

Opentrad

Traducció automàtica en codi obert
Traducción automática en código abierto
Itzulpen automatikoa kode irekian
Traducción automática en código abierto

Barcelona, 30 de maig de 2006

Novedades Novetats Berritasunak Novidades

- Características generales
- Opciones de traducción
- Resultados
- Casos prácticos
- Mirando al futuro

- Primer y único traductor automático entre **el español y el euskara**

eu

- Primer traductor automático que traduce **del gallego al español**

gl

- Primer traductor automático de código libre entre **el español y el catalán**

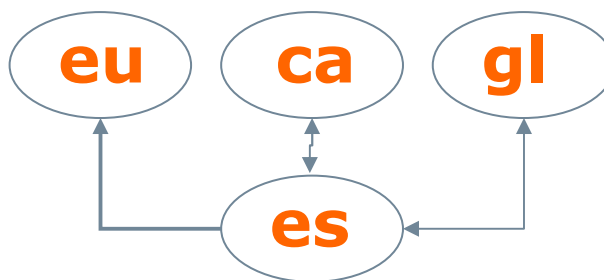
ca

- Primer sistema que aglutina **cuatro lenguas oficiales** del estado

④ lenguas llengües hizkuntza linguas

Características generales

- Un proyecto entre **cuatro lenguas**



- Desarrollado en **código abierto**



• **Aplicación y datos lingüísticos básicos** se pueden descargar en SourceForge



- Se utiliza la **tecnología de transferencia**



- El proyecto ha recibido la ayuda económica **del Ministerio de Industria, Comercio y Turismo** (programa PROFIT)



Características generales

Opciones de traducción

Resultados

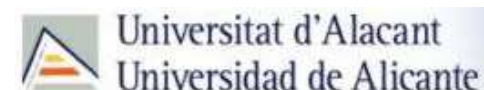
Casos prácticos

Mirando al futuro

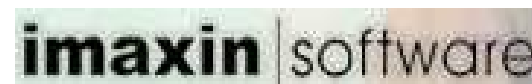
Características generales

Participantes: en el proyecto participamos 4 universidades y 3 empresas.

