** La alteración y movimiento fuerte del estómago, causado de la repugnancia y aversión que se tiene à alguna cosa, que provoca à vómito, ò causa notable defazon ò disgusto. Covarr. ~~diccionario~~ que puede venir de la voz Griega *Zjebos*, que vale cola fúcia, porque naturalmente mueve à asco; pero mas verisimil parece que se haya formado por la figura Onomatopéya del sonido violento que hace la garganta al tiempo de semejante descomposición y movimiento. Lat. *Nausea*. C. LUCAN. cap. 3. Cuidando que lo facian por *asco* que de él tomaban, comenzó à llorar, y quejarse de gran pesar y quebranto del *asco* que de él ovieron. ALFAR. fol. 211. Haciéndola fuerza la obligó à comerlo contra toda su voluntad y con *asco*.

Es un *asco*. Modo de hablar con que se dá à entender que alguna cosa es fúcia y provocativa à náusea: y tambien se usa de esta phrase quando está tan mal hecha, y tan toscamente executada, que causa fastidio y disgusto el verla. Lat. *Res fæda est: Nauseam movet*. ANT. AGUST. Dial. fol. 141. Hai otras piedras tan mal dibujadas, que *es un asco* el verlas cabe las otras.

Hacer *ascos* de alguna cosa. Es despreciarla y reputarla por ruin, y de poco ò ningun aprécio, como por lo contrario: No hai que hacer ascos, es decir no hai que desestimarla, ni hacer desprécio. Lat. *Tadio, nausea afflii. Nauseare. Fastidire*. ALFAR. fol. 213. Si dixéres que *hago ascos* de mi próprio trato.... ò que hago melindres, pesaráme de que lo juzgues à tal. QUEV. Mus. 6, Rom. 78.

*Quando ella mirlada hacia ascos, torciendo el hocico.*

Marcial Valladares Núñez  
(1884): *Diccionario gallego-castellano*. Santiago: Tipografía del Seminario Conciliar.

**ASCO com.** Alteracion del estómago, por la repugnancia que se tiene à alguna cosa. Hastío. Fastidio.

Real Academia Española (1899<sup>13</sup>):  
*Diccionario de la lengua castellana*.  
Madrid: Hernando y Cía.

**Balar.** (Del lat. *balāre*.) n. Dar balidos. ||  
fig. y fam. Seguido de la prep. *por*, desear  
con ansia alguna cosa.

**Balastar.** a. Tender el balasto.

**Balasto.** (Del ingl. *ballast*, lastre.) m. Capa  
de grava ó de piedra machacada, que se tien-  
de sobre la explanación de los ferrocarriles  
para asentar y sujetar sobre ella la vía.

**Balata.** (Del ital. *ballata*; del lat. *ballāre*, bai-  
lar.) f. Composición poética que en un prin-  
cipio se hacía para ser cantada al son de la  
música de los bailes.

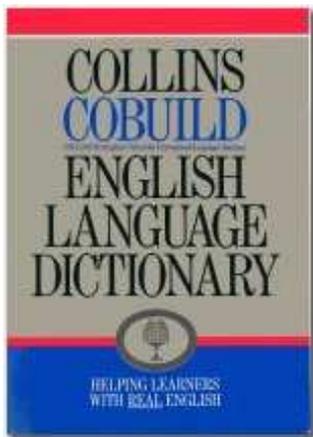
**Balate.** (Del ár. *بلاط*, *balat*, suelo allanado.)  
m. Parata cuyo corte exterior no está forma-  
do por hiladas artificiales de piedra, sino  
que, naturalmente ó por trabajo del hombre,

Real Academia Galega (1913-  
1928): *Diccionario gallego-  
castellano [a-cativo]*. A Coruña:  
Imprenta y fotograbado de  
Ferrer.

**BALATA** s. f. Balata.  
Composición poética que  
se cantaba al son de la  
música de los bailes

Eladio Rodríguez González  
(1958-1961): *Diccionario  
enciclopédico gallego-  
castellano*, Vigo: Galaxia.

**BALATA** s. f. Composición  
poética que se cantaba al  
son de la música de los  
bailes.



## analyst

46

*right track and I think in the final analysis people will understand that... Violence in the last analysis produces more violence.*

**analyst** /æˈnəlɪst/ (**analysts**) **1** An **analyst** is a person whose job is to analyse a subject and give opinions about it. □ *...a political analyst.*

◆◆◆  
N-COUNT

**2** An **analyst** is someone, usually a doctor, who examines and treats people who are emotionally disturbed.

N-COUNT  
= psycho-  
analyst

**analytic** /æˈnəlɪtɪk/ **Analytic** means the same as **analytical**. [mainly AM] ADJ

**analytical** /æˈnəlɪtɪkəl/ **1** An **analytical** way of doing something involves the use of logical reasoning. □ *I have an analytical approach to every survey.* ◆ **analytical|ly** /æˈnəlɪtɪkli/ A teacher can encourage children to think **analytically**.

ADV:  
ADV with v,  
ADV adj  
ADJ: ADJ n

**2** **Analytical** research involves using chemical analysis. □ *All raw materials are subjected to our latest analytical techniques.*

**analyze** /æˈnəlaɪz/ → see **analyse**.

**anarchic** /æˈnɑːrkiːk/ If you describe someone or something as **anarchic**, you disapprove of them because they do not recognize or obey any rules or laws. □ *...anarchic attitudes and complete disrespect for authority.*

ADJ:  
usu ADJ n  
disapproval

**anarchism** /æˈnɑːkɪzəm/ **Anarchism** is the belief that the laws and power of governments should be replaced by people working together freely. N-UNCOUNT

**anarchist** /æˈnɑːkɪst/ (**anarchists**) **1** An **anarchist** is a person who believes in anarchism. □ *...a well-known anarchist poet.*

N-COUNT:  
oft N n

**2** If someone has **anarchist** beliefs or views, they believe in anarchism. □ *He was apparently quite converted from his anarchist views.* **3** If you say that someone is an **anarchist**, you disapprove of them because they seem to pay no attention to the rules or laws that everyone else obeys. □ *He was a social anarchist.*

