

ElIEO Hitzaldiak – Conferencias COIIE 2004

Informatikako Ingeniariei Erronkak. / Retos de los Ingenieros en Informática.”

Europa. La ingeniería Informática tiene un mercado abierto:

TECNOLOGÍA DE LA LENGUA

Ponente: **Kepa Sarasola**

Cargo Profesor de la UPV-EHU

Miembro del grupode investigación IXA



EUSKADIKO INFORMATIKAKO
INGENIARIEN ELKARGO OFIZIALA

COLEGIO OFICIAL DE INGENIEROS
EN INFORMÁTICA DEL PAÍS VASCO

¿Qué es la industria de la lengua?

Busqueda de información remota

Traducción de documentos

Escritura y corrección de textos

Consulta de diccionarios y enciclopedias

Electronic messaging

Aprendizaje de segundas lenguas

Servicios telefónicos automáticos

Cada vez más personas usan ordenadores en su vida diaria. Muchas de las aplicaciones utilizadas incluyen procesamiento del lenguaje



Motivación

- Las *tecnologías de la lengua* son un elemento fundamental para el éxito de la llamada *sociedad de la información*
- La comunicación entre personas y máquinas, a medio plazo, podrá ser llevada a cabo en lenguaje natural, en nuestra propia lengua.
- Serán **herramientas limitadas** y siempre con un grado de error, pero aún así su uso será de **gran ayuda**.



Motivación

Aplicaciones en uso

- En la actualidad ya hay múltiples servicios lingüísticos que facilitan el trabajo al usuario humano
 - Ayuda a la edición y comprensión de textos
 - Traducción automática
 - Tratamiento de grandes masas de texto
 - Tratamiento de voz
 - Enseñanza de segundas lenguas
 - ...



Motivación


Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
 - Correctores ortográficos y de estilo
 - Sistemas de ayuda para la consulta de diccionarios,
- Traducción automática
- Tratamiento de grandes masas de texto
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...

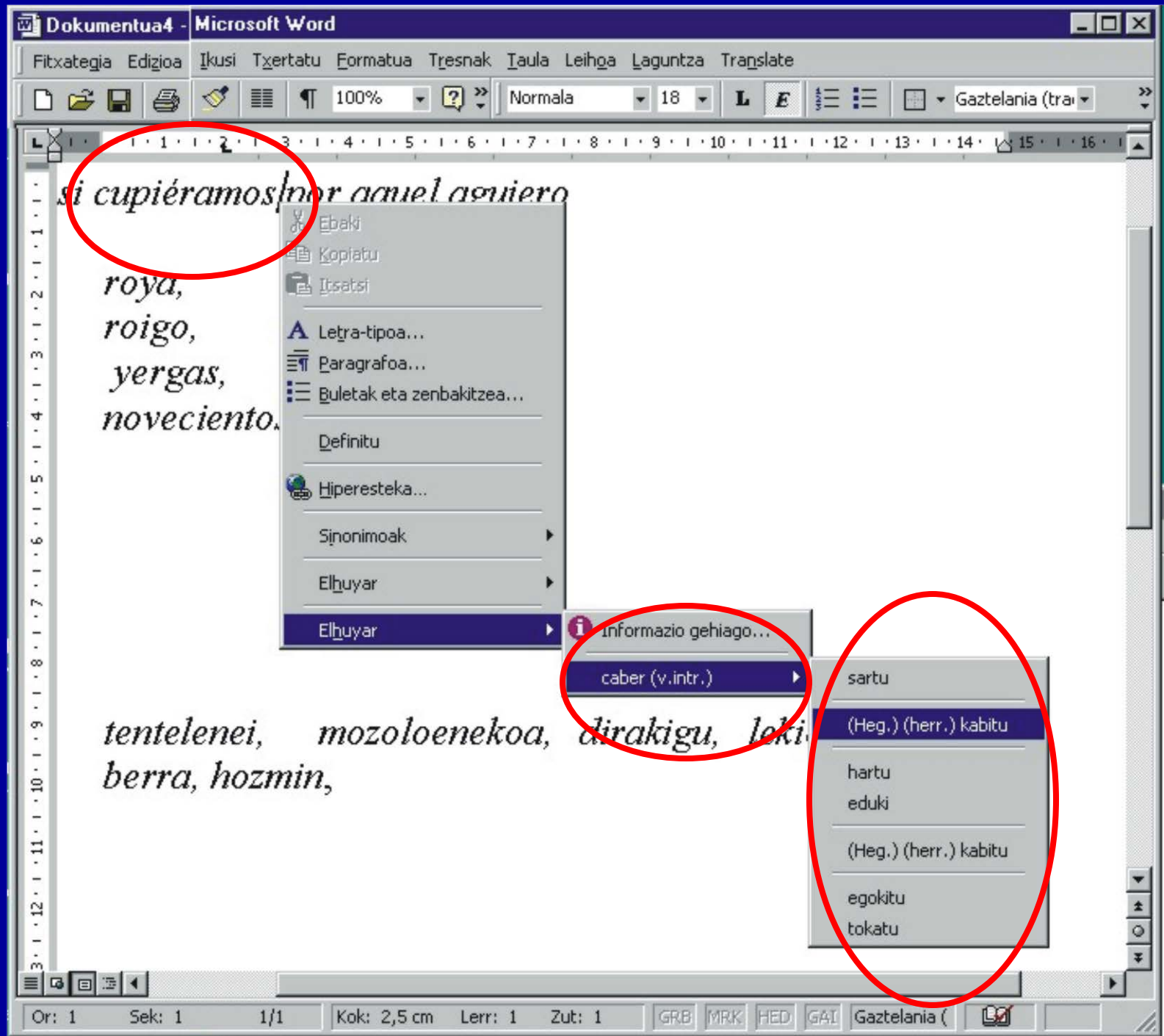


Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
 - Correctores ortográficos y de estilo
 - Sistemas de ayuda para la consulta de diccionarios 
 - On-line con editor de textos. Elhuyar(Cast-Eusk), UZEI sinón.
 - Diccionarios electrónicos avanzados: eEH Euskal Hiztegia
 - Muchísimos diccionarios: www.yourdictionary.com
- Traducción automática
- Tratamiento de grandes masas de texto
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...







your dictionary.com™
The last word in words

[About](#) [Contact](#) [Register](#) [Store](#)

Multilingual Software Testing
World-wide Tester Network
Professional Translating
For individuals and businesses

The Dictionaries: [Language](#) [Multilingual](#) [Phrases & such](#) [Translation](#) [Specialty](#) [Search](#)
[Grammars](#) [Aqora](#) [Archives](#) [Endangered Languages](#) [Gameroom](#) [Lang Guessers](#) [Library](#) [Research](#) [More](#)

WIZCOM
Infoscan™ Infoscan takes notes for you, then guess what?

Relax and Learn
in our
Game Room

Language Dictionaries

There are [6,800 known languages](#) spoken in the 200 countries of the world. 2,261 have writing systems (the others are only spoken) and about 300 are represented by on-line dictionaries as of May 11, 2004. Below are the ones we currently list. New languages and dictionaries are constantly being added to yourDictionary.com; as a result, we have the widest and deepest set of dictionaries, grammars, and other language resources on the web. **New language dictionaries added recently are marked by a • bullet.**

If you are on line, on line resources, traditional, cooperative, to have in a from resources support yourDictionary.com, please go to WorldLanguage through our website.