• Universidades



• Empresas



Características generales

Opciones de traducción

Resultados

Casos prácticos

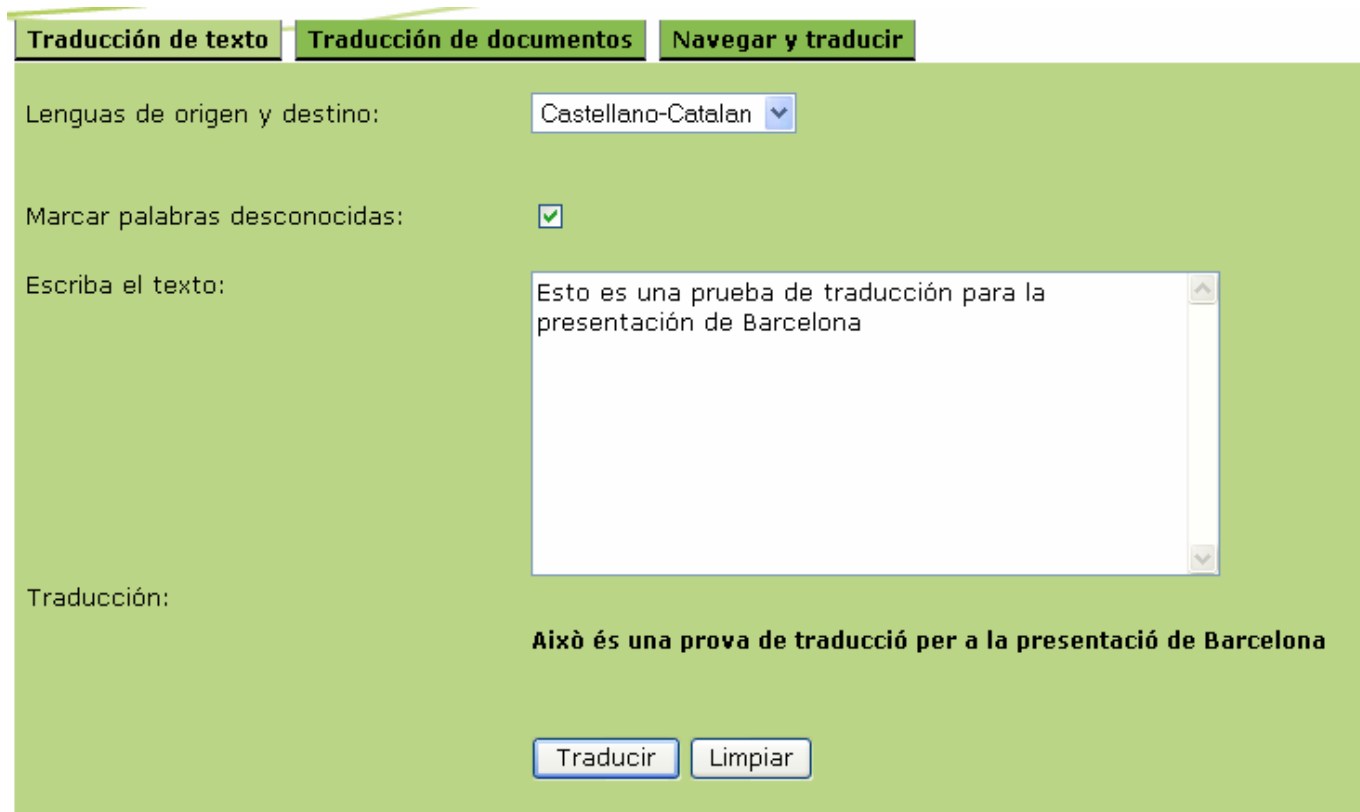
Mirando al futuro

Plazos de proyecto: inicio mediados del 2004 y fin primer trimestre del 2006

Tarea	2004	2005	2006
Especificación y análisis			
Desarrollo de los motores y recursos lingüísticos			
Pruebas y validación del sistema			

Opciones de traducción

Traducción de texto libre: el usuario escribe en la casilla el texto que desea traducir



Traducción de texto Traducción de documentos Navegar y traducir

Lenguas de origen y destino: Castellano-Catalan ▾

Marcar palabras desconocidas:

Escriba el texto:

Esto es una prueba de traducción para la presentación de Barcelona

Traducción:

Això és una prova de traducció per a la presentació de Barcelona

Traducir Limpiar

■ Características generales

■ **Opciones de traducción**

■ Resultados

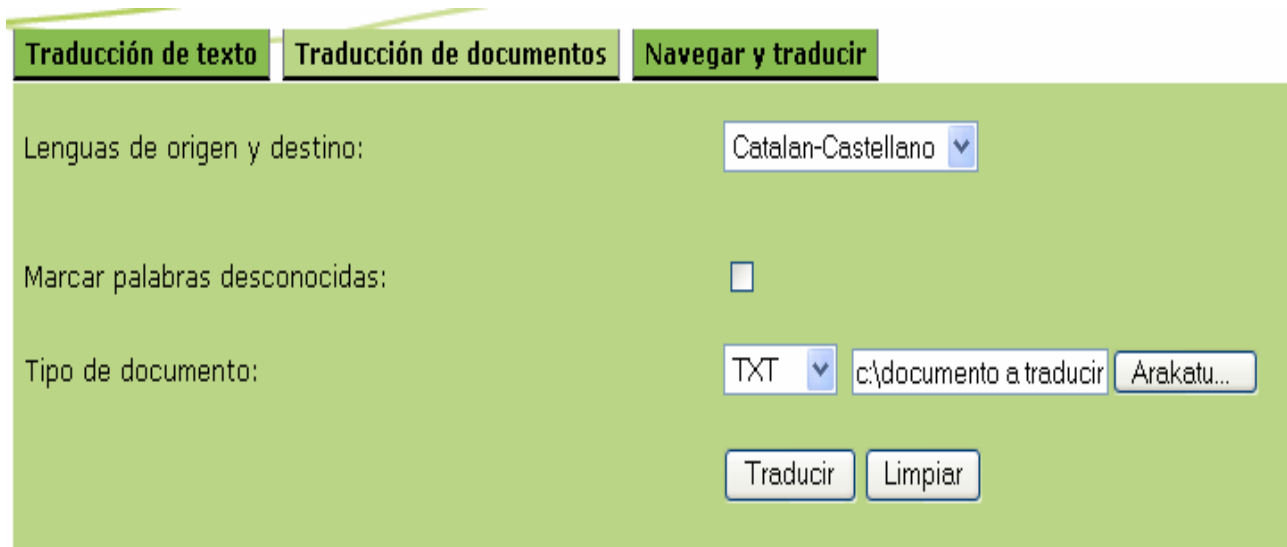
■ Casos prácticos

■ Mirando al futuro

Como máximo se traducen 10.000 caracteres

Opciones de traducción

Traducción de documentos: el usuario indica el documento cuyo contenido desea traducir



Traducción de texto Traducción de documentos Navegar y traducir

Lenguas de origen y destino: Catalan-Castellano ▾

Marcar palabras desconocidas:

Tipo de documento: TXT ▾ c:\documento a traducir Arakatu...

Traducir Limpiar

- Características generales
- **Opciones de traducción**
- Resultados
- Casos prácticos
- Mirando al futuro

Se aceptan los formatos TXT, RTF y HTML. Próximamente se incluirá el formato DOC (de Microsoft Word)

Como máximo se traducen 10.000 caracteres

Opciones de traducción


Navega y traduce: el usuario indica la página web que desea traducir. El sistema va traduciendo a medida que el usuario navega

Traducción de texto | **Traducción de documentos** | **Navegar y traducir**

Lenguas de origen y destino:

Marcar palabras desconocidas:

Dirección URL:

 Esta página ha sido traducida automáticamente, ver página original

Portal Institucional de Gipuzkoa | eSuscripciones | contacte | tots els enllacs | ajuda | accessibilitat | en | es | fr | en

Ciudadans
* Ajudes i Subvencions
* Conèixer Gipuzkoa
* Fiscalitat i Impostos
* Informació Territorial
[més temes ciutadans](#)

Associacions
Empreses
Visitants

En un només clic
* Butlletí Oficial de Gipuzkoa
* Ajudes i subvencions

Actualitat
*** Conveni per a escometre l'excavació integral del solar de Zarautz Jauregia en Getaria!**
La investigació arqueològica d'aquest jaciment permetrà la reconstrucció històrica de Getaria i un avanç decisiu en el coneixement de l'època antiga i medieval de Gipuzkoa. La Diputació Foral de Gipuzkoa aportarà 215.000 euros i l'Ajuntament de Getaria, 72.000 euros.