ADJ: ADJ n

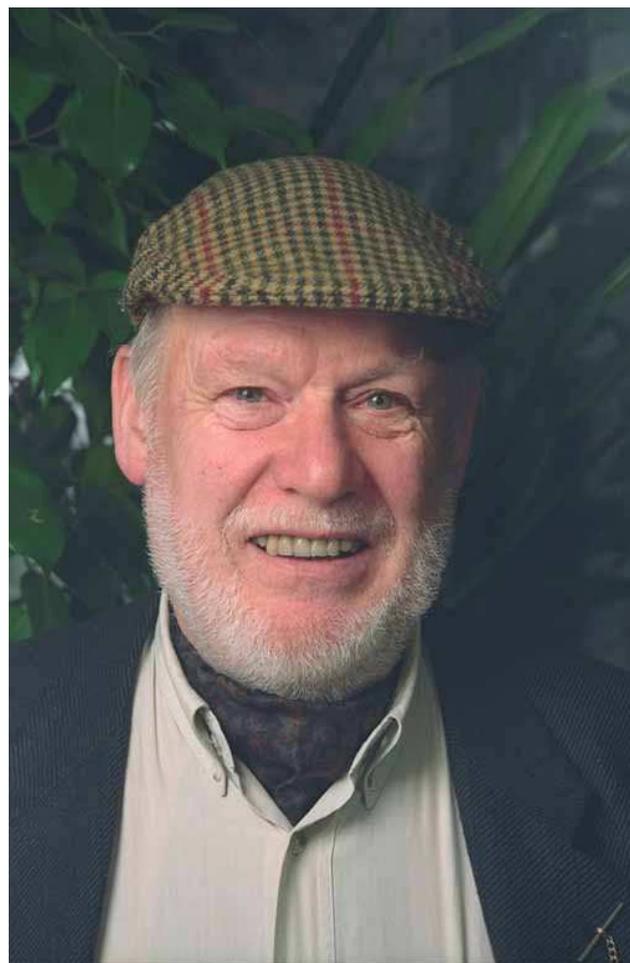
N-COUNT  
disapproval

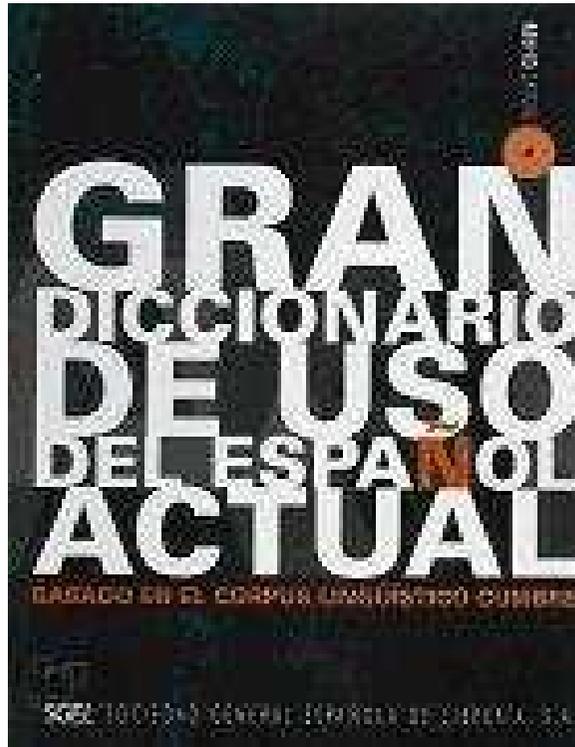
**anarchistic** /æˈnɑːkɪstɪk/ **1** An **anarchistic** person believes in anarchism. **Anarchistic** activity or literature promotes anarchism. □ *...an anarchistic revolutionary movement.*

ADJ:  
usu ADJ n

**2** If you describe someone as **anarchistic**, you disapprove of them because they pay no attention to the rules or laws that everyone else obeys. □ *The Hell's Angels were once the most notorious and anarchistic of motorbike gangs.*

ADJ:  
usu ADJ n  
disapproval





Universidade de Vigo



Corpus Textual Informatitzat de la Llengua Catalana - Mozilla Firefox

Ficheiro Editar Ver Histórico Marcadores Ferramentas Ajuda

http://ctilc.iec.cat/

Institut d'Estudis Catalans

Corpus Textual Informatitzat de la Llengua Catalana

Diccionari del Català Contemporani

Presentació Instruccions Usuari: Anònim

Selecció de lemes i formes / Llistat de concordances

Reconstrucció de context

la actitud d'en Mañé y del "Diari" se pot condensar en aquests	<b>termes:</b> --Això era previst; això es un fet; el deure de tots es durlo
per correspondre a la de l'artista en els següents precisos	<b>termes:</b> --Aneu-hi vos, que ara tinc feina.-- L'artista --fos un dels
ha estat emprat per tal d'obtenir la imputació territorial, en	<b>termes:</b> --ara sí-- d'incidència, de la despesa pública més clarament
de respondre la pregunta: --On va Catalunya?, amb aquests	<b>termes:</b> --Catalunya, cansada de suportar la tirania d'Espanya, cansada
quan cal presentar l'evocació d'objectes situats a diferents	<b>termes:</b> --com passa en el paisatge-- el pintor hagi cercat els efectes
encara tens ocasió d'eixir-te'n. --Planteja-m'ho en altres	<b>termes:</b> --contestà el gerent, encetant un punt de negociació--. Tu
en les seves explicacions. I assentava un peu damunt dels	<b>termes:</b> --dels termes que es trobaven en llocs enlairats--, i estenia el
escuradents d'entre geniva i queixal, se'm dirigís en aquests	<b>termes:</b> --I de què va morir la vostra mare? O bé: --I amb què compteu
fitxa solemne: significa la taxativa "creació" del país en els	<b>termes:</b> --ja que no amb els límits-- en què avui subsisteix. No va
nocturna) m'estava esperant, i m'adreçà la paraula en aquests	<b>termes:</b> --Jove, em sap greu de veure la vostra davallada. Però quina

Pàgina 1 / 407 ( Nombre de resultats: 4062 ) > >>

[Requisits tècnics](#) C. del Carme, 47. 08001 Barcelona | Telèfon +34 932 701 620 / Fax +34 932 701 180 | [dcc@iec.cat](mailto:dcc@iec.cat)

**Avis legal:** Són rigorosament prohibides, sense l'autorització escrita dels titulars del copyright, l'extracció, la reutilització i la reproducció en qualsevol mena de suport o el tractament informàtic del contingut de les dades del CTILC. També n'és prohibit el lloguer, el préstec i l'accés no autoritzat per xarxa telemàtica. Les infraccions d'aquests drets estan sotmeses a les sancions establertes per la llei.

javascript:C.setPropertyValue("idLocali","7807668");C.invokeMethodInModel("veureFitxa");




# EUSKALTZAINDIA

Bilatu

>> EUSKARA >> ESPAÑOL >> FRANÇAIS >> ENGLISH

- ▶ ERAKUNDEA
- ▶ ARGITALPENAK
- ▶ AZKUE BIBLIOTEKA
- ▶ HIZKUNTZA-KONTSULTAK
- ▶ IKER/JAGON-TEGIAK
- ▶ E-MAIL



## Orotariko Euskal Hiztegia

>> Hiztegia bilatu

>> Argibideak

>> Jaisteko gunea

**galiziar**

**galiziar** (Lar), *galiziatar, galizitar.*

♦ Gallego. v. **gallego (2)**, *galiziano.* \* *Ausaz Galiziarren bat da gure zimaur zale au? Lar DT CCXVIII. Bonboa esaten da gure itxaserrietan Galiziarren Falucho, ingelesen Vessel-gaiti. A BGuzur 141. Galizitar erari saltzallea. Ag G 125. Galiziar bati cuarénticincu entzuten badiote, irri egiñen diote. Or in Gazt MusIx 16. Euskaldun semea galizitarren etxean oso ondo artua izan zan. NETx LBB 46. Cela galiziarra da. MIH 362. v. tbn. Berron Kijote 184 (galiziatar).*



Orrialde honetako dokumentu, zerrenda eta datu-baseak **Creative Commons** Aitortu-EzKomertziala-PartekatuBerdin baimen baten mende daude.