Diccionarios
accesibles por Internet
para 300 lenguas
www.yourdictionary.com

Get Cheap [Car Insurance](#) [WebGuru—answers to your questions](#) Create your own [Italian charm bracelet on-line!](#)

Language Dictionaries

Abenaki	Afrikaans	Ainu	Aklanon	Alabama	Albanian	Algerian	Amharic
Apache	Arabic	Aramaic	Armenian	ASL	Avestan	Ayapathu	Aymara
Azeri	Balinese	Bamwe	Bantu	Basa	Basque	Belarusan	Bemba
Bengali	Berber	Bicol	Bisaya	Bobangi	Brahui	Breton	Butuanon

Abenaki	Afrikaans	Ainu	Aklanon	Alabama	Albanian	Algerian	Amharic
Apache	Arabic	Aramaic	Armenian	ASL	Avestan	Ayapathu	Aymara
Azeri	Balinese	Bamwe	Bantu	Basa	Basque	Belarusan	Bemba
Bengali	Berber	Bicol	Bisaya	Bobangi	Brahui	Breton	Butuanon
Bukusu	Bulgarian	Burmese	Cantonese	Carib	Catalan	Catawba	Cayuga
Cebuano	Chamorro	Chechen	Cherokee	Chewa	Cheyenne	Chinese	Chinook
Ch'ol	Chorti	Cispa	Coptic	Cree	Creole	Croatian	Czech
Danish	Delaware	Demonh'ka	Denya	Duma	Dutch	Eggon	Ejagham
Egyptian	Ekegusii	Emakua	English	Eskimo	Esperanto	Estonian	Etruscan
Fang	Faroese	Farsi	Finnish	Flemish	Frankish	French	Frisian
• Fulfulde	Gaelic	Galician	Gaulish	Gamilaraay	Ganda	Gbari	Georgian
German	Gevove	Gilbertese	Ginyama	Gothic	Greek	Guarani	Gullah
Haida	Hakka	Hawaiian	Hausa	Hebrew	Hiligaynon	Hmong	Hindi
Holoholo	Hungarian	Icelandic	Igbo				
Inzebi	Irish	Ishkashmi	Italian				
Kannada	Kapampangan	Karelian	Kamviri				
Khakas	Khmer	Khowar	Kiga				
Kongo	Korean	Koromfé	Koyo				
Lingala	Lithuanian	Lojban	Lozi				
Malagasy	Malay	Maliseet	Malayalam				
Manx	Maori	Mapudungun	Marathi				
Miami	Miskitu	Mixtec	Mohawk	Mongolian	Mpongwe	Nanuati	Nande
Nanticoke	Nauruan	Navajo	Ndebele	Nepalese	Newari	Nhirrpi	Norwegian
Nyambo	Nyamwezi	Occitan	Ojibwe	Olkola	Olutec	Onondaga	Oriya
Oscan	Oygangand	Pahlavi	Pakahn	Pali	Papiamentu	Pashto	Pende
Passam...dy	Phoenician	Phrygian	Pidgin	Pirahã	Polish	Popoluca	Portuguese
Potawatomi	Prussian	• Pulaar	Punjabi	Quechua	• Quenya	Rapa Nui	Rasta
Rejang	Romanian	Romany	Romansh	Roviana	Rotuman	Russian	Saanich
Saami	Samoan	Sanskrit	Sardinian	Seneca	Serbian	Seri	Sherpa
Shi	Shona	Shoshoni	Sign	• Sindarin	Sinhalese	Slovak	Slovene
Sogdian	Somali	Sorbian	Spanish	Sranan	Sûdovian	Sumerian	Swabian

Diccionarios
 accesibles por Internet
 para 300 lenguas
 www.yourdictionary.com

Manx	Maori	Mapudungun	Marathi	Masaba	Mayan	Mayangna	Mawu
Miami	Miskitu	Mixtec	Mohawk	Mongolian	Mpongwe	Nahuatl	Nande
Nanticoke	Nauruan	Navajo	Ndebele	Nepalese	Newari	Nhirrpi	Norwegian
Nyambo	Nyamwezi	Occitan	Ojibwe	Olkola	Olutec	Onondaga	Oriya
Oscan	Oyikangand	Pahlavi	Pakahn	Pali	Papiamento	Pashto	Pende
Passam...dy	Phoenician	Phrygian	Pidgin	Pirahã	Polish	Popoluca	Portuguese
Potawatomi	Prussian	Pulaar	Punjabi	Quechua	Quenya	Rapa Nui	Rasta
Rejang	Romanian	Romany	Romansh	Roviana	Rotuman	Russian	Saanich
Saami	Samoan	Sanskrit	Sardinian	Seneca	Serbian	Seri	Sherpa
Shi	Shona	Shoshoni	Sign	Sindarin	Sinhalese	Slovak	Slovene
Sogdian	Somali	Sorbian	Spanish	Sranan	Sûdovian	Sumerian	Swabian
Swahili	Swedish	Tagalog	Taiwanese	Tamasheq	Tahitian	Taino	Tamazight
Tamil	Tarahumara	Tarok	Tatar				
Tibetan	Tlingit	Tocharian	Tongan				
Ulwa	Umbrian	Üqoi	Urdu				
Vietnamese	Votic	Vulcan	Wagiman				
Wyandot	Xhosa	Yaka	Yao				
Zoque	Zulu	.	.				

Diccionarios
 accesibles por Internet
 para 300 lenguas
 www.yourdictionary.com

NEW! Universal Cur
 Candid Reviews of Boston Hot

your  dictionary.com™
 The last word in words

Home | The Dictionaries: [Language](#) | [Specialty](#) | [Multilingual](#) | [Translation](#) | [Software Testing](#)
[Nyms & such](#) | [Grammars](#) | [Language Identifiers](#) | [Research](#) | [Library](#) | [Fun & Games](#) | [More](#)
[About YDC](#) | [Lookup Button](#) | [Store](#) | [Register](#) | [Endangered Languages](#) | [Search](#) | [Contact us](#)

Paid Ads: [Cheap Car Insurance](#) | [Mortgage Advice](#) | [Personal Loans](#) | [Credit Cards](#) | [Concert Tickets](#)
[UK Loans](#) | [Pellicerie-Furs](#) | [Los Angeles California](#) | [Travel Worldwide!](#) | [MMS Logos](#) | [Colorado Architect](#)
 27,000 Reference Articles | [ADVLab](#) | [Fast Search](#) | [Hardware](#) | [Splitta](#) | [Nysite](#) | [St. Augustine Beach FL](#)
[Holiday Insurance](#) | [Mortgages](#) | [Home Insurance](#) | [Mortgage Lenders](#) | [Buy to let Mortgages](#) | [Car insurance](#)
[Gatwick Airport Parking](#) | [Heathrow Airport Parking](#) | [Manchester Airport Parking](#) | [Luton Airport Parking](#)



Basque • Profile

- [Morris English-Basque-English Online Dictionary](#)

