



**EUSKO JAURLARITZA**  
**GOBIERNO VASCO**

HEZKUNTZA, HIZKUNTZA POLITIKA  
ETA KULTURA SAILA  
Hizkuntza Politikarako Sailburuordetza

DEPARTAMENTO DE EDUCACIÓN,  
POLÍTICA LINGÜÍSTICA Y CULTURA  
Viceconsejería de Política Lingüística



## **Léxicos bilingües abiertos del traductor automático**

**Descripción de los léxicos bilingües con euskera en formato TBX**

## TABLA DE CONTENIDOS

---

<b>INTRODUCCIÓN.....</b>	<b>3</b>
<b>DESCRIPCIÓN DE LOS LÉXICOS BILINGÜES ABIERTOS.....</b>	<b>4</b>
LÉXICOS BILINGÜES ENTREGADOS .....	4
DESCRIPCIÓN DEL FORMATO TBX LOS LÉXICOS .....	5
<i>Campos relevantes de cada entrada.....</i>	<i>5</i>

## Introducción

Este documento describe el contenido y el formato de los léxicos bilingües español→euskera, euskera→español, inglés→euskera y euskera→inglés, extraídos de los léxicos computacionales del traductor automático instalado en el Gobierno Vasco para dichos pares de lenguas, y que se van a poner en abierto próximamente.

## Descripción de los léxicos bilingües abiertos

### Léxicos bilingües entregados

Se han puesto en formato abierto los léxicos bilingües desarrollados hasta la fecha para las cuatro direcciones de traducción con el euskera para el Gobierno Vasco:

- Español→euskera
- Euskera→español
- Inglés→euskera
- Euskera→inglés

Los léxicos contienen entradas bilingües con información sobre el lema, categoría, área temática y dialecto (inglés británico o inglés americano, en el caso del léxico euskera→inglés), para las categorías abiertas: nombre, adjetivos, verbos y adverbios.

El número de entradas de cada léxico es:

LÉXICO BILINGÜE	NÚMERO DE ENTRADAS
Español→euskera	62.271 entradas
Euskera→español	67.691 entradas
Inglés→euskera	44.406 entradas
Euskera→inglés	50.045 entradas

## Descripción del formato TBX los léxicos

Los léxicos se suministran en cuatro ficheros – uno por dirección de traducción -, cada uno de ellos en formato TBX (TermBase eXchange). TBX es un estándar ISO basado en XML – y por lo tanto editable con un editor de texto o de XML - para el intercambio de terminología y léxicos (cf. [https://en.wikipedia.org/wiki/TermBase\\_eXchange](https://en.wikipedia.org/wiki/TermBase_eXchange), <http://www.ttt.org/oscarstandards/> ).

Cada entrada bilingüe tiene el siguiente formato:

(ejemplo para la entrada “aurkezpen” [nombre] → “presentation” [nombre] {Vocabulario General} del léxico **euskera** → **inglés**)

```
<termEntry>
<descrip type="subjectField">General Vocabulary</descrip>
<langSet xml:lang="eu">
<tig>
<term>aurkezpen</term>
<termNote type="partOfSpeech">noun</termNote>
</tig>
</langSet>
<langSet xml:lang="en">
<tig>
<term>presentation</term>
<termNote type="partOfSpeech">noun</termNote>
</tig>
</langSet>
</termEntry>
```

## Campos relevantes de cada entrada

Lengua a la que pertenece el lema

Está definida entre las etiquetas <langSet xml:lang="LENGUA"> ... </langSet>

Los valores posibles para LENGUA son:

**eu** = euskera

**es** = español

**en** = inglés (sin especificación de dialecto)

**en-uk** = inglés británico

**en\_us** = inglés americano

## Lemas de la entrada bilingüe

Los lemas de la entrada bilingüe están definidos entre las etiquetas `<term>LEMA</term>`, dentro de su sección `<langSet>` correspondiente.