BUSCAR

nova busca

desde  ata

Ano	Autor - Obra	Paxina/Linha	Palabra
1853	<a href="#">PINGAI853</a>	71 19	gallego; ou pon+me en castellano pola túa régoa estes gallegos <b>termos</b> «aceite», «seis» e «rei», e máis aínda o «peite» con
1853	<a href="#">PINGAI853</a>	24 7	por exemplo, ou digan que «taberna», «pan», e «carne» son <b>termos</b> castellanos non gallegos: aqueles que sentenzan que
1853	<a href="#">PINGAI853</a>	147 0	Conque logo ¿dis que moitos non entenden eses <b>termos</b> , e que outros en castellano non acertan a poñé-los?
1853	<a href="#">PINGAI853</a>	100 18	medio do adro aquel trompeteiro publicar as novas nestes mesmos <b>termos</b> : Xa os mouros van de vencida polos nosos mariñeiros que
1853	<a href="#">PINGAI853</a>	34 19	toda esa carta en gallego tomando ben o sentido das súas frases e <b>termos</b> ? ¿Ou fas como <3*Catuxiña3> que andou a ler muito tempo e
1858	<a href="#">PINCON858</a>	176 1	da Verga3> mui tenro contestou+lle de contado seica nestes mesmos <b>termos</b> : Non temas, miña ruliña, que falte ó prometimento, que
1858	<a href="#">PINCON858</a>	336 25	ollos se lle enchendo, mira para o seu amante falando+lle nestes <b>termos</b> ; Cala <3*Pedro3>, non agoires prenósticos tan funestos,
1858	<a href="#">PINCON858</a>	150 40	o que din os mariñeiros cantando polo peirán así nestes mesmos <b>termos</b> ; Non che se dea, miniña, polo que a xente marmula
1861	<a href="#">PINGAL861</a>	365 62	1> está reinante, e nos códigos mesmos castelanos a miles hai os <b>termos</b> galicianos.
1861	<a href="#">IGSGAL861</a>	371 59	do catro de setembro do ano de trece, falou dos gallegos nestes <b>termos</b> : «Guerreiros do mundo cevilizado: dependede a sé-lo
1881	<a href="#">PICCAR881</a>	249 33	frondosas ramas, deitados s'o tenro e verde mosgo, recrebado con <b>termos</b> melosos ás zagás, guapas com'estrelas e fachendosas;
1882	<a href="#">POPPOE882</a>	55 1	o laverco que só quer beillar «dancitas» como el latrica en seus <b>termos</b> ; Máis valera alá non fora, traballara o seu terreo, e
1886	<a href="#">gal111886</a>	211 3	carpinteiro, e pondo+lle diante a víutema, di+lle mesmo nistes <b>termos</b> : -Quero te encargues ti de faguer iste enterramento
1888	<a href="#">FULO17888</a>	4 7	que non hai, escomezarei a hestoria ou conto no mesmo orden e <b>termos</b> qu'emprou o bo do vello, cando eu agarrando+me con
1888	<a href="#">VATNOI888</a>	76 11	que non hai, escomezarei a hestoria ou conto no mesmo orden e <b>termos</b> qu'emprou o bo do vello, cando ou agarrando+me con
1889	<a href="#">LABBAL889</a>	141 6	E soilo, alá de vés en cando, mira d'ermo sin fin antr'os escuros <b>termos</b> , outras qu'atrás lle berran: iredemámo-nos! con vista
1889	<a href="#">LABBAL889</a>	41 2	por un gallego!- con perdón da Tauromaquia, escomecei nestes <b>termos</b> : <1*Noia1>.- Agosto vintecatro.
1890	<a href="#">LVABRI890</a>	78 7	tódo-los demos, encarando+s'hacia á porta falou+lles en estos <b>termos</b> : -Eso é non saber crianza, meus mociños de <1*Louredo1>;
1891	<a href="#">TÍAO18891</a>	4 28	se tira ¡puñesflero! ás vellas, cicais por estarmos encorreadas e <b>termos</b> algún ballón, tira+se, soilo, ás mociñas imá centella o
1891	<a href="#">BRCCON891</a>	216 2	de que as costumes canónicas, ca axuda de Dios, reducen+se ós <b>termos</b> paternaes mediante nosa groria; pero ante todo
1891	<a href="#">TÍAO18891</a>	213 6	Iba moi tenra, como que os <b>termos</b> todos de que se compuña saían+lle da ialma; e non tiña
1891	<a href="#">TÍAO18891</a>	4 11	no pior das súas apuradas, persentou+se a el e falou+lle nestes <b>termos</b> : -<3*Xan3>, disque vas a morrer? -¡Sí, muller... sí!
1891	<a href="#">BRCCON891</a>	215 13	, o mesmo santísimo Príncipe, falou ó venerable Concilio nestes <b>termos</b> : «Non penso, reverendísimos Sacerdotes, que
1894	<a href="#">RIBFER894</a>	11 20	chame isto, qu'eu do nome non coído, veu cando debía polos seus <b>termos</b> contados, e non é, pese a quen pese, fogo fatuo nascido
1894	<a href="#">SALLEN894</a>	13 13	de chegando o momento» «de escomezar xogo e baila» «despois de <b>termos</b> o choio» «chegado ó fin da xornada», «vostede acabando



40 zurekin



EUSKALTAINDIA  
REAL ACADEMIA DE LA LENGUA VASCA  
ACADÉMIE DE LA LANGUE BASQUE

## Diccionario CLUVI Inglés-Galego

 2ª edición

Xavier Gómez Guinovart (coord.)

Equipo de redacción:

Alberto Álvarez Lugrís  
Eva Díaz Rodríguez  
Xavier Gómez Guinovart



Seminario de Lingüística Informática  
Universidade de Vigo

## Diccionario CLUVI inglés-galego

- Diccionario elaborado a partir de los datos del Corpus CLUVI, con 20.000 entradas, 30.000 traducciones y 60.000 ejemplos documentados en el corpus
- De libre consulta en **<http://sli.uvigo.es/diccionario>**



Universidade de Vigo

# Dicionario CLUVI inglés-galego, 2ª edición



(Baseado no Corpus CLUVI da Universidade de Vigo)

Seminario de Lingüística Informática, 2005-2008  
Universidade de Vigo

[Portada](#)

[Limiar](#)

[Nova consulta](#)

[tentacle](#), [tentative](#), [tenth](#), [tenuous](#), [tenure](#), [tepee](#), [tepid](#), [tequila](#), [tercentenary](#), [term](#), [termagant](#),  
[terminal](#), [terminate](#), [termination](#), [terminology](#), [termite](#), [terrace](#), [terracotta](#), [terrain](#)

## term

■ noun

✓ trimestre

**EN** *First came the vacation and then the next **term** and then vacation again and then again another term and then again the vacation.*

**GL** *Primeiro virían as vacacións e despois o seguinte **trimestre**, logo volta coas vacacións e outro trimestre e máis vacacións.*

► Fonte: RET (259)

✓ termo

**EN** *There seemed to be few, if any, abstract **terms**, or little use of figurative language.*

**GL** *Parecía haber poucos **termos** abstractos, se é que os había, e tampouco se usaba a lingua figurada.*

► Fonte: TEM (877)

...

&lt;tu&gt;

<tuv xml:lang="en"><seg>The Time Traveller (for so it will be convenient to speak of him) was expounding a recondite matter to us.</seg></tuv>

<tuv xml:lang="gl"><seg>O Viaxeiro do Tempo (este será o xeito máis conveniente de referírmonos a el) estaba expoñéndonos un problema cargado de misterio para nós.</seg></tuv>

&lt;/tu&gt;

&lt;tu&gt;

<tuv xml:lang="en"><seg>His grey eyes shone and twinkled, and his usually pale face was flushed and animated.</seg></tuv>

<tuv xml:lang="gl"><seg>Os seus ollos grises brillaban e escintilaban e víaselle a cara, normalmente pálida, acorada e animada.</seg></tuv>

&lt;/tu&gt;

...

## Corpus CLUVI

- Corpus paralelo, de traducciones al gallego o del gallego, con 23 millones de palabras
- Literatura, derecho, informática, *Consumer Eroski*, *El Correo de la Unesco*
- De libre consulta en <http://sli.uvigo.es/CLUVI>



# Pescudas no Corpus Paralelo CLUVI



(Corpus Lingüístico da Universidade de Vigo)

<a href="#">Como facer buscas</a>	<a href="#">Obras dispoñíbeis</a>	<a href="#">Máis información</a>	<a href="#">Corpus Técnico do Galego</a>	<a href="#">In English</a>
-----------------------------------	-----------------------------------	----------------------------------	--	----------------------------

## Seccións principais do CLUVI:

- Corpus xurídico LEGA galego-español (6.582.415 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus UNESCO inglés-galego-francés-español de divulgación científica (3.724.620 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus LOGALIZA de localización de software inglés-galego (3.526.850 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus literario TECTRA inglés-galego (1.476.020 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus literario FEGA francés-galego (1.648.272 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus CONSUMER español-galego-catalán-euskara de información sobre consumo (5.586.431 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus xurídico LEGE-BI euskara-español (2.384.053 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus TURIGAL de turismo portugués-inglés (1.285.764 palabras) [[Buscar](#)] [[Composición](#)]

## Seccións do CLUVI en construción:

- Corpus EGAL de economía galego-español (718.642 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus literario TECTRA inglés-portugués (875.595 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus literario TECTRA inglés-español (122.251 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus literario DEGA alemán-galego (76.364 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus VEIGA de subtítulos de películas inglés-galego (285.645 palabras) [[Buscar](#)] [[Composición](#)]
- Corpus PALOP de literatura poscolonial portugués-español (566.590 palabras) [[Buscar](#)] [[Composición](#)]



Universidade de Vigo

## Resultados da pescuda no Corpus Paralelo CLUVI



(Corpus Lingüístico da Universidade de Vigo)

Os resultados das buscas efectuadas no CLUVI poden ser usados con fins educativos e de investigación, sempre que se mencione a fonte. Se desexa citar exemplos tirados do corpus, calque na ligazón que hai na esquerda de cada equivalencia para obter a súa referencia completa. A cela da esquerda da táboa contén o número de orde do exemplo, seguido do código da obra mais o número da unidade de tradución da equivalencia no texto aliñado. Se desexa consultar o contexto (frase anterior e posterior) dunha equivalencia, calque na frecha da dereita. Para se referir ao corpus como un todo, cite: *Corpus Paralelo CLUVI 2.1* - <http://sli.uvigo.es/CLUVI/>. Se desexa realizar outra pescuda no Corpus Paralelo CLUVI, pode calcar [aquí](#).