- [Short Basque Word List](#)
- [Basque-Esperanto-Basque Word Lists](#)
- [Euskera-okzitano hiztegia](#)
- [Euskara/Gaztelania-Castellano/Vasco Elhuyar Hiztegia](#)
- [Castellano-Basque Dictionary](#)
- [Basque-Castellano Dictionary](#)
- [Basque Thesaurus](#)
- [Harluxet Hiztegi Entziklopedikoa](#)
- [Basque Language Books and Other Resources](#)

For related Basque language resources see:
[Basque grammars, news, fonts](#)
[Basque Language Resources](#)

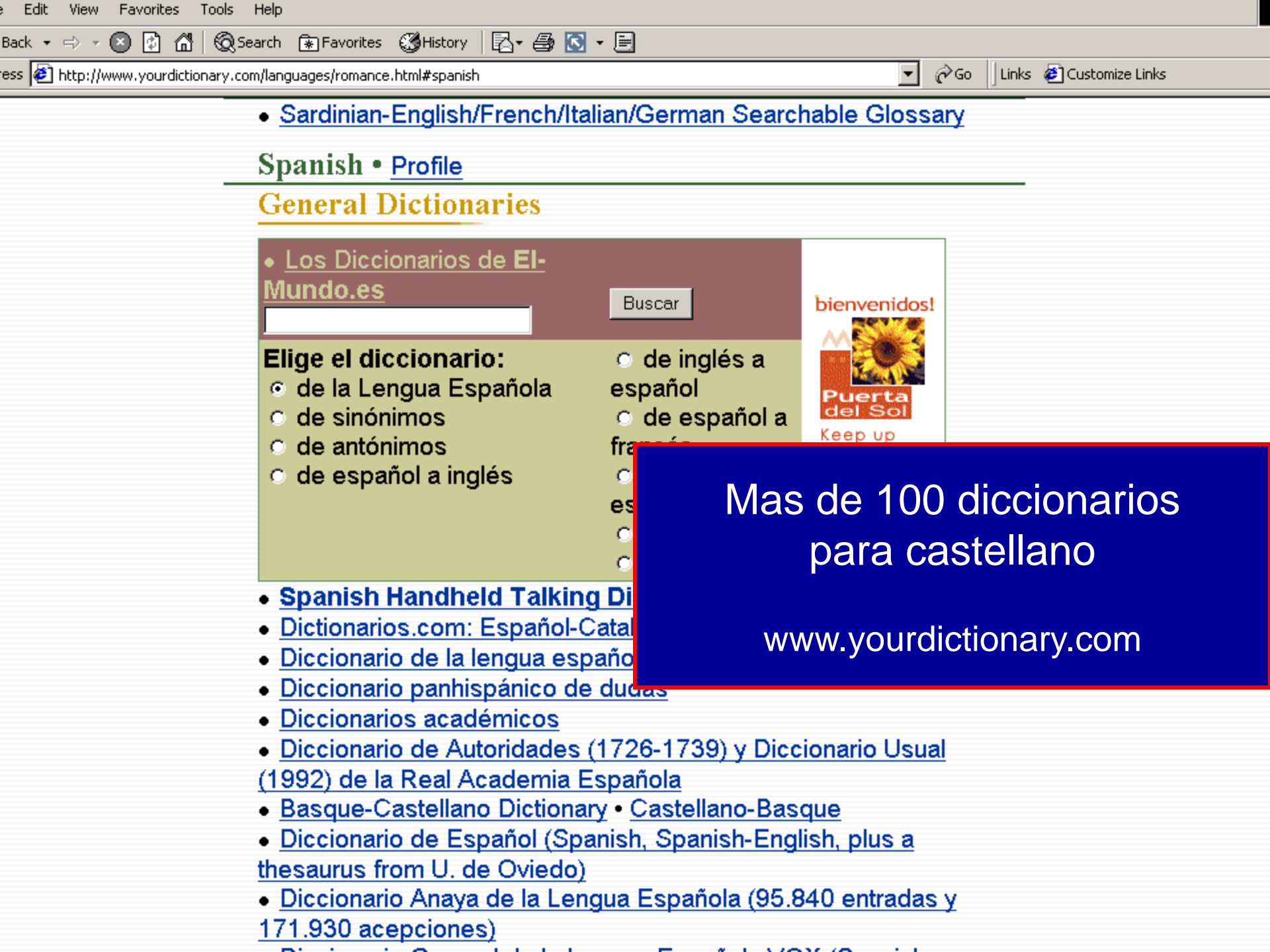
Caucasian

- [Chechen-German Online Dictionary](#)
- [English-Chechen Lexical Database](#)
- [Chechen Language Books and Other Resources](#)
- [Ingush-English Lexical Database](#) • [Profile](#) (currently unavailable)
- [Russian-Ingush Vocabulary](#)
- [Ingush-Russian Vocabulary](#)
- [Russian-Ingush, Ingush-Russian Dictionaries \(RTF format, zip file\)](#)

For related Ingush language resources see:
[Ingush grammars, news, and fonts](#)

9 diccionarios
para euskara

www.yourdictionary.com



- [Sardinian-English/French/Italian/German Searchable Glossary](#)

Spanish • [Profile](#)

General Dictionaries

- [Los Diccionarios de El-Mundo.es](#)

Buscar

Elige el diccionario:

- de la Lengua Española
- de sinónimos
- de antónimos
- de español a inglés

de inglés a español

de español a francés

de francés a español

de español a portugués

de portugués a español

de español a gallego

de gallego a español

bienvenidos!



Puerta del Sol

Keep up

Mas de 100 diccionarios para castellano

www.yourdictionary.com

- [Spanish Handheld Talking Dictionary](#)
- [Diccionarios.com: Español-Catalán](#)
- [Diccionario de la lengua española](#)
- [Diccionario panhispánico de dudas](#)
- [Diccionarios académicos](#)
- [Diccionario de Autoridades \(1726-1739\) y Diccionario Usual \(1992\) de la Real Academia Española](#)
- [Basque-Castellano Dictionary • Castellano-Basque](#)
- [Diccionario de Español \(Spanish, Spanish-English, plus a thesaurus from U. de Oviedo\)](#)
- [Diccionario Anaya de la Lengua Española \(95.840 entradas y 171.930 acepciones\)](#)

[Spanish Language Books and Other Resources](#)

[Specialty Dictionaries](#)

- [Biotic Guia de Recursos Glosario](#)
- [Diccionarios y Glosarios de Biología \(click on Biologicas and find the pages you want\)](#)
- [Diccionario de la contaminación](#)
- [Diccionaria de la Diabetes](#)
- [Diccionario Castellano-Inglés • Inglés-Castellano](#) (Lists of Spanish shipping terminology)
- [Diccionario de Empresa \(Castellano-Portugués\)](#)
- **NEW!** [Enciclopedia Ilustrada de Salud](#)
- **NEW!** [Diccionario médico](#)
- **NEW!** [Diccionario Médico](#)
- **NEW!** [Diccionario de la Salud](#)
- **NEW!** [Medicina de Familia](#)
- [Diccionario ilustrado de términos](#)
- [Spanish Handheld Talking Di](#)
- [Diccionario Terminológico de O](#)
- [English-Spanish Dictionaries of](#)
- [Diccionario de Sinónimos Universidad de Oviedo](#)
- [Diccionario Universal de Términos Parlamentarios](#)
- [Diccionario del Vino Español](#)
- [English to Spanish Business Dictionary \(8,000 entries\)](#) **MIRROR**
- **NEW!** [Glossario de Smileys](#)
- [Spanish Hardback and CD Specialty Dictionaries](#)
- [English-Spanish Gymnastics Terminology and Phrases](#)