## Categoría sintáctica

La categoría sintáctica de cada lema está definida entre las etiquetas

`<termNote type="partOfSpeech">CATEGORÍA</termNote>`

donde CATEGORÍA puede tener los siguientes valores:

**noun** = nombre

**adjective** = adjetivo

**verb** = verbo

**adverb** = adverbio

**determiner** = determinante

**pronoun** = pronombre

## Dominio (área temática)

El valor de dominio indica a qué área temática pertenece la entrada bilingüe. Este valor está definido entre las etiquetas:

`<descrip type="subjectField">DOMINIO</descrip>`

Donde DOMINIO puede ser una de las siguientes áreas temáticas:

INGLÉS	ESPAÑOL	EUSKERA
"General Vocabulary"	"Vocabulario general"	"Hiztegi orokorra"
"Common Social Vocabulary"	"Vocabulario social general"	"Hiztegi sozial orokorra"
"Art & Literature"	"Arte y literatura"	"Arte eta literatura"
"Architecture"	"Arquitectura"	"Arkitektura"
"Music"	"Música"	"Musika"
"Art, User Entries"	"Arte, entradas de usuario"	"Artea, erabiltzailearen sarrerak"
"Ecology, Environment Protection & Environmental Problems"	"Ecología, medio ambiente"	"Ekologia eta ingurumena"
"Ecology, User Entries"	"Ecología, entradas de usuario"	"Ekologia, erabiltzailearen sarrerak"
"Economy & Trade"	"Economía y comercio"	"Ekonomia eta merkataritza"

"Business Management"	"Gestión empresarial"	"Enpresa-kudeaketa"
"Finance & Banking"	"Banca y finanzas"	"Finantzak"
"Insurance"	"Seguros"	"Aseguruak"
"Political Economics"	"Ciencias políticas"	"Ekonomia politikoa"
"Economy, User Entries"	"Economía, entradas de usuario"	"Ekonomia, erabiltzailearen sarrerak"
"Law & Legal Science"	"Derecho"	"Zuzenbidea"
"Law, User Entries"	"Derecho, entradas de usuario"	"Zuzenbidea, erabiltzailearen sarrerak"
"Government, Public Authorities, International Relations"	"Instituciones públicas y gubernamentales, relaciones internacionales"	"Gobernua, erakunde publikoak, nazioarteko harremanak"
"Government etc., User Entries"	"Instituciones públicas etc., entradas de usuario"	"Gobernua, erabiltzailearen sarrerak"
"Recreation, Sports, Games & Hobbies"	"Ocio, deportes y juegos"	"Kirolak eta aisialdia"
"Recreation etc., User Entries"	"Ocio etc., entradas de usuario"	"Kirolak eta aisialdia, erabiltzailearen sarrerak"
"Social Sciences"	"Ciencias sociales"	"Soziologia"
"Linguistics"	"Lingüística"	"Hizkuntzalaritza"
"Media, Publishing, Libraries & Documentation"	"Editoriales, medios de comunicación, bibliotecas y documentación"	"Komunikabideak, argitaletxeak, liburutegiak eta dokumentazioa"
"Social Sciences, User Entries"	"Ciencias sociales, entradas de usuario"	"Soziologia, erabiltzailearen sarrerak"
"Common Technical Vocabulary"	"Vocabulario técnico general"	"Hiztegi tekniko orokorra"
"Agriculture & Fishing"	"Agricultura, ganadería y pesca"	"Nekazaritza eta arrantza"
"Agriculture & Fishing, User Entries"	"Agricultura, ganadería y pesca, entradas de usuario"	"Nekazaritza eta arrantza, erabiltzailearen sarrerak"
"Civil Engineering"	"Ingeniería civil"	"Ingeniaritza zibila"
"Civil Engineering, User Entries"	"Ingeniería civil, entradas de usuario"	"Ingeniaritza zibila, erabiltzailearen sarrerak"
"Data Processing"	"Informática"	"Informatika"
"Computer Hardware"	"Hardware"	"Hardwarea"
"Computer Software"	"Software"	"Softwarea"
"Data Transmission"	"Transmisión de datos"	"Datu-transmisioa"
"Data Processing, User Entries"	"Informática, entradas de usuario"	"Informatika, erabiltzailearen sarrerak"
"Electrical Engineering"	"Electrotecnia"	"Elektroteknia"
"Semiconductor Technology"	"Semiconductores"	"Erdieroaleak"
"Electrical Engineering, User Entries"	"Electrotecnia, entradas de usuario"	"Elektroteknia, erabiltzailearen sarrerak"
"Mechanical Engineering"	"Electromecánica"	"Elektromekanika"
"Mechanical Engineering, User Entries"	"Electromecánica, entradas de usuario"	"Elektromekanika, erabiltzailearen sarrerak"
"Medical Science"	"Medicina"	"Medikuntza"
"Medical Science, User Entries"	"Medicina, entradas de usuario"	"Medikuntza, erabiltzailearen sarrerak"
"Mining, Raw Material Industry"	"Minería, industria de materias primas"	"Meatzaritza, lehengaien industria"
"Mining, User Entries"	"Minería, entradas de usuario"	"Meatzaritza, erabiltzailearen sarrerak"
"Natural Sciences & Mathematics"	"Ciencias naturales y matemáticas"	"Natur zientziak eta matematika"
"Biology"	"Biología"	"Biologia"
"Chemistry"	"Química"	"Kimika"
"Geography, Geology & Meteorology"	"Geografía, geología y meteorología"	"Geografia, geologia eta meteorologia"