Está a procurar equivalencias de tradución **inglés - galego** no Corpus literario TECTRA inglés-galego (1.476.020 palabras).  
Equivalencia de tradución buscada: [en] \bterm\b = [gl] .

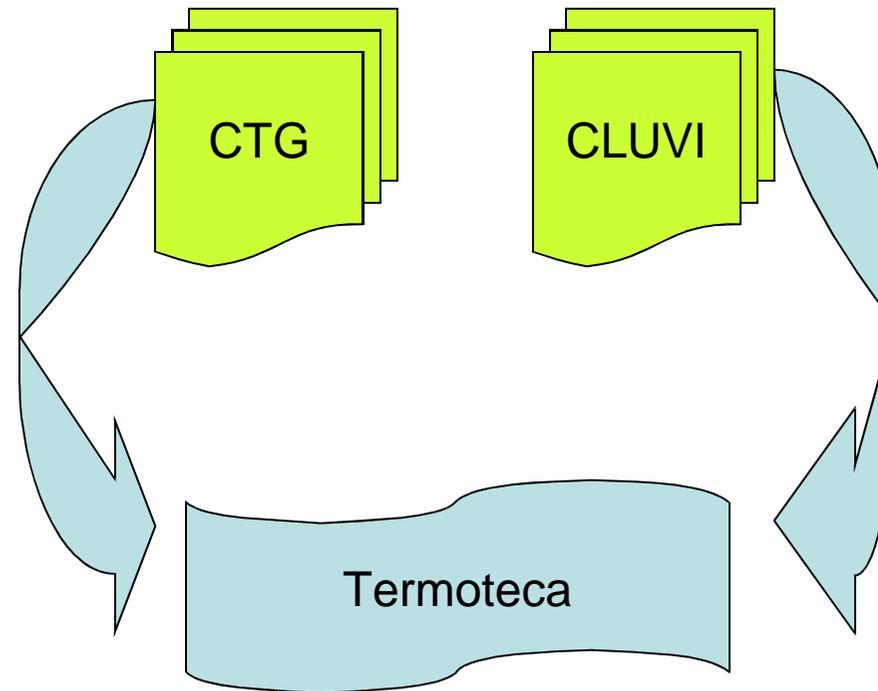
1-TER (209)	I met him my first <b>term</b> at school.	Encontrámonos no primeiro curso da escola.	
2-RET (259)	First came the vacation and then the next <b>term</b> and then vacation again and then again another <b>term</b> and then again the vacation.	Primeiro virían as vacacións e despois o seguinte trimestre, logo volta coas vacacións e outro trimestre e máis vacacións.	
3-RET (261)	<b>Term</b> , vacation; tunnel, out; noise, stop.	Trimestre, vacacións, túnel, fóra; ruído, alto.	
4-RET (1658)	It was towards the close of his first <b>term</b> in the college when he was in number six.	Fora cara ao final do seu primeiro trimestre no colexio cando estaba no sexto.	
5-RET (4061)	The esthetic emotion (I used the general <b>term</b> ) is therefore static.	A emoción estética (empreguei o termo xeral) é, polo tanto, estática.	
6-RET (4243)	Aquinas uses a <b>term</b> which seems to be inexact.	Aquino emprega un termo que semella inexacto.	
7-RET (4258)	...re speak of beauty in the second sense of the <b>term</b> our judgement is ... ...ed in the first place by the art itself and by the form of that art.	Cando falamos da beleza neste senso, o noso xuízo está influenciado en primeiro lugar pola propia arte e a súa forma.	



Universidade de Vigo

# Termoteca

- Banco de datos terminológico de la Universidad de Vigo
- 7.137 registros con información sobre 13.752 términos extraída de los corpus CLUVI y CTG
- De libre consulta en <http://sli.uvigo.es/termoteca>



# Termoteca - Banco de Datos Terminolóxico da Universidade de Vigo



Baseado no Corpus Técnico do Galego e no Corpus CLUVI

---

Procurar termo:	<input type="text" value="invernadoiro"/>
Lingua:	<input type="text" value="Galego"/>
No ámbito:	<input type="text" value="Todos os ámbitos"/>
Patrón categorial:	<input type="text" value="Calquera"/>
	<input type="checkbox"/> Ver só neónimos
	<input type="button" value="Vai --&gt;"/>

---

Explorar áreas:	<input type="text" value="Seleccionar área"/>
	<input type="button" value="Vai --&gt;"/>



---

Universidade de Vigo

<i>Ref.</i> 666406	
<i>Campo temático:</i>	< <a href="#">AUGA.3.2.3.3/3.2.3.3. Gases contaminantes.</a> > < <a href="#">AUGA.3.2.1.2/3.2.1.2. Axentes</a> >
<i>Definición:</i>	Os <a href="#">gases de efecto invernadoiro</a> son compoñentes gasosos da atmosfera, tanto naturais como antropoxénicos, que absorben e emiten radiación en determinadas lonxitudes de onda do espectro de radiación infravermella emitido pola superficie da Terra, a atmosfera e as nubes. - <a href="#">POF</a>
<i>Relacións semánticas:</i>	[caus => <a href="#">66673</a> ] [hipo => <a href="#">666405</a> ]
<i>Termo GL:</i>	<b>gas de efecto invernadoiro</b>
<i>Categoría:</i>	m
<i>Morfosintaxe:</i>	gas/gas_NCMS0 de/de_S efecto/efecto_NCMS0 invernadoiro/invernadoiro_A0MS
<i>Variante:</i>	com
<i>Frecuencia relativa:</i>	184.307631145
<i>Contexto de uso:</i>	O metano (CH <sub>4</sub> ) é un <i>gas de efecto invernadoiro</i> moito máis potente có dióxido de carbono e, aínda que a súa concentración na atmosfera é relativamente baixa, está aumentando en case un 1% cada ano. [A523]
<i>Máis exemplos:</i>	<a href="#">No CTG</a> (algún dos exemplos do termo no corpus poden non corresponder ao concepto da ficha)
<i>Termo GL:</i>	<b>gas invernadoiro</b>
<i>Categoría:</i>	m
<i>Morfosintaxe:</i>	gas/gas_NCMS0 invernadoiro/invernadoiro_A0MS
<i>Variante:</i>	morf
<i>Frecuencia relativa:</i>	1.70261091127
<i>Contexto de uso:</i>	O dióxido de carbono (CO <sub>2</sub> ) é o principal <i>gas invernadoiro</i> e emítese, de forma inevitábel, ao queimarmos combustíbeis fósiles (carbón, petróleo e gas). [A031]
<i>Máis exemplos:</i>	<a href="#">No CTG</a> (algún dos exemplos do termo no corpus poden non corresponder ao concepto da ficha)

...