**The
Oxford
Business
Spanish
Dictionary**



Diccionarios especializados para castellano
www.yourdictionary.com

"Just when I think I've produced the clearest possible draft, it points out



- [Negligencias.com Online Medical Concordance](#)
- [Pequeño Diccionario Médico Etimológico](#)
- [Random Spanish Idiom Generator](#)
- [Railcar Dictionary - English to Spanish](#)
- [Spanish Food Dictionary](#)
- [Spanish-English Glossary of Mat...](#)
- [Spanish Hardback and CD Sp...](#)
- [A Spanish Morpheme Dictionary](#)
- [Tesoro de la Astronomía \(Español\)](#)
- [Vademécum de Español Urgente](#)
- [Vocabulario Técnico de Marketi...](#)

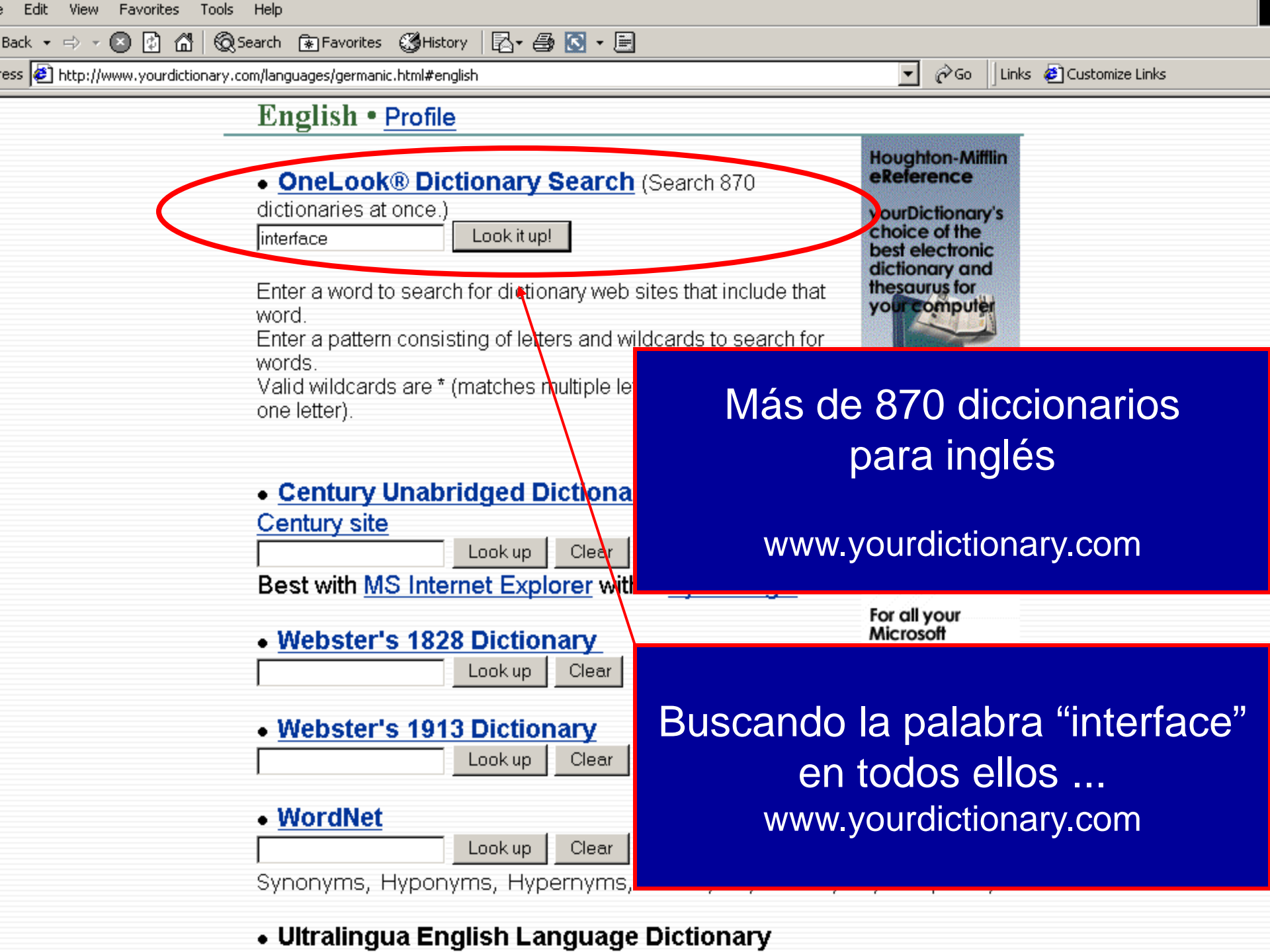
Diccionarios
para variantes del castellano
www.yourdictionary.com

Spanish Dialects

- [Diccionario de Regionalismos de la Lengua Española \(Contiene más de 30.000 acepciones\)](#)
- [Diccionariu Asturianu-Castellán](#)
- [Diccionarios de Jerga, Argot, Slang, Modismo \(with Java midis\)](#)
- **NEW!** [Glossario Argentino-Español](#)
- **NEW!** [Glosario de Jerga Guayaca](#)
- [Glossario del Español usual en México](#)
- **NEW!** [Diccionario del Español usual en México](#)
- [Localismos utilizados en La Ramajería y los Arribes \(Salamanca\)](#)
- [Spanish of the Canary Islands](#)
- [Vocabulario Caló \(Gitano\)- Español \(PDF files online\)](#)
- [Diccionario des modismos \(Chile\)](#)
- **NEW!** [Dictionnaire des régionalismes de la langue espagnole](#)
- **NEW!** [Diccionario de Venezolanismos](#)

For related Spanish language resources see:

Spanish grammars, news, and fonts



English • Profile

- **OneLook® Dictionary Search** (Search 870 dictionaries at once.)

Enter a word to search for dictionary web sites that include that word.
 Enter a pattern consisting of letters and wildcards to search for words.
 Valid wildcards are * (matches multiple letters) and ? (matches one letter).