"Mathematics & Statistics"	"Matemáticas y estadística"	"Matematika eta estatistika"
"Physics"	"Física"	"Fisika"
"Natural Sciences & Mathematics, User Entries"	"Ciencias naturales y matemáticas, entradas de usuario"	"Natur zientziak eta matematika, erabiltzailearen sarrerak"
"Energy Technology, Power Engineering"	"Energía, ingeniería energética"	"Energia"
"Energy Technology, User Entries"	"Energía, entradas de usuario"	"Energia, erabiltzailearen sarrerak"
"Technology, Industry, Crafts & Trades"	"Tecnología, industria y artesanía"	"Teknologia, industria eta eskulangintza"
"Aeronautic & Space Industry"	"Aeronáutica, industria espacial"	"Aeronautika eta astronautika"
"Motor Car Industry"	"Automoción"	"Automobilgintza"
"Ceramics & Glass Industry"	"Cerámica e industria del vidrio"	"Zeramika eta beiraren industria"
"Food, Drink & Tobacco Industries"	"Alimentación e industrias tabaqueras"	"Elikagaigintza eta tabakoaren industria"
"Metallurgy"	"Metalurgia"	"Metalurgia"
"Military Industry"	"Industria militar"	"Industria militarra"
"Norms & Standards"	"Normas y estándares"	"Arauek eta estandarrak"
"Optical Industry"	"Óptica"	"Optika"
"Rubber & Plastics Industry"	"Plásticos"	"Plastikoen eta kautxuen industria"
"Railway Engineering"	"Ferrocarriles"	"Trenbideak"
"Shipbuilding"	"Construcción naval"	"Itsasontzigintza"
"Textile & Leather Industry"	"Industria textil y del cuero"	"Ehungintza eta larrugintza"
"Transportation & Traffic"	"Transportes y tráfico"	"Garraioak"
"Typographic, Graphic & Reprographic Industries"	"Industrias gráficas"	"Industria grafikoak"
"Wood Industry, Carpentry & Joinery"	"Industria maderera, carpintería"	"Zurgintza"
"Technology, User Entries"	"Tecnología, entradas de usuario"	"Teknologia, erabiltzailearen sarrerak"
"Telecommunications"	"Telecomunicaciones"	"Telekomunikazioak"
"Fibre Optics"	"Fibras ópticas"	"Zuntz optikoak"
"Radio Technology"	"Radiotecnía"	"Irrati-ingeniaritza"
"Telecommunications, User Entries"	"Telecomunicaciones, entradas de usuario"	"Telekomunikazioak, erabiltzailearen sarrerak"
"eXtra"	"Otras"	"Beste batzuk"
"Synonyms"	"Sinónimos"	"Sinonimoak"