<frase>As medidas que se deben adoptar para o proceso de descarga, moedura e preparación do alimento son: </frase>

<frase>Os materiais pulverulentos ou potencialmente pulverulentos débense almacenar en colectores, en silos deseñados para este propósito ou baixo cuberta.</frase>

<frase>A distribución do alimento dende o lugar de almacenamento ata as instalacións onde comen os animais debe levarse a cabo de tal maneira que se prevexan ou minimicen as emisións ao aire de partículas, é dicir, preferentemente con sistemas pechados.</frase>

<frase>A moedura e a mestura do alimento haberá que realizala de tal forma que se prevexan ou minimicen as emisións de partículas ao aire.</frase>

<frase>Os tanques de almacenamento deberán estar protexidos de posibles danos por colisión.</frase>

...

## Corpus Técnico do Galego (CTG)

- Corpus textual de registros especializados del gallego contemporáneo, de 14,5 millones de palabras
- Derecho, informática, sociología, economía, ecología y medicina
- De libre consulta en **<http://sli.uvigo.es/CTG>**





## Corpus Técnico do Galego (CTG)

### Pescudas no CTG

---

Palabra ou expresión en galego:

Busca no CTG

Cinguir a busca a un só ámbito do CTG:

- Dereito
- Ecoloxía
- Economía
- Informática
- Socioloxía
- Medicina



## Corpus Técnico do Galego



### Resultados da pescuda

Os resultados das buscas efectuadas no CTG poden ser usados con fins educativos e de investigación, sempre que se mencione a fonte. Se desexa consultar a referencia e o contexto dun exemplo, calque no símbolo [+] na cela da dereita. Para se referir ao corpus como un todo, cite: *Corpus Técnico do Galego (CTG)* - <http://sli.uvigo.es/CTG/>. Se desexa realizar outra pescuda no Corpus CTG, pode calcar [aquí](#).

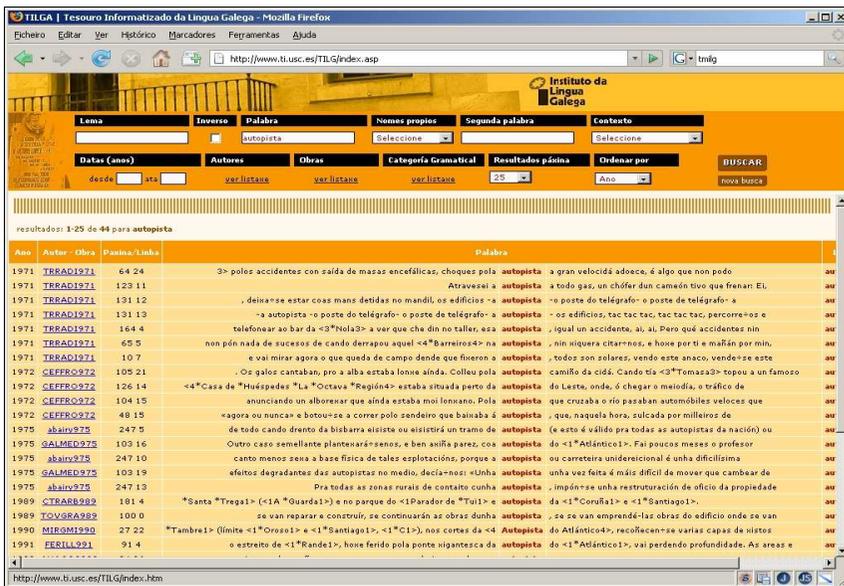
Está a procurar contextos do uso de **gas de efecto invernadoiro** nos textos de ecoloxía e ciencias ambientais do Corpus Técnico do Galego (CTG).  
Número de contextos atopados: 5

1 - [1018/119]	l) Tonelada equivalente de dióxido de carbono: unha tonelada métrica de dióxido de carbono (CO <sub>2</sub> ) ou unha cantidade de calquera outro <b>gas de efecto invernadoiro</b> recollido no anexo II cun potencial equivalente de queentamento do planeta.	[+]
2 - [314/80]	Mentras que o desenvolvemento da silvicultura supón un beneficio global, pois potencia o secuestro de CO <sub>2</sub> nos bosques que actúan como sumidoiros deste <b>gas de efecto invernadoiro</b> (se a madeira e os seus subproductos non se queiman despois), as plantacións selvícolas exercen unha influencia de sentido e intensidade variables, según a(s) especie(s) e o manexo, sobre a calidade e conservación do solo, os ciclos de auga e nutrientes, as especies da flora e fauna silvestres e a paisaxe.	[+]
3 - [333/270]	<del>Vemos agora os factores de emisión utilizados en función do combustible para cada <b>gas de efecto invernadoiro</b> (táboas 49, 50 e 51).</del>	[+]
4 - [516/24]	O metano (CH <sub>4</sub> ) é un <b>gas de efecto invernadoiro</b> moito máis potente có dióxido de carbono e, aínda que a súa concentración na atmosfera é relativamente baixa, está aumentando en case un 1% cada ano [18].	[+]
5 - [523/30]	<del>O metano (CH<sub>4</sub>) é un <b>gas de efecto invernadoiro</b> moito máis potente que o dióxido de carbono e, aínda que a súa concentración na atmosfera é relativamente baixa, está aumentando en case un 1% cada ano.</del>	[+]





# Tesouro informatizado da lingua galega



- Corpus etiquetado y lematizado del gallego moderno de 20 millones de palabras
- Contiene casi la totalidad de las obras en gallego entre 1612 y 1980
- De libre consulta en la web en <http://www.ti.usc.es/TILG> y <http://sli.uvigo.es/TILG>



Universidade de Vigo

1853	PINGAI853	185 19	Mui enfiados en fronteiras ringles fan sitio os <b>árbores</b> para o
1858	GAMPOE858	83 5	ist ' aire susurrante qu ' os <b>arboriños</b> peina , a púrpura bicou da
1858	PINCON858	336 11	ceo , nos <b>árbores</b> , e nas frores , no campo , no mar inmenso , na
1861	FMOENS861	207 11	dos subiaos , follas d ' <b>árboles</b> e feitos , e con pallas dos rastrollos
1861	FMOENS861	258 5	A vida dos mentecatos pasaba , ós <b>árboles</b> trochos tirando ; murando
1861	IGSGAL861	370 45	sangue a <b>arbre</b> da relixión acó prantada polo fillo do <3 Zebedeu 3> ,
1861	IGSGAL861	370 46	<b>arbre</b> qu ' arraigou fortemente e deu froitos escolleitos de bendición
1878	PRZVER878	17 1	con negra furia o ramalliño de vès ; e aquel <b>arbre</b> , que agarimo
1878	PRZVER878	20 0	¡ Cómo decir ÷che , meniña , que contra o <b>arbre</b> querido un labrego
1878	PRZVER878	21 4	do <b>arbre</b> , do rego e frores , ¡ feitos do voso afán ! Praceres
1880	LAMSAU880	36 3	<b>álbores</b> apacibre sombra .
1880	LAMSAU880	63 6	na terra nosa son estos <b>álbores</b> símbolo da inmortalidad ' e gloria ,
1881	PICCAR881	248 28	i <b>álbores</b> de toda cras de frutas , peras , mazás , castañas , noces ,
1882	IGLFON882	7 1	AUTO PRIMEIRO Decoración de praia con <b>albres</b> á dereita e esquerda .
1882	IGLFON882	7 3	da esquerda do fondo . Unha <b>albre</b> no escomenzo da subida . Casa
1883	MTNPOE883	71 5	Ciruxán , ou manate con guedellas Das raspas do xa podre <b>árbol</b>
1886	CURAIR886	5 13	en qu ' as fontes solouzan e marmullan entr ' os follosos <b>álbores</b>
1886	CURAIR886	27 13	<b>álbores</b> i as prantas esfollaron .