- **Century Unabridged Dictionary**
[Century site](#)

Best with [MS Internet Explorer](#) with

- **Webster's 1828 Dictionary**

- **Webster's 1913 Dictionary**

- **WordNet**

Synonyms, Hyponyms, Hypernyms,

- **Ultralingua English Language Dictionary**

Houghton-Mifflin
eReference

YourDictionary's
choice of the
best electronic
dictionary and
thesaurus for
your computer

Más de 870 diccionarios
para inglés

www.yourdictionary.com

For all your
Microsoft

Buscando la palabra "interface"
en todos ellos ...

www.yourdictionary.com

Never stop learning! OneLook is sponsored in part by [KnowledgeNews](#).
KnowledgeNews brings the fascinating world of history, science, and culture right to your inbox every weekday.
[Click here to become a free introductory member today!](#)

Word or phrase:

- Find definitions Find translations Search all dictionaries

Jump to: [General](#), [Art](#), [Business](#), [Computing](#), [Medicine](#), Miscellaneous, Religion, [Science](#), Slang, Sports, [Tech](#), [Phrases](#)

Found 49 dictionaries with English definitions that include the word *interface*.

Click on the first link on a line below to go directly to a page where "interface" is defined.

General (13 matching dictionaries)

- [interface](#): Merriam-Webster's Online Dictionary, 1
- [interface](#): Encarta® World English Dictionary, No
- [interface](#): Cambridge International Dictionary of E
- [interface](#): The Wordsmyth English Dictionary-The
- [interface](#): The American Heritage® Dictionary of t
- [interface](#): Infoplease Dictionary [[home](#), [info](#)]
- [interface](#): Dictionary.com [[home](#), [info](#)]
- [interface](#): UltraLingua English Dictionary [[home](#), [in](#)
- [interface](#): Cambridge Dictionary of American English [[home](#), [info](#)]
- [Interface \(object-oriented programming\)](#), [Interface](#): Wikipedia, the Free Encyclopedia [[home](#), [info](#)]
- [interface](#): Rhymezone [[home](#), [info](#)]
- [interface](#): WordNet 1.7 Vocabulary Helper [[home](#), [info](#)]
- [interface](#): LookWAYup Translating Dictionary/Thesaurus [[home](#), [info](#)]

Buscando "interface"...
Aparece en 49 diccionarios
www.yourdictionary.com

(Interface)
... surface forming a
... between two things
... or chemical
... (science) computer
... hardware and
... links one device
... with another (especially a computer and a
... hard disk drive or other peripherals)
• **noun:** the overlap where two
theories or phenomena affect each other
or have links with each other (Example:
"The interface between chemistry and

➔ **Business** (4 matching dictionaries)

18. [Interface](#) : MoneyGlossary.com [[home](#), [info](#)]
19. [interface](#) : Travel Industry Dictionary [[home](#), [info](#)]
20. [Interface](#) : Homeglossary.com [[home](#), [info](#)]
21. [Interface](#) : Construction Term Glossary [[home](#), [info](#)]

➔ **Computing** (11 matching dictionaries)

22. [interface](#) : Free On-line Dictionary of Computing [[home](#), [info](#)]
23. [interface](#) : Netlingo [[home](#), [info](#)]
24. [interface](#) : Hutchinson Dictionary of Computers, Multimedia, and the Internet [[home](#), [info](#)]
25. [interface](#) : CCI Computer [[home](#), [info](#)]
26. [Interface](#) : Cybernetics and Systems [[home](#), [info](#)]
27. [Interface](#) : Butterfly Glossary (networking terminology) [[home](#), [info](#)]
28. [Interface](#) : TECHNICAL [[home](#), [info](#)]
29. [interface](#) : Dictionary of Computing and Digital Media
30. [interface](#) : whatis? [[home](#), [info](#)]
31. [Interface](#) : Karbo's Dictionary [[home](#), [info](#)]
32. [interface](#) : Webopedia [[home](#), [info](#)]

➔ **Medicine** (3 matching dictionaries)

In general usage, an **interface** is the point, area, or surface along which two substances or other qualitatively different things meet; it is also used metaphorically for the juncture between items.\n\nThe verb **to interface** means to interconnect two or more entities at a common point or shared boundary, or to prepare either entity for that purpose.\n\nThe word **interface** also has...
([continued](#) at Wikipedia)

Buscando "interface"...
aparece en 11 diccionarios
sobre "Computing"

Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
 - Generación de borradores para traducción ←
 - Comprensión superficial de documentos (Web)
 - Memorias de traducción
- Tratamiento de grandes masas de texto
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...





Sistema de traducción automática castellano-catalán

➔ Revise el texto traducido

Compruebe también la ortografía del texto

Hola *Montse,
On aneu de vacances aquest any?
Nosaltres anirem A Galícia al setembre

➔ Continúe probando la versión actual

Introduzca o pegue un texto ANSI (hasta 1000 caracteres). Indique el sentido de la traducción, el tipo de salida

Sentido

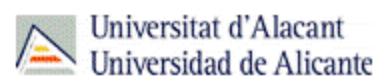
Tipo de salida

Hola Montse,
¿Donde vais de vacaciones este año?
Nosotros iremos a Galicia en setiembre.

- Menú principal
- ➔ Versión estándar
- Versión grandes volúmenes
- Solicitar acceso completo
- Evaluaciones
- Introduzca palabras
- El equipo de interNOSTRUM
- Información técnica
- Contacte con nosotros
- Libro de visitas
- Enlácenos
- Preguntas más frecuentes
- Analizadores morfológicos

- Traducir...
- ➔ Texto
- Correo
- Documentos y websites
- Traducir y navegar
- Chat

Traducción instantanea por web y gratis catalán<->castellano
www.internostrum.com



TIC-2000-1599-C02-02

Desarrollado por

Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
 - Generación de borradores para traducción
 - Comprensión superficial de documentos (Web) ←
 - Memorias de traducción
- Tratamiento de grandes masas de texto
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...





La Web [Imágenes](#) [Grupos](#) [Directorio](#) [News ¡Nuevo!](#)

[Búsqueda Avanzada](#)[Preferencias](#)

Búsqueda: la Web páginas en español páginas de España

La Web

Resultados 1 - 10 de aproximadamente 594,000 de J.K. Rowling . (0.18 segundos)

[J.K. Rowling Official Site - Harry Potter and more](#) - [[Traduzca esta página](#)]

Jump aboard the Hogwart Express to come on a fantastic adventure at the official JK Rowling Website. See what's on my desk. ... All rights reserved JK Rowling. ...

[www.jkrowling.com/](#) - 4k - 21 Jun 2004 - [En caché](#) - [Páginas similares](#)

[Harry Potter: Meet JK Rowling](#) - [[Traduzca esta página](#)]

JK Rowling has won the Hugo Award, the Bram Stoker Award, the Whitbread Award for Best Children's Book, a special commendation for the Anne Spencer Lindbergh ...

[www.scholastic.com/harrypotter/author/](#) - 23k - [En caché](#) - [Páginas similares](#)

[Harry Potter](#) - [[Traduzca esta página](#)]

The Scholastic Store. Select. ...

[www.scholastic.com/harrypotter/home.asp](#) - 16k - 21 Jun 2004 - [En caché](#) - [Páginas similares](#)

[[Más resultados de www.scholastic.com](#)]

[JK Rowling](#) - [[Traduzca esta página](#)]

Like that of her own character, Harry Potter, JK Rowling's life has the luster of a fairy tale. ... Courtesy of Book Magazine. An Interview with JK Rowling. ...