Agenda Gipuzkoa
Maig 2006 << >>
LuMa **El** Ju **Veig** SaDo
meu veure
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 **24** 25 26 27 28
29 30 31

EL TEMPS AVUI
ESTAT DEL TRAFIC
EL ESTADO DEL TRAFICO

Esdeveniments recomanats
Concurs de Pintura

WEBCAM

Características generales

Opciones de traducción

Resultados

Casos prácticos

Mirando al futuro

Resultados

Resumen de resultados técnicos:

▣ español-catalán:	Errores 3,8% Palabras desconocidas, transferencias estructurales
▣ español-gallego	Errores 4,8% Palabras desconocidas
▣ español-euskara	Errores 24,8% Analizador del español, acepciones incorrectas, flexiones incorrectas, orden

- ▣ Características generales
- ▣ Opciones de traducción
- ▣ **Resultados**
- ▣ Casos prácticos
- ▣ Mirando al futuro

Resultados es-ca: Apertium

Apertium és:

- una caixa d'eines de codi obert per a construir sistemes de traducció automàtica entre llengües emparentades, que inclou:
 - un motor (programa) de traducció de codi obert
 - programes per a la gestió de les dades lingüístiques
 - dades lingüístiques (diccionaris i regles) per a diversos parells de llengües

- Características generales
- Opciones de traducción
- **Resultados**
- Casos prácticos
- Mirando al futuro

Resultados es-ca: Apertium

Apertium també és, per tant:

- un sistema de traducció automàtica entre llengües emparentades
 - ràpid (10.000 mots per segon)
 - precís (fins al 96% de text correcte amb dades adequades)
 - gratuït i obert
 - per a traduir entre les llengües romàniques de l'Estat espanyol (espanyol, català/valencià, gallec, etc.)
 - que reconeix els formats de text més usuals

- Características generales
- Opciones de traducción
- **Resultados**
- Casos prácticos
- Mirando al futuro

Resultados es-ca: Apertium

El paquet Apertium de dades espanyol-català-espanyol (versió actual, 1.0).

- té 13.000 entrades de diccionari bilingüe
- 50 regles gramaticals de traducció
- reconeix entre el 90% i el 97% dels mots dels textos, dependent de la temàtica
- genera entre el 90% i el 96% de text correcte, dependent de la temàtica

- Características generales
- Opciones de traducción
- **Resultados**
- Casos prácticos
- Mirando al futuro

Resultados es-ca: Apertium

Com tots els paquets d'Apertium, el paquet de dades espanyol-català-espanyol està disponible i es pot provar en:

www.opentrad.com

i

www.apertium.org

on també es pot usar en línia per a traduir documents, webs i textos curts.

- Características generales
- Opciones de traducción
- **Resultados**
- Casos prácticos
- Mirando al futuro

Resultados es-ca: Apertium

Apertium es pot usar per a altres parells de llengües emparentades.

De fet, ja està disponible per a la traducció espanyol—portugués—espanyol.

Pròximament estarà disponible per a la traducció català—aranés—català (finançat per la Generalitat de Catalunya)

- Características generales
- Opciones de traducción
- **Resultados**
- Casos prácticos
- Mirando al futuro