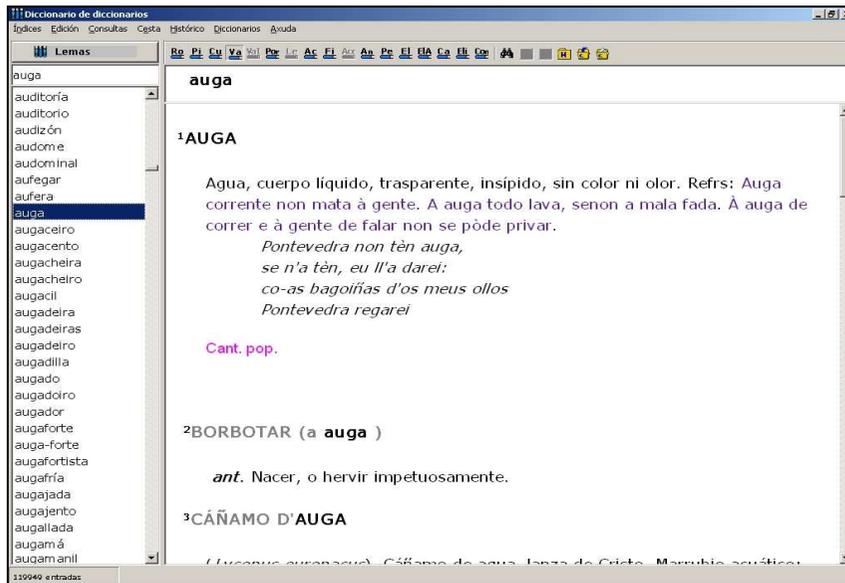
<b>árbore (1006) - árbores (1355)</b>	<b>loureiro (492) - loureiros (262)</b>	<b>ámbar (39) - ámbares (5)</b>
arbre (275) - arbres (466)	laureiro (11) - laureiros (16)	ambre (6) - ambres (4)
albre (194), albres (263)	lourel (9) - loureles (1)	ambra (3)
árbol (121) - árboles (75)	doureido (1)	
álbor (4), álборе (96) - álbores (183)	lureiros (1)	
álbere (2) - álberes (2)		
arbro (2)		
ialbre (1)		





# Diccionario de diccionarios

- Recopilación de 25 obras lexicográficas de los siglos XIX y XX
- Permite la consulta de más de 300.000 entradas
- Disponible en CD-ROM y en línea en <http://sli.uvigo.es/DdD>



Corpus lexicográfico da lingua galega

Está a procurar a palabra **arbo??** como *lema* no *Diccionario de Dicionarios*, sen ter en conta a acentuación da palabra nin a diferenza *b/v*.

- Número de acepcións atopadas: 96.

- Distribución por dicionarios: Sarmiento(1746-1755c) (2), Sarmiento(1754-1758) (16), Sobreira(1790c) (2), Sobreira(1792-1797) (9), Sobreira(1794) (2), Pintos(1865c) (1), Valladares(1884) (3), Porto(1900c) (7), RAG(1913-1928) (8), Filgueira(1926) (1), Carré(1928-1931) (4), Acevedo(1932) (1), Ibañez(1950) (2), Eladio(1958-1961) (18), Franco(1972) (3), Panisse(1977) (2), Carré(1979) (5), CGarcía(1985) (3), Losada(1992) (5), Sarmiento(d1762) (2).

Se desexa realizar outra pescuda, pode calcar [aquí](#).

**Martín Sarmiento (1746-1755c): Catálogo de voces y frases de la lengua gallega, ed. de J. L. Pensado Tomé (U. de Salamanca en 1973)**

**arbor**  
**fria**

Llaman así hacia **Orense** a la especie de álamo, que tiembla con sus hojas; y se llama *arbor fria*, porque, aunque haga infinito calor siempre tiembla. CatálogoVF 1745-1755

**árbol de**  
**San**  
**Gregorio**

Dicen es árbol único en Galicia y que sólo se halla en un monte de Deza, camino de Santiago, y que un peregrino le descubrió la propiedad que tomando en bebida el palo, hojas, etc., en infusión, cura el mal de la rabia, y la preserva. Es común, y tienen este palo de prevención en las casas. Es poco mayor que un sauce, y las hojas algo parecidas al cerezo. Oí que lo había en Millaráda, hacia Sotelo. CatálogoVF 1745-1755

- (1) **álbore**. *s. f.* Corrupción popular de ÁRBORE. Árbol. V. ÁRBORE. (exs. de Lamas e Curros).
- (2) **albre**. *s. f.* Contracción de *álbore*. Árbol. V. ÁRBORE.
- (3) **árbol**. *s. f.* “Árbol. Árbole. Tiene el árbol: *raíz, toro, polas, pingólas, follatos, gamállos, fillos, codia, madeira, leña, páos o páus, garabullos, folla, gromos, canas, pé, ramallos.*” P. SOBREIRA. V. ÁRBORE.
- (4) **árbole**. *s. f.* “Árbol”. P. SOBREIRA. V. ÁRBORE.
- (5) **árbor**. V. ÁRBORE.- “Et as seetas et os dardos mais espesos caan entre eles. que a arbor quando he entre si muy'to espesa.” (Y las saetas y los dardos, caían más espesos entre ellos, que el árbol cuando es entre sí muy espeso). C. T., vol. II, p. 41. En lo antiguo también se escribía ÁRVOR.
- (6) **árbore**. *s. f.* Árbol. El mayor de los vegetales leñosos [seguen exs. ant. e mod.]; 2. *náut.* Cualquiera de los palos o masteleros de una nave u otra embarcación. 3. En *tipografía*, pieza de hierro en la parte superior del husillo de la prensa de imprimir. Var: ÁLBORE, ALBRE, ÁRBOL, ÁRBOLE, ÁRBOR y ARBRE. Antiguamente solía escribirse ARVORE.
- (7) **arbre**. Contracción de ÁRBORE. V. [seguen exs. de Rosalía, Pondal e García Ferreiro].
- (8) **arvor**. V. ARBORE.
- (9) **arvore**. V. ARBORE [seguen exs. ant.].

Real Academia Galega (1913-1928):  
*Diccionario gallego-castellano [a-cativo]*. A  
Coruña: Imprenta y fotograbado de Ferrer.

# Corpus y planificación lingüística

- usos en la **planificación del corpus**
  - la lingüística de corpus como base teórica y metodológica para la elaboración de **recursos lingüísticos** (gramáticas, diccionarios...)
  - los corpus como **base empírica** de las decisiones normativas
- usos en la **planificación del estatus**
  - aplicaciones de los corpus en las **tecnologías de la lengua**, fundamentales para el uso y prestigio de la lengua
  - corpus para el **aprendizaje e investigación** lingüística
  - la **función simbólica** de los corpus nacionales





Universidade de Vigo

# Corpus lingüísticos y planificación del estatus

- como base para la elaboración de herramientas básicas para el procesamiento lingüístico del lenguaje, imprescindible para la construcción de aplicaciones más complejas de las tecnologías de la lengua
  - etiquetadores gramaticales (*taggers*)
  - identificación de nombres de entidades (*named entity recognition*)
  - analizadores sintácticos (*parsers*) probabilísticos
  - etiquetadores semánticos, desambiguación del significado léxico (*word sense disambiguation*)





## FreeLing Galego

Análise lingüística automática do galego

(c) [Seminario de Lingüística Informática \(SLI\)](#), Universidade de Vigo, 2006

(c) [Centre TALP](#), Universitat Politècnica de Catalunya, 2006

Con tecnoloxía [FreeLing 1.5](#) de código aberto



[In English](#)

### Escriba aquí as frases

A cativa ceou moitas patacas

### Opcións de análise

- Identificación de locucións
- Identificación de números
- Identificación de data/hora
- Cantidades, proporcións e porcentaxes
- Identificación de entidades definidas

### Tipo de análise

Análise morfolóxica desambiguada en contexto

Analizar

## Resultado da análise

### Sentence #1

<b>A</b>	<b>cativa</b>	<b>ceou</b>	<b>moitas</b>	<b>patacas</b>
<i>o</i>	<i>cativo</i>	<i>cear</i>	<i>moito</i>	<i>pataca</i>
DA0FS0	NCFS000	VMIS3S0	DI0FP0	NCFP000

<http://sli.uvigo.es/lingua>

# Corpus lingüísticos y planificación del estatus

- como fundamento básico de muchas aplicaciones de las tecnologías de la lengua
  - aplicaciones de texto predictivo (teléfonos móviles)
  - correctores ortográficos, gramaticales y de estilo
  - programas de resumen de textos
  - recuperación de información, respuestas a preguntas
  - traducción automática mediante ejemplos o estadística
  - tecnologías del habla, sistemas de dictado





Universidade de Vigo



Sen título 1 - OpenOffice.org Writer

Echeiro Editar Ver Inserir Formato Táboa Ferramentas Golfiño Ventá Axuda

Predefinido Times New Roman 12 B / U

A sal é un bo condimento.