[www.kidsreads.com/harrypotter/jkrowling.html](#) - 26k - [En caché](#) - [Páginas similares](#)

[JK Rowling Teacher Resource File](#) - [[Traduzca esta página](#)]

JK Rowling Teacher Resource File. Welcome to the Internet School Library Media

Esta página ha sido [traducida automáticamente](#) del inglés. [Ver la Página Web Original](#)

Versión para Imprimir Volver a los Resultados

Hogar Escolástico | Sobre Nosotros | Mapa Del Sitio



Búsqueda | Aislamiento | Servicio De Cliente


Kids Fun Online HOME

SELECT YOUR FAVORITE

Selecto

[GAMES & CONTESTS](#)
[CARD FACTORY](#)
[DOWNLOADS 260](#)
[KIDS COUNCIL](#)
[KID NEWS](#)
[BOOK CENTRAL](#)
[TV TUNE-IN](#)

GO

- Home
- 
- About the Books**
- [Pronunciation Guide](#)
 - [Discussion Guides](#)
 - +** [Meet J.K. Rowling](#)
 - [Meet Mary GrandPré](#)
- Reference Guide**
- [Discussion Chamber](#)
 - [Wizard Challenge](#)
 - [Screensaver](#)

Harry Potter

About the Books:
Meet J.K. Rowling

"soy una persona extraordinario afortunada, haciendo lo que amo lo más mejor posible en el mundo. Soy seguro que seré siempre escritor. Era bastante maravilloso justo que se publicará. La recompensa más grande es el entusiasmo de los lectores."

Wizard Challenge

See what questions are new in the Harry Potter Wizard Challenge!




- Live Interview Transcripts
- +** [For Comic Relief March 12, 2001](#)
 - +** [On Scholastic.com Oct. 16, 2000](#)
 - +** [On Scholastic.com from Feb. 3, 2000](#)



HOME

GAMES &
CONTESTSCARD
FACTORYDOWNLOADS
260KIDS
COUNCILKID
NEWSBOOK
CENTRALTV
TUNE-IN

THE SCHOLASTIC STORE

SELECT YOUR FAVORITE

Select

GO

Home



About the Books

[Pronunciation Guide](#)
[Discussion Guides](#)
[Meet J.K. Rowling](#)
[Meet Mary GrandPré](#)
[Reference Guide](#)
[Discussion Chamber](#)
[Wizard Challenge](#)
[Screensaver](#)

Harry Potter

About the Books: Meet J.K. Rowling

"I am an extraordinarily lucky person, doing what I love best in the world. I'm sure that I will always be a writer. It was wonderful enough just to be published. The greatest reward is the enthusiasm of the readers."

Harry Potter's magic has touched a huge audience of all ages all over the world. In America, there

Wizard Challenge

See what questions are new in the Harry Potter Wizard Challenge!



Live Interview
Transcripts

[For Comic Relief
March 12, 2001](#)

[On Scholastic.com
Oct. 16, 2000](#)

[On Scholastic.com
from Feb. 3, 2000](#)

[Read 'Conversations
with J.K. Rowling'](#)



Traducción
de páginas web
por web
y gratis
catalán<->castellano
www.internostrum.com

- 01.05.2004 [El Gnome 2.6](#)
- 22.04.2004 [.cat i .ct, per què dos dominis?](#)
- 20.04.2004 [Recursos lingüístics en codi lliure](#)

XML
>>> [més articles](#)

ROVETATS

CatScript.
Facilita als programadors

Utilitats

Eines de fitxers, discs...

Documentació

Manuals, tutorials i documents.

Fines d'Internet. Clients de correu electrònic, FTP, IRC..

Educació. Matemàtiques, ciències, llengües...

Llengua. glossaris, correctors, traductors...

Ofimàtica. processadors de textos, fulls de càlcul...

Notícies

L'OpenOffice.org cada vegada més útil per a les llengües minoritàries i minoritzades

Cerca dins de Softcatalà

Cerca

TOP 10

1. Open Office	2337
2. Emule	1958
3. WinRAR	1510
4. Mozilla	1300
5. WordCorrect	1254
6. FireFox	867
7. Music Time	832
8. Navegador 4.76	603
9. Salt 2.0	586
10. Ara	510

DESTACATS

Ajuda a promocionar Softcatalà. [Més...](#)



Esta página ha sido traducida automáticamente

Página traducida de www.softcatalà.com

www.internostrum.com



Softcatalà.f.Organización sin ánimo de lucro que fomenta el uso del catalán en la web que recopila los proyectos de Softcatalà y todo el software existente en catalán

ARTICLES

- 06.06.2004 [Los símbolos de Catalunya a la Red](#)
- 23.05.2004 [Declaración de Barcelona para el adelanto del software libre](#)
- 01.05.2004 [El Gnome 2.6](#)
- 22.04.2004 [.cat y .ct, por qué dos dominios?](#)
- 20.04.2004 [Recursos lingüísticos en código libre](#)

El Rebot

programari en català

Navegadores de Internet , el Mozilla, el Navegador...	Herramientas de Internet , Clientes de correo electrónico, FTP, IRC..
Gestión , facturación, contabilidad...	Educación , Matemáticas, ciencias, lenguas...
Utilidades , Herramientas de ficheros, discos...	Lengua , glosarios, correctores, traductores...
Documentación , Manuales, tutoriales y documentos.	Ofimática , procesadores de textos, hojas de cálculo...

Cerca

TOP 10

1 . Open Office	2337.
2 . Emulo	1958.
3 . WinRAR	1510.
4 . Mozilla	1300.
5 . WordCorrect	1254.
6 . FireFox	867.
7 . Music Time	832.
8 . Navegador 4.76	603.
9 . Salt 2.0	586.
10 . Ahora	510.