Resultados es-eu

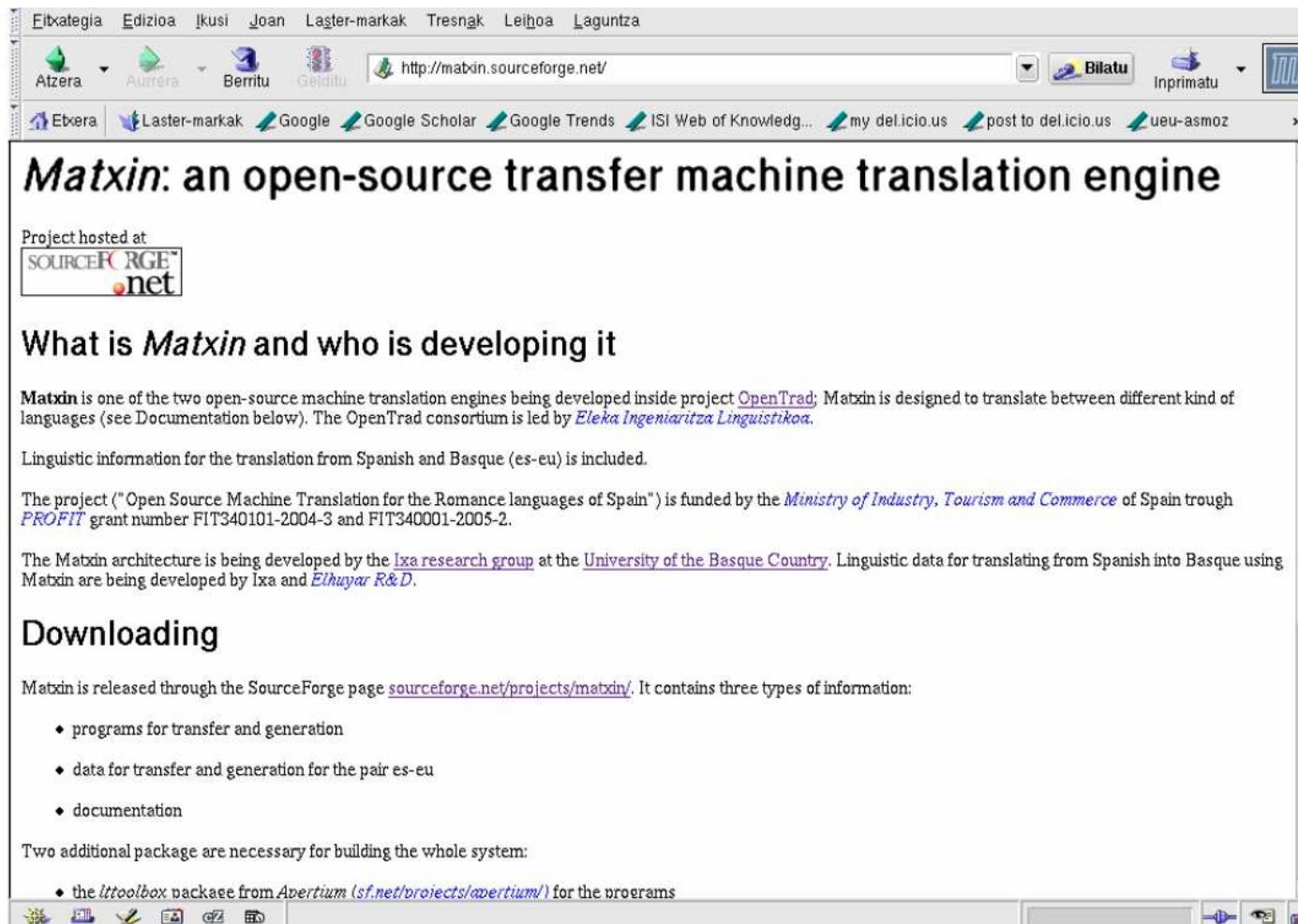
Características
generales

Opciones de
traducción


Resultados

Casos prácticos

Mirando al futuro



The screenshot shows a web browser window with the address bar displaying <http://matxin.sourceforge.net/>. The page title is "Matxin: an open-source transfer machine translation engine". The content includes the SourceForge logo, a heading "What is Matxin and who is developing it", and several paragraphs of text describing the project's goals, funding, and development. A "Downloading" section follows, listing the components included in the release and additional packages needed for building the system.

Project hosted at


What is *Matxin* and who is developing it

Matxin is one of the two open-source machine translation engines being developed inside project [OpenTrad](#); Matxin is designed to translate between different kind of languages (see Documentation below). The OpenTrad consortium is led by [Elexa Ingeniaritza Linguistikoa](#).

Linguistic information for the translation from Spanish and Basque (es-eu) is included.

The project ("Open Source Machine Translation for the Romance languages of Spain") is funded by the [Ministry of Industry, Tourism and Commerce](#) of Spain trough [PROFIT](#) grant number FIT340101-2004-3 and FIT340001-2005-2.

The Matxin architecture is being developed by the [Ixa research group](#) at the [University of the Basque Country](#). Linguistic data for translating from Spanish into Basque using Matxin are being developed by Ixa and [Eihuyar R&D](#).

Downloading

Matxin is released through the SourceForge page sourceforge.net/projects/matxin/. It contains three types of information:

- ♦ programs for transfer and generation
- ♦ data for transfer and generation for the pair es-eu
- ♦ documentation

Two additional package are necessary for building the whole system:

- ♦ the *ltxtoolbox* package from [Apertium](#) (sf.net/projects/apertium/) for the programs