**Golfiño: corrector gramatical d...**

Orixinal  
A

Observacións  
O artigo debe concordar en xénero e número co substantivo que acompaña

Substituír por  
O

Suxestións  
O

Iniciar  
Substituír  
Ignorar  
Ignorar todo  
Cancelar  
Estilos normativos

  
XUNTA DE GALICIA  
CONSELLERÍA DE INNOVACIÓN  
E INDUSTRIA  
Dirección Xeral de Promoción Industrial  
e da Sociedade da Información  
0% analizado



Universidade de Vigo

## TA basada en corpus

- TA mediante ejemplos
  - objetivo: descomponer la oración de la LO en fragmentos, traducir estos fragmentos por analogía a partir de un corpus paralelo, y componer los fragmentos traducidos en una oración bien construida en la LT
  - el corpus paralelo modelo tiene que estar alineado a nivel de fragmentos



## TA basada en corpus

- ejemplo: traducir del inglés al japonés:
  - He buys a book on international politics
- a partir de estos dos ejemplos:
  - (He buys) a notebook = (Kare ha) nouto (wo kau)
  - I read (a book on international politics) = Watashi ha (kokusaiseiji nitsuite kakareta hon) wo yomu
- resulta en:
  - (Kare ha) (kokusaiseiji nitsuite kakareta hon) (wo kau)

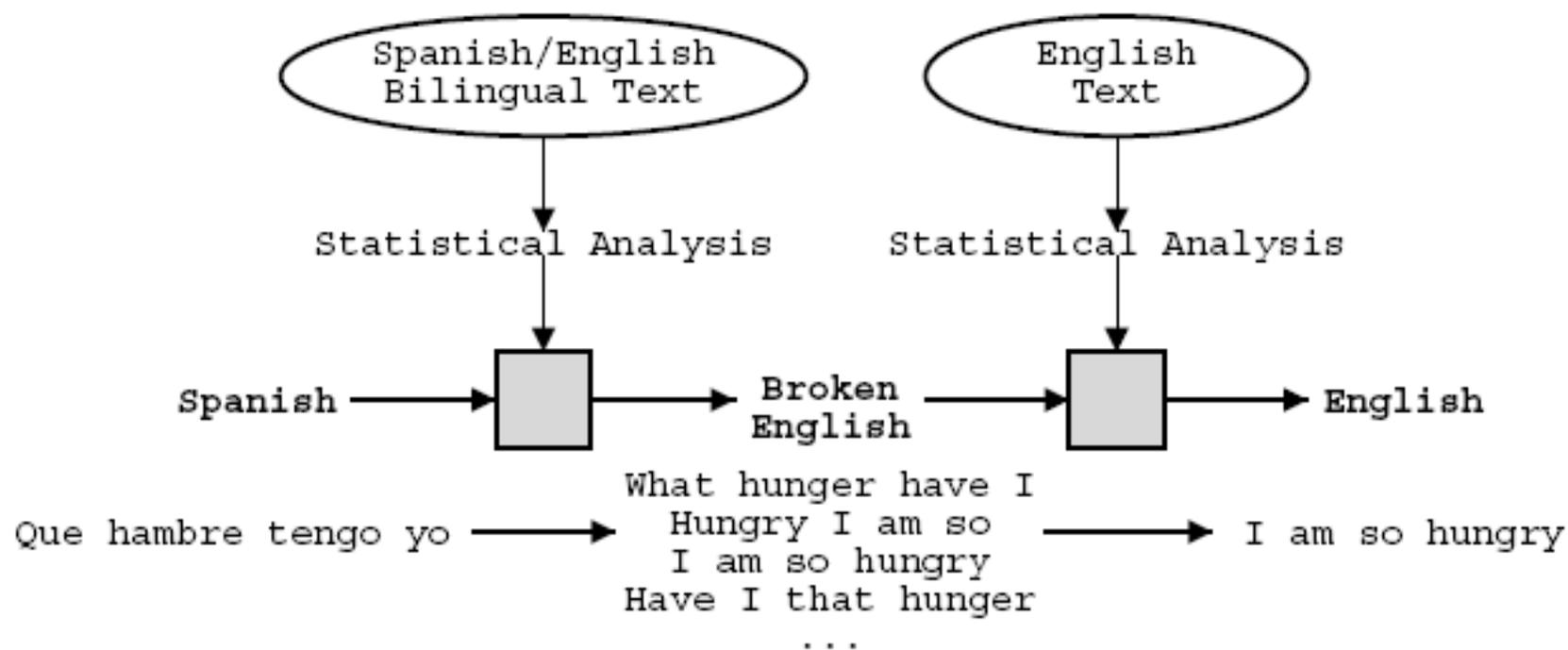


# TA basada en corpus

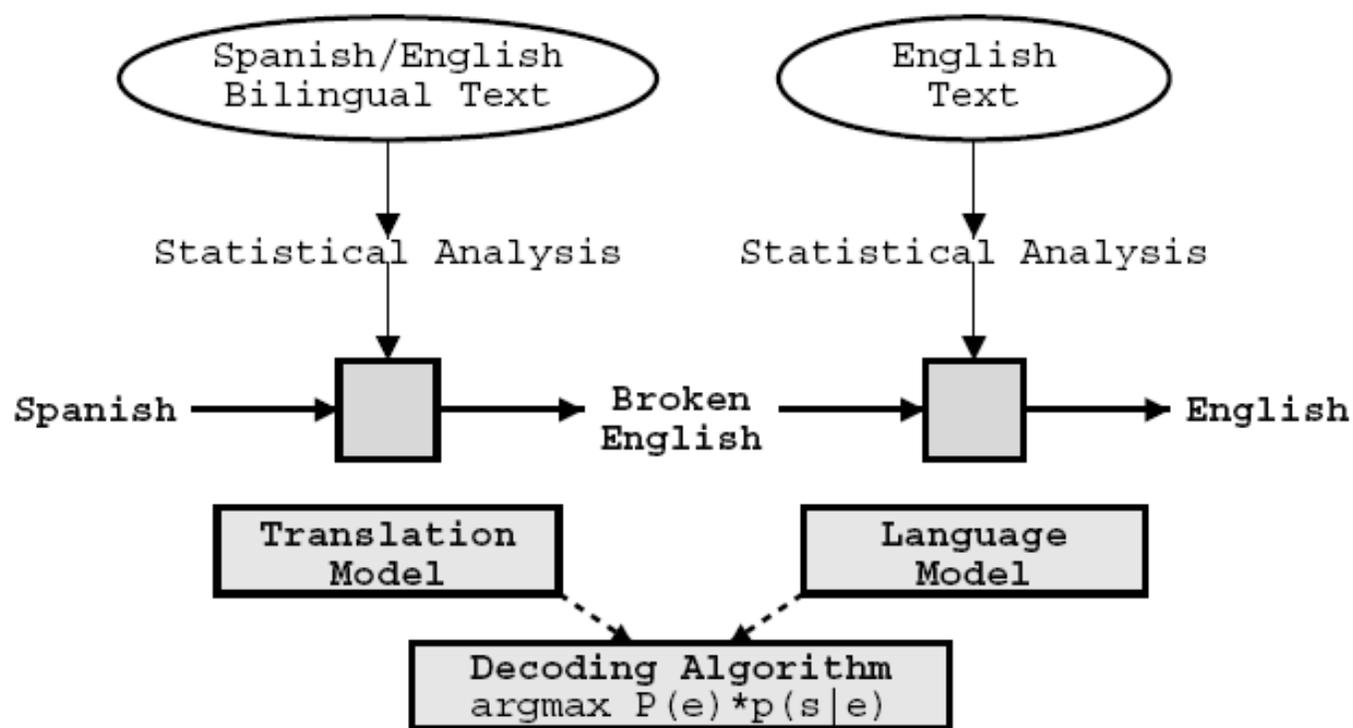
- TA estadística
  - obxectivo: descubrir la frase más probable en LT dada una frase original en LO
  - Las probabilidades se calculan a partir de un modelo estadístico de traducción basado en el corpus paralelo alineado a nivel de palabra