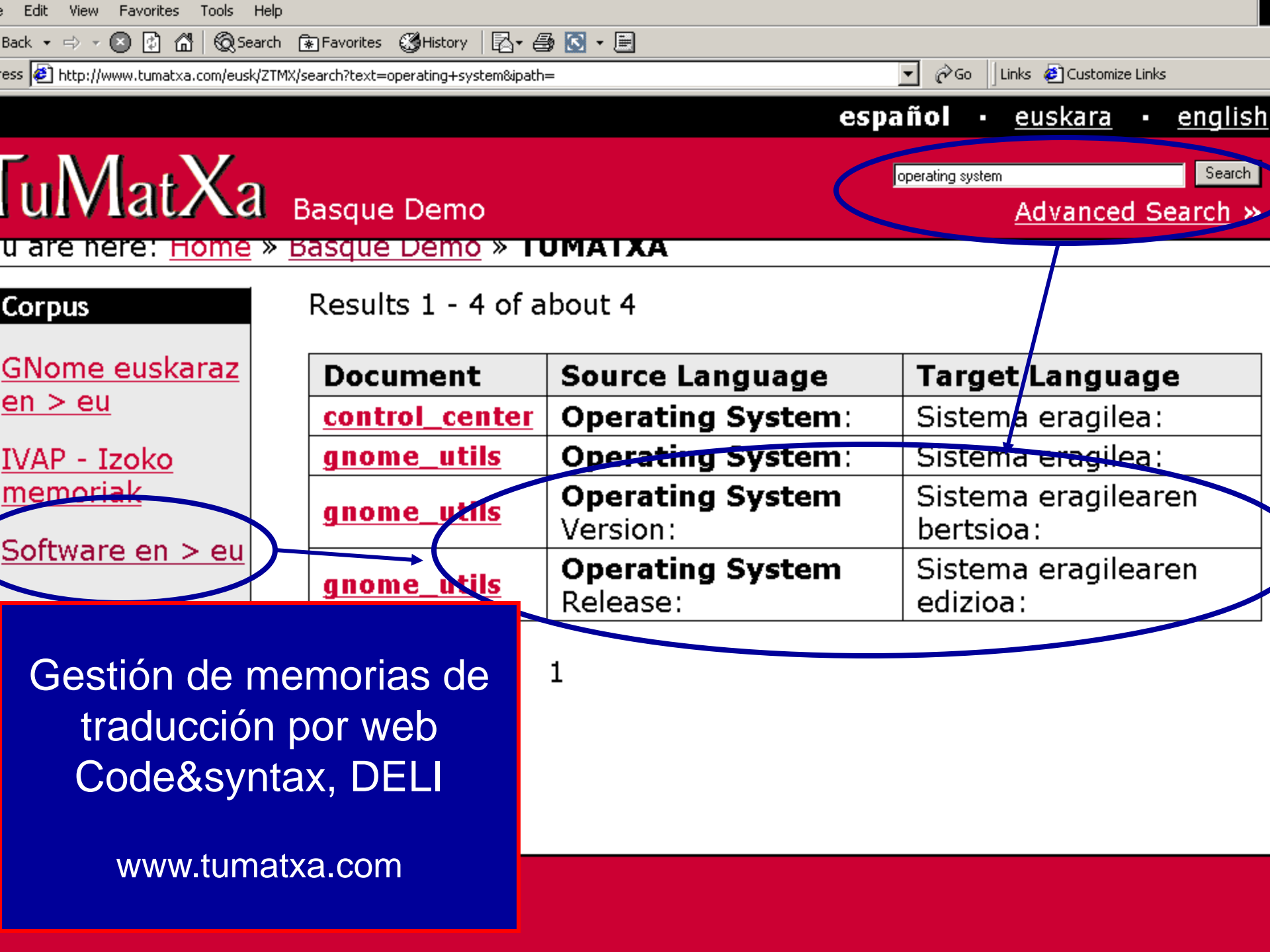
Noticies

Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
 - Generación de borradores para traducción
 - Comprensión superficial de documentos (Web)
 - Memorias de traducción
- Tratamiento de grandes masas de texto
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...





Corpus

[GNome euskaraz en > eu](#)

[IVAP - Izoko memoriak](#)

[Software en > eu](#)

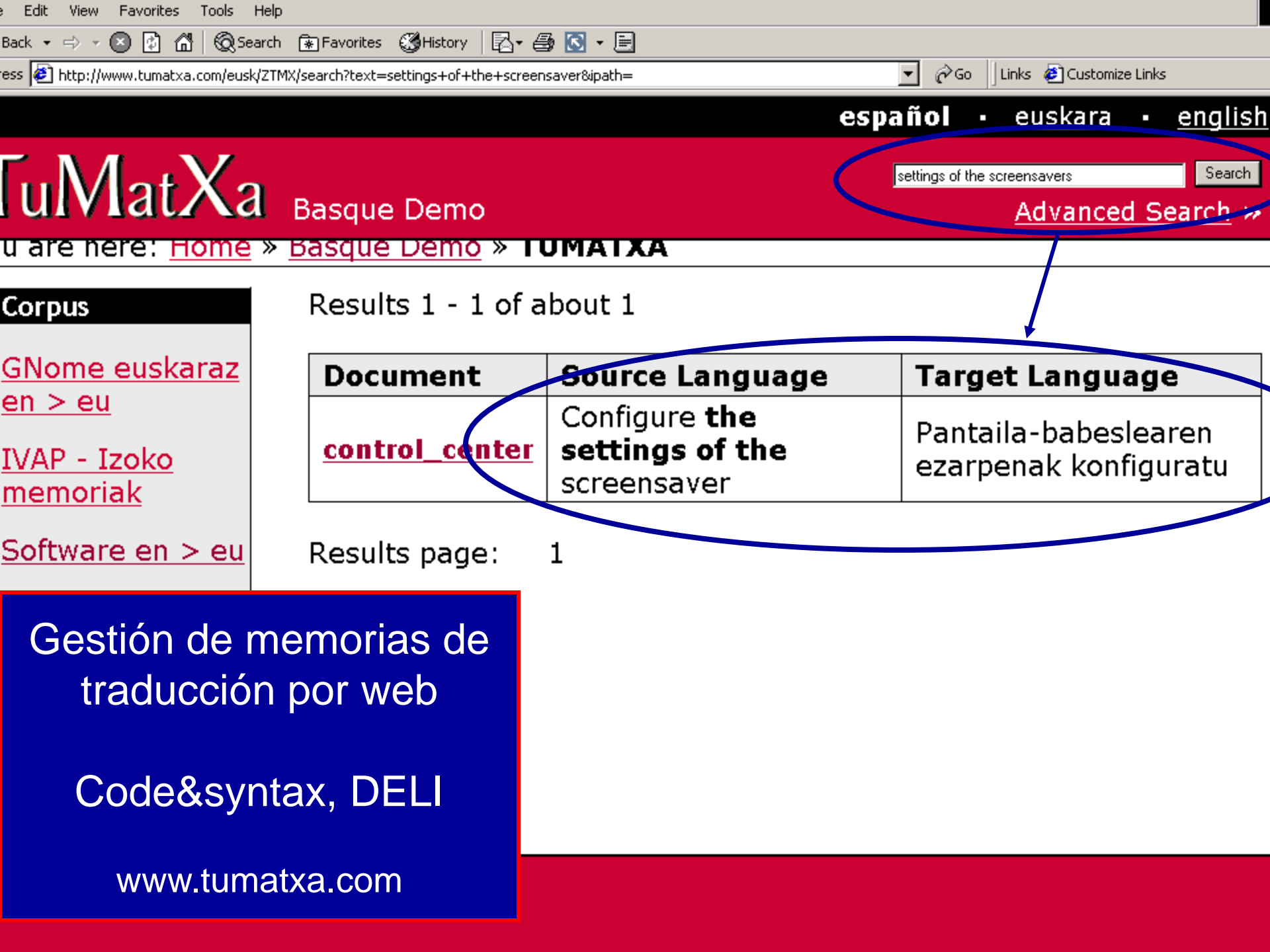
Results 1 - 4 of about 4

Document	Source Language	Target Language
control_center	Operating System:	Sistema eragilea:
gnome_utils	Operating System:	Sistema eragilea:
gnome_utils	Operating System Version:	Sistema eragilearen bertsioa:
gnome_utils	Operating System Release:	Sistema eragilearen edizioa:

1

Gestión de memorias de traducción por web
Code&syntax, DELI

www.tumatxa.com



[Advanced Search](#)

you are here: [Home](#) » [Basque Demo](#) » **TUMATXA**

Corpus

[GNome euskaraz en > eu](#)

[IVAP - Izoko memoriak](#)

[Software en > eu](#)

Results 1 - 1 of about 1

Document	Source Language	Target Language
control center	Configure the settings of the screensaver	Pantaila-babeslearen ezarpenak konfiguratu

Results page: 1

Gestión de memorias de traducción por web

Code&syntax, DELI

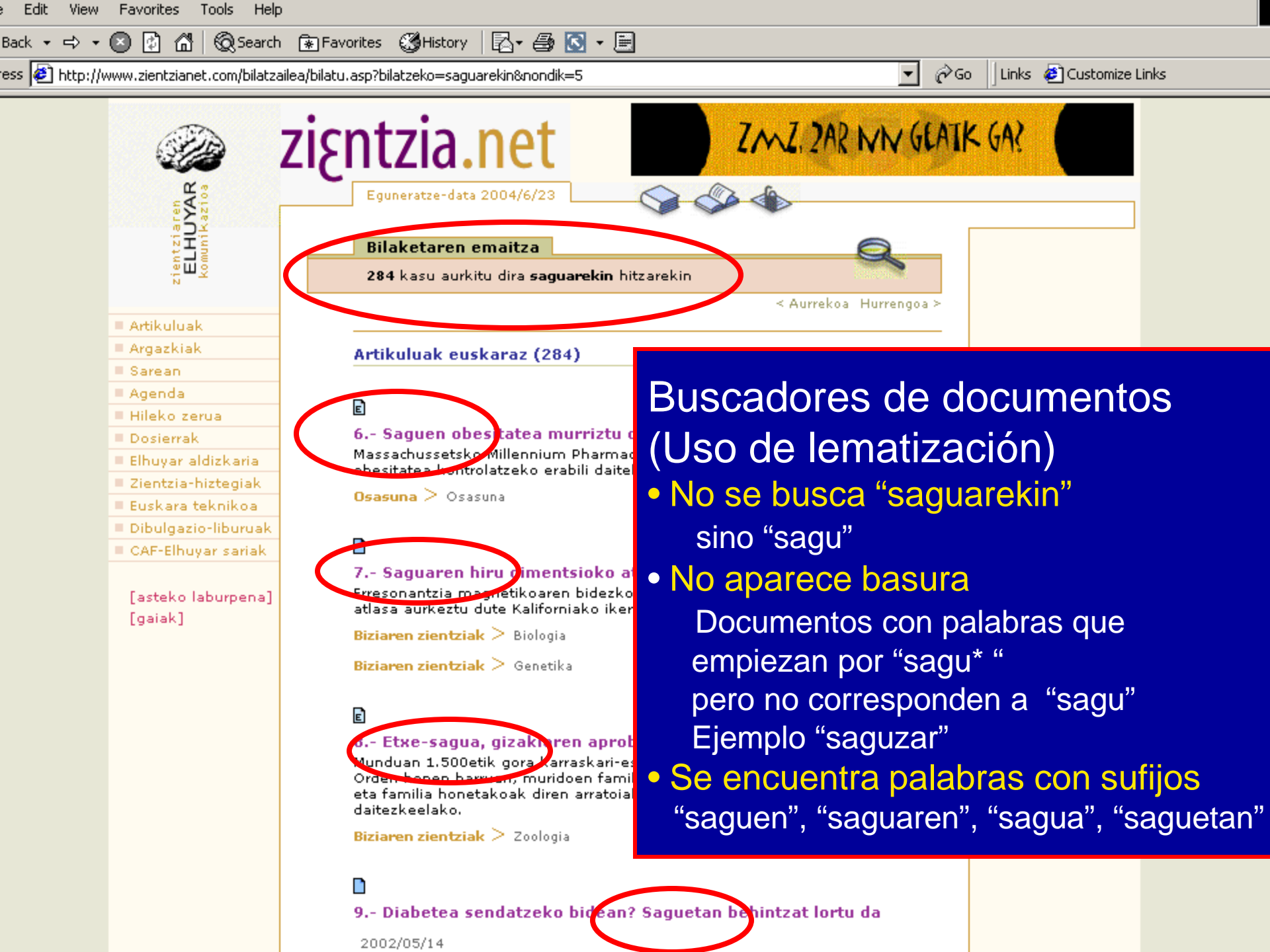
www.tumatxa.com

Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
- Tratamiento de grandes masas de texto
 - Buscadores de documentos (~Google) ←
 - Buscadores de datos en documentos
 - Resúmenes automáticos
 - Question-Answering
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...





zientziaren
ELHUYAR
komunikazioa

zientzia.net

Eguneratze-data 2004/6/23

Bilaketaren emaitza

284 kasu aurkitu dira **saguarekin** hitzarekin

< Aurrekoa Hurrengoa >

Artikuluak euskaraz (284)



6.- Saguaren obesitate murriztu...

Massachusettsko Millennium Pharmaco...
obesitatea kontrolatzeko erabili daite...

Osasuna > Osasuna



7.- Saguaren hiru gimentsioko a...

Eresonantzia magnetikoaren bidezko
atlasa aurkeztu dute Kaliforniako iker...

Biziaren zientziak > Biologia

Biziaren zientziak > Genetika



8.- Etxe-sagua, gizakiaren aprob...

Munduan 1.500etik gora barraskari-es...
Orden honen barruan, muridoen fami...
eta familia honetakoak diren arratoia...
daitezkeelako.

Biziaren zientziak > Zoologia



9.- Diabetea sendatzeko bidean? Sagueta...

2002/05/14

Buscadores de documentos (Uso de lematización)

- No se busca "saguarekin" sino "sagu"
- No aparece basura
Documentos con palabras que empiezan por "sagu*" pero no corresponden a "sagu"
Ejemplo "saguzar"
- Se encuentra palabras con sufijos
"saguen", "saguaren", "sagua", "saguetan"



ORDIMENAREN LAPURRAK 27,95 € ELHUYAR Fundazioaren CD-ROM berria

Eguneratze-data 2004/6/23

Bilaketaren emaitza

33 kasu aurkitu dira **saguzar** hitzarekin



Hurrengoa >

- Artikuluak
- Argazkiak
- Sarean
- Agenda
- Hileko zerua
- Dosierrak
- Elhuyar aldizkaria
- Zientzia-hiztegiak
- Euskara teknikoa
- Dibulgazio-liburuak
- CAF-Elhuyar sariak

Artikuluak euskaraz (33)

1.- Airean harrapatutako txoriak jaten dituzten saguzarrak

2001/08/07
Espainiako hainbat zientzi dituen txoriak jaten dituel...

Biziaren zientziak > Biolo...

Biziaren zientziak > Zoolo...

2.- Saguzarren birsa

Australiako bitanleek kon saguzarrak bizi diren ingur batek harramazka egin or...

Osasuna > Osasuna

3.- Saguzar nekazarale

Saguzar askok intsektuak jaten dituzte, baina badaude begetarriak diren beste batzuk ere.

Biziaren zientziak > Botanika

Biziaren zientziak > Zoologia

Buscadores de documentos (Uso de lematización)

- No aparece basura

Los 33 documentos encontrados para "saguzar" no aparecían al buscar "sagu" (284 documentos)

Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
- Tratamiento de grandes masas de texto
 - Buscadores de