Resultados es-eu

Características generales

Opciones de traducción

Resultados

Casos prácticos

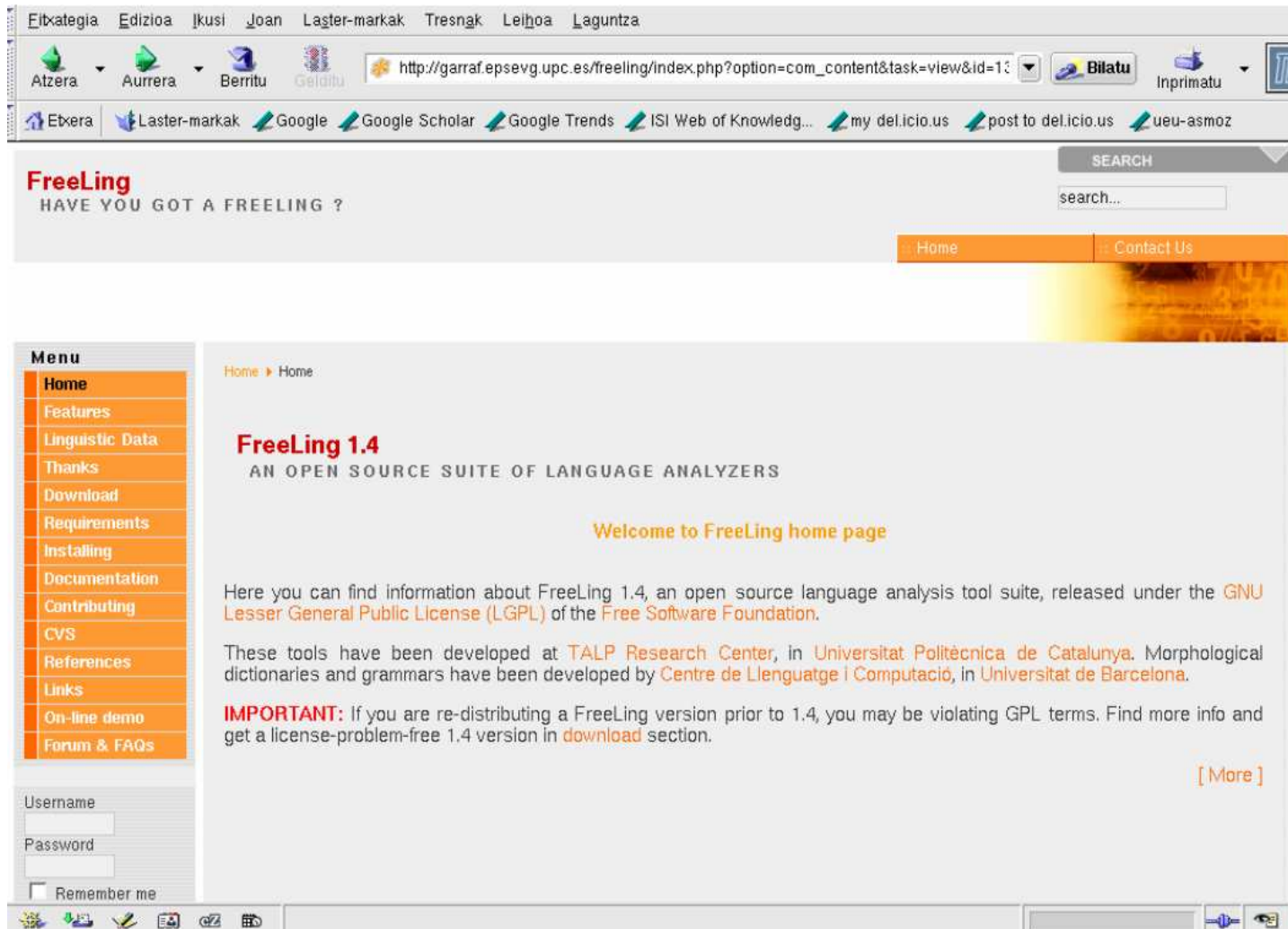
Mirando al futuro

The screenshot shows the SourceForge project page for 'mabxin'. The page includes a navigation menu, a search bar, and a main content area with sections for 'About mabxin', 'Latest File Releases', and 'File Releases'. The 'Latest File Releases' section contains a table with the following data:

Package	Release	Date	Notes / Monitor	Downloads
mabxin	0.7	May 5, 2006	-	Download

The 'File Releases' section below lists the latest release: '0.7 [Notes] (2006-05-05 07:36)'.

Resultados es-eu



The screenshot shows a web browser window displaying the FreeLing website. The browser's address bar shows the URL: `http://garraf.epsevg.upc.es/freeling/index.php?option=com_content&task=view&id=10`. The website header includes the text "FreeLing HAVE YOU GOT A FREELING ?" and a search bar. A navigation menu on the left lists various sections: Home, Features, Linguistic Data, Thanks, Download, Requirements, Installing, Documentation, Contributing, CVS, References, Links, On-line demo, and Forum & FAQs. The main content area features the heading "FreeLing 1.4 AN OPEN SOURCE SUITE OF LANGUAGE ANALYZERS" and a welcome message: "Welcome to FreeLing home page". Below this, there is a paragraph of introductory text and an "IMPORTANT" notice regarding the GNU Lesser General Public License (LGPL) and the Free Software Foundation. A "[More]" link is visible at the bottom right of the main content area. The browser's taskbar at the bottom shows several application icons.

Características
generales

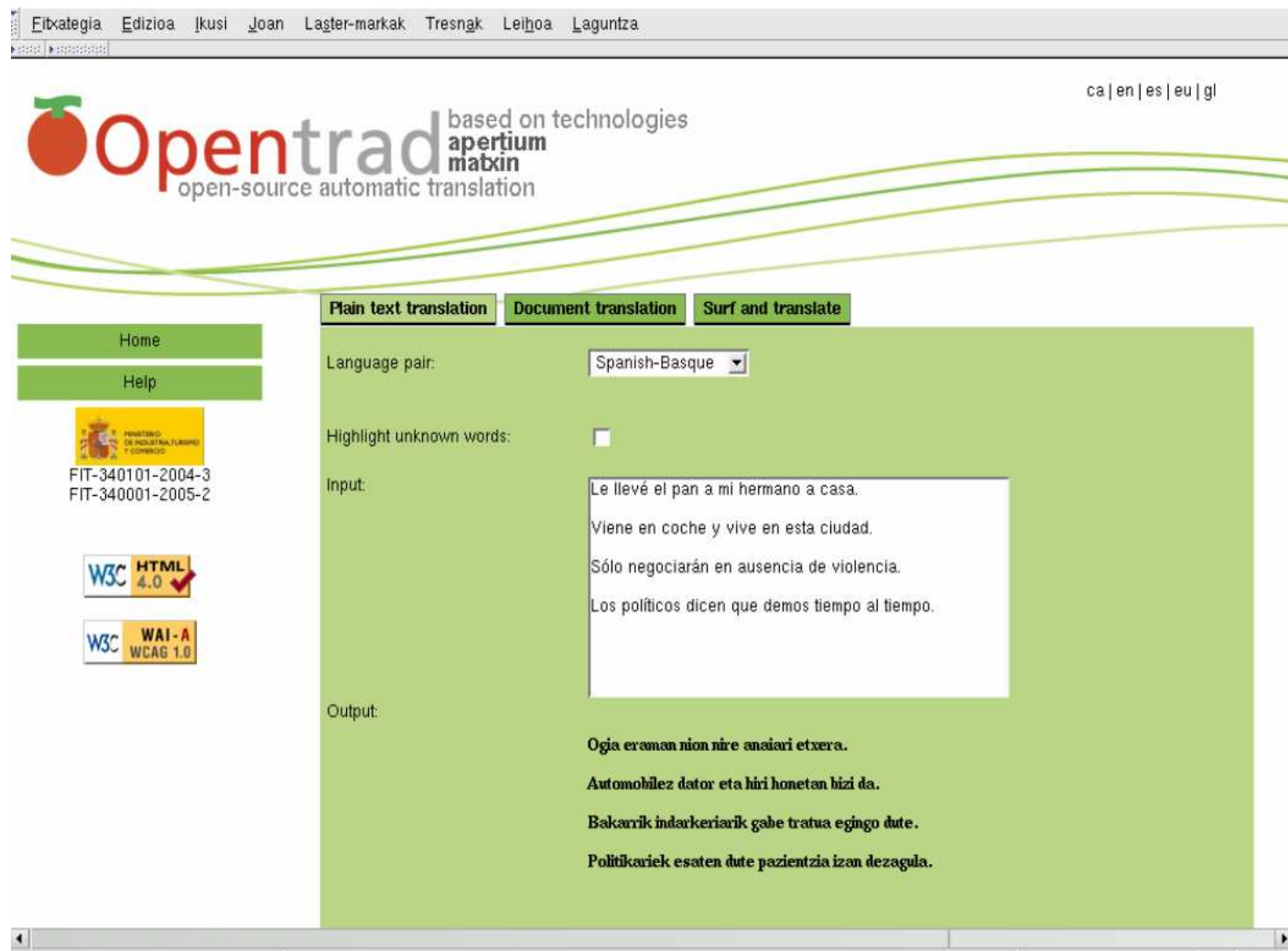
Opciones de
traducción

Resultados

Casos prácticos

Mirando al futuro

Resultados es-eu



The screenshot shows the Open trad web interface. At the top, there is a navigation menu with links: Fitxategia, Edizioa, Ikusi, Joan, Laister-markak, Tresngak, Leihoa, Laguntza. The language selection is set to 'ca | en | es | eu | gl'. The main header features the Open trad logo and the text 'based on technologies apertium mabin open-source automatic translation'. Below the header, there are three tabs: 'Plain text translation' (selected), 'Document translation', and 'Surf and translate'. On the left side, there are links for 'Home' and 'Help', and a list of FIT-340101-2004-3 and FIT-340001-2005-2. There are also logos for W3C HTML 4.0 and W3C WAI-A WCAG 1.0. The main content area shows the translation process. The 'Language pair' is set to 'Spanish-Basque'. The 'Highlight unknown words' checkbox is unchecked. The 'Input' field contains the Spanish text: 'Le llevé el pan a mi hermano a casa. Viene en coche y vive en esta ciudad. Sólo negociarán en ausencia de violencia. Los políticos dicen que demos tiempo al tiempo.' The 'Output' field shows the translated Basque text: 'Ogia eraman nion nire anaiari etxera. Automobílez dator eta hiri honetan bizi da. Bakarrik indarkeriarik gabe tratua egingo dute. Politikariek esaten dute pazientzia izan dezagula.'

Características
generales

Opciones de
traducción

Resultados

Casos prácticos

Mirando al futuro

Resultados es-eu



The screenshot shows the Open trad web interface. At the top, there is a navigation menu with links: [Fitxategia](#), [Edizioa](#), [Ikusi](#), [Joan](#), [Laster-markak](#), [Tresnak](#), [Leihoa](#), and [Laguntza](#). The language selection is set to 'ca | en | es | eu | gl'. The main header features the Open trad logo and the text 'based on technologies apertium matxin open-source automatic translation'. Below the header, there are three tabs: 'Plain text translation' (selected), 'Document translation', and 'Surf and translate'. On the left side, there are two buttons: 'Home' and 'Help'. Below these are logos for the Spanish Ministry of Industry, Tourism and Commerce, and two W3C standards: HTML 4.0 and WAI-A WCAG 1.0. The main content area shows the translation process. The 'Language pair' is set to 'Spanish-Basque'. The 'Highlight unknown words' checkbox is unchecked. The 'Input' field contains the text: '¿Te preocupan los virus informáticos? Cuatro nuevas sucursales de Correos se abrirán en la capital. El hospital tendrá 48 nuevas habitaciones individuales en 2009. ¿Quién crees que está ganando la batalla de las consolas portátiles?'. The 'Output' field shows the translated text: 'Birus informatikoak kezkatzen dituzte? Correos-en 4 sukertsal berri hiriburuan ireldko dira. Ospitaleak 48 Banako Gela berri izango du 2009TAN. Nork sinesten duzu kotsola eramangarrien batalla irabazten ari dela?'. The browser's status bar at the bottom shows the page is 100% zoomed.

Características
generales

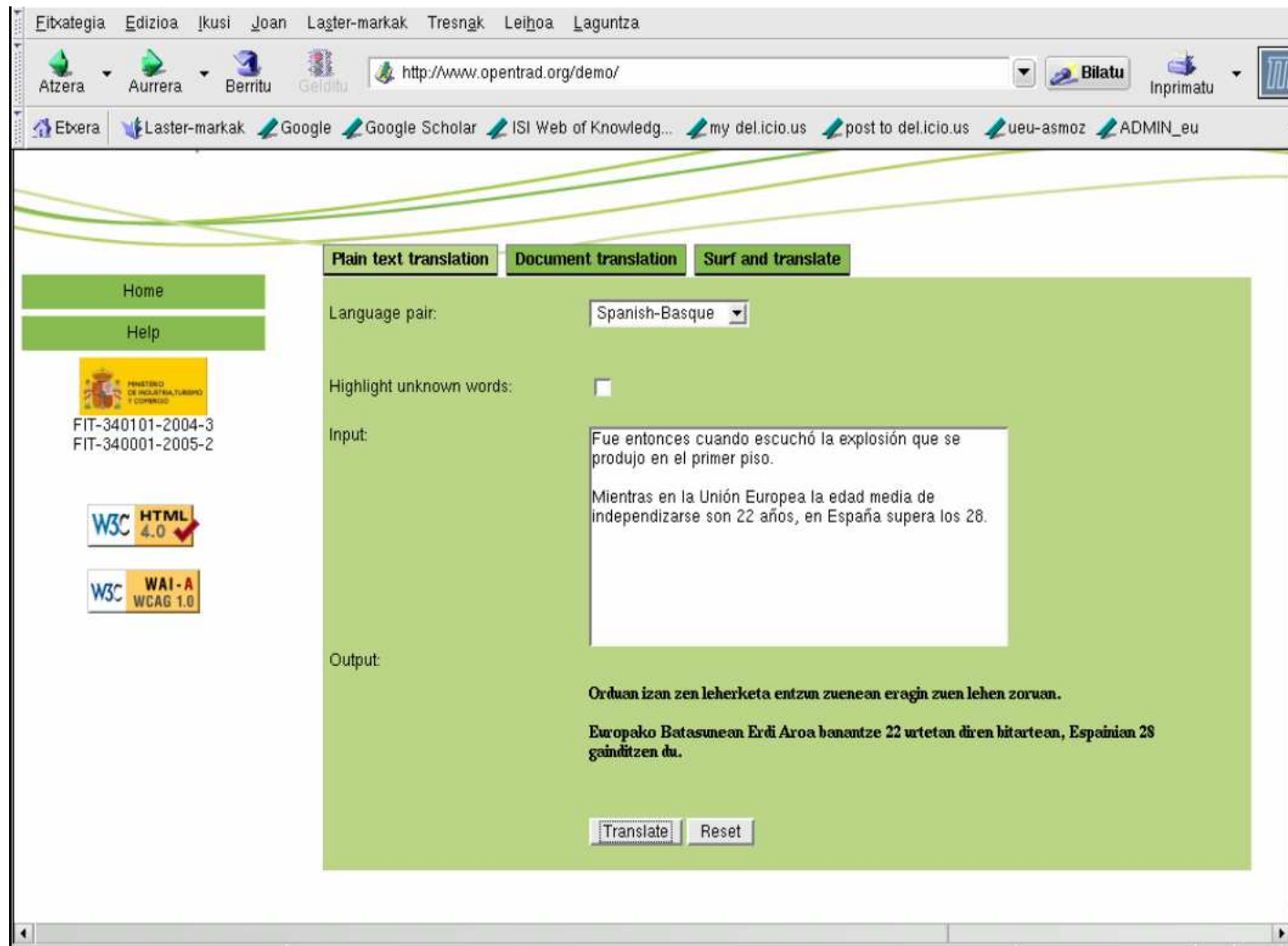
Opciones de
traducción

Resultados

Casos prácticos

Mirando al futuro

Resultados es-eu



The screenshot shows a web browser window with the URL <http://www.opentrad.org/demo/>. The browser's address bar and menu bar are visible. The main content area features three tabs: "Plain text translation", "Document translation", and "Surf and translate". The "Plain text translation" tab is active. The interface includes a "Language pair" dropdown menu set to "Spanish-Basque", a "Highlight unknown words" checkbox, and an "Input" text area containing the Spanish text: "Fue entonces cuando escuchó la explosión que se produjo en el primer piso. Mientras en la Unión Europea la edad media de independizarse son 22 años, en España supera los 28." Below the input is an "Output" area displaying the Basque translation: "Orduan izan zen leherketa entzun zuenean eragin zuen lehen zoruan. Europako Batasunean Erdi Aroa banantze 22 urtetan diren hitartean, Espainian 28 gaintzen du." At the bottom of the output area are "Translate" and "Reset" buttons. On the left side of the page, there is a sidebar with navigation links: "Home", "Help", and a list of items including "FIT-340101-2004-3", "FIT-340001-2005-2", "W3C HTML 4.0", and "W3C WAI-A WCAG 1.0".

Características
generales

Opciones de
traducción

Resultados

Casos prácticos

Mirando al futuro

Resultados es-eu

The screenshot shows the EL PAÍS.es website interface. At the top, there is a navigation bar with links like 'Eitxategia', 'Edizioa', 'Ikusi', 'Joan', 'Laster-markak', 'Tresngak', 'Leihoak', and 'Laguntza'. Below this is a search bar with the URL 'http://www.elpais.es/'. The main header features the 'Especial Directivos' logo and the text 'Revistas Especializadas para Profesionales'. A banner for 'RETRANSMISION' mentions 'Santiago Segurola comentará en directo y con los lectores el Barcelona-Milán a partir de las 20.45 - 18:37 h.'. The main content area is titled 'EL PAÍS.es' and includes a search bar and navigation tabs like 'la portada', 'lo último', 'el indice', 'lo más...', 'mi País', and 'el archivo'. The date is 'Miércoles, 26 de abril de 2006, actualizado a las 18:51 h.'. The main article is titled 'Zidane seguirá ligado al Madrid en un "proyecto relacionado con los niños"'. Other sections include '24HORAS', 'Destacamos', 'FOTOGALERÍA', and 'ENTREVISTA'. The footer contains various icons and a 'ya.com' logo.

Características
generales

Opciones de
traducción

Resultados

Casos prácticos

Mirando al futuro

Resultados es-eu

Fitxategia Edizioa Ikusi Joan Lagter-markak Tresnak Leihoa Laguntza

 This page was translated automatically, this is the [original page](#)

EMANKIZUNA *Santiago Seguro-la Zuzenean iruzkina egingo du eta irakurleekin 2045TARIK AURRERA - *Barcelona-milán 18: 37H.

 + uñetas
EDICIÓN IMPRESA + autonomías
+ suplementos

 **HEMEROTECA**
Archivo de EL PAÍS
desde su fundación

EL PAÍS.es

[Azkena](#) [erakuslea areago](#) [nire herria](#) [artxiboa](#) [azala](#)

Asteazkenean, 2006ko Apirilaren 26an, 18RI: eguneratu 48H.

Nazioartekoa
 España 
 Iritzia
 Gizartea
 Teknologia
 Ekonomia
 Kirolak
 Kultura
 Jendea eta *TV

> *ADSL Banda Zabal *\\
> [Egunekoa](#)
[Titularrak](#)
> [hakarrik](#) *Testo

24HORAS
El primer periódico gratuito actualizado permanentemente
Patrocinado por **IBERIA.COM**
> Elegir otra edición

lo descargas lo imprimes lo lees

FUTBOLA
*Zidane-k lotura jarraikiko dio Madrili " asmo bat erlazionatu haurrekin"

ekin eta ekiten du
*Supremo-ren erabakia errespetatuko duen *OPA DE GAS NATURAL-EN GAINETIK Zapataria


lo descargas lo imprimes lo lees

 **BIDEOA**
*Zidane-k futbolaren agur egiten du

 *FOTOGALERÍA
*Villarreal, *Champions-i bigarrenaren

 **EL KARRIZKETA**
*Marisa Paredes Aktorea

*Dumefald eta *Dica *Ink hitatzen dieta *EE

Características generales

Opciones de traducción

Resultados

Casos prácticos

Mirando al futuro

Aplicación real

Casos prácticos

Ventajas competitivas

- Traducción bidireccional
- Permite la total integración (técnica, lingüística)
- Software libre
- Independencia del proveedor
- Futuros desarrollos de la comunidad de usuarios

Resistencia

- Resistencia al cambio
- Amenaza percibida por lingüistas
- Desconocimiento de la traducción automática

Entorno I. El medio

- La integración en medios de comunicación
- Entornos no restringidos
- La integración en el flujo de trabajo, factor clave de éxito

Entorno II. Postedición

- Múltiples salidas de formato
 - Edición en papel
 - Edición en Web

Una vez implantado

- Permite la perfecta integración
- Facilidad de mantenimiento
- Plataformas abiertas permiten mayor desarrollos y mejor aplicabilidad
- Transparente para los usuarios
- No news, good news

Caso de éxito

- “La Voz de Galicia”
- Integración en La Voz digital
- Integración en el sistema editorial

Opentrad

Traducció automàtica en codi obert
Traducción automática en código abierto
Itzulpen automatikoa kode irekian
Traducción automática en código abierto

Gràcies