# TA basada en corpus



# TA basada en corpus



# Google tradutor

Tradución

[Procura traducida](#)

[Ferramentas de tradución](#)

[Ferramentas e recursos](#)

## Traduza texto, páxinas web e documentos

Introduza texto ou unha páxina web URL, ou [envíe un documento](#).

Traducir do:

Traducir ao:

tradución de

[Suxira un...](#)

- Dinamarqués
- Eslovaco
- Esloveno
- Español
- Estoniano
- Finés
- Francés
- galego
- Galés
- Grego
- Hebreo
- Hindú
- Holandés
- Húngaro
- Indonesio
- Inglés
- Irlandés**
- Islandés
- Italiano
- Letón

MANUEL GONZÁLEZ, PRESIDENTE DA RAG

## 'O borrador transpira pouco afecto polo galego'

A Real Academia Galega (RAG) considera 'inaceptábel' e 'inconsistente legalmente' o borrador do decreto do galego no... máis

2 Comentarios



- Análise das 'bases para o decreto do plurilingüismo' da RAG
- A RAG ve no decreto do PP 'unha involución' do galego

MARISOL SONEIRA, DEPUTADA DO PSDEG

## 'O Goberno merca coches de luxo e os Audis podrecen'

A voceira de Pesca do PSdeG, Marisol Soneira, pide 'explicacións' no Parlamento galego pola compra dun coche... máis

1 Comentario



'E O GRAN REGULADOR É O BANCO DE ESPAÑA'  
Caamaño: 'A competencia'

XOSÉ FONTES (CC 00)

## 'O decreto fai dos centros de ensino lugares de conflito'

O responsábel de Ensino Público de CCOO, Xosé Fontes, xustifica a adhesión da súa agrupación sindical á folga... máis

2 Comentarios



- A Xunta non fala do decreto con Queremos Galego
- CC 00 tamén se suma á folga no ensino o 21X
- 'Non se pode tratar o inglés como ás linguas oficiais'

EDUCACIÓN INICIA A ROLDA DE CONTACTOS

## A Xunta non fala do decreto con Queremos Galego

A Consellaría de Educación comeza hoxe a rolda de contactos para presentarlle o borrador do polémico decreto do... máis

7 Comentarios



- A AELG apoia a folga no ensino
- 'O decreto fai dos centros de ensino lugares de conflito'
- UPyD reclama a súa presenza na rolda de contactos sobre o decreto do idioma

Google tradutor

http://www.anosaterra.org/

Traducir

Traducir do: galego

Traducir ao: Irlandés

MANUEL GONZALEZ, UACHTARÁN NA RAG

## 'Exudes an dréacht-affection beag ar an Gailísis'

An lucht acadúil Real Galega (RAG) Measann 'inghlactha' agus 'go deathach ar neamhréir' 'san fholcadán dréacht de na Gailísis ... níos mó

2 comments



Texto orixinal en galego:

Google

A RAG ve no decreto do PP 'unha involución' do galego

+ Suxira unha tradución mellor

+ RAG féach san fholcadán de involucion ar an PP 'na Gailísis'

MARISOL SONEIRA, BALL DE PSDEG

## "Buys rialtas na gluasteáin agus só na podrecen Audi "

Iarrann A urlabhraí don PSdeG lascaigh, Marisol Soneira, 'buzzwords' i bParlaimint na Gailísis do cheannach de ghluasteáin ... níos mó

Nota 1



JOSÉ FONTES (CC 00)

## 'Is é an fholcadán an ionaid oideachais áitiúla ar choimhlint'

Údar an ceann an Oideachais Poiblí de CCOO, José Fontes, an aontachais ar a aontas ghrúpa dul ar stailc ... níos mó

2 comments



- Ní dhéanann an Bord ar labhairt na fholcadán againn Galego
- CC Oo Cuireann freisin leis an stailc ag teagasc 21X
- 'Ní féidir leat déileáil le Béarla mar theanga oifigiúil'

OIDEACHAS THOSAÍONN BHABHTA DE TEAGMHÁLACHA

## Ní dhéanann an Bord ar labhairt na fholcadán againn Galego

An Roinn Oideachais dtosaíonn liosta de na teagmhálacha a chur i láthair an dréacht de na fholcadán conspóideach ... níos mó

7 Comments





## Corpus lingüísticos y planificación del estatus

- los corpus textuales constituyen una base muy importante para el desarrollo de las tecnologías de la lengua, que a su vez contribuyen al uso de la lengua nacional en ámbitos como el ordenador, el teléfono móvil, la agenda electrónica, el ordenador de bolsillo, etc., ámbitos en los que la lengua gana utilidad y prestigio entre un amplio sector de usuarios de interés estratégico para la planificación lingüística



---

Universidade de Vigo



40 zurekin



EUSKALTZAINDIA  
REAL ACADEMIA DE LA LENGUA VASCA  
ACADÉMIE DE LA LANGUE BASQUE

## Corpus lingüísticos y planificación del estatus

- en la planificación de la adquisición: como base para la elaboración de materiales didácticos, orientados a ampliar la extensión de la lengua tanto a nivel nacional como internacional
- como material de soporte de una política lingüística de fomento de la investigación centrada en la lengua propia
- los corpus nacionales como imagen pública simbólica de la unidad, dignidad, independencia y fortaleza de una lengua (función simbólica semejante a la del diccionario “oficial” de la lengua)



---

Universidade de Vigo

## Conclusiones

- la explotación de corpus permite beneficiarse de datos lingüísticos reales inalcanzables por otras vías, como la frecuencia de uso de las palabras y construcciones, su potencialidad semántica, su dispersión textual, su datación temporal o su combinatoria
- aun así, debe observarse que la extracción de información lingüística de los corpus puede complementar el trabajo de investigación humano, pero nunca lo hace innecesario (especialmente, cuando se esperan resultados con calidad humana, como en la traducción o en la generación de diccionarios para uso humano)



## Conclusiones

- hoy en día, no es posible concebir la planificación lingüística sin tener en cuenta la construcción y explotación de corpus textuales de la lengua monolingües y plurilingües, contemporáneos e históricos, generales y especializados.
- el papel de los corpus textuales en la planificación lingüística es central, y es relevante en todas sus vertientes: planificación del corpus, planificación de la adquisición y planificación del estatus



## Conclusiones

- debido a su enorme incidencia social, la presencia de una lengua en las tecnologías es un hecho determinante para alcanzar (o para conservar) el estatus de lengua normalizada
- las tareas de normalización de una lengua deben incluir entre sus objetivos el de dotar a la sociedad del conjunto necesario de tecnologías lingüísticas
- las tecnologías de la lengua constituyen un requisito imprescindible para que una comunidad lingüística pueda acceder en pie de igualdad y en su propia lengua a todos los bienes y servicios de la sociedad de la información





# Corpus textuales y planificación lingüística

Eskerrik asko!



---

Universidade de Vigo

**TALG**

Grupo de investigación  
Tecnoloxías e Aplicacións  
da Lingua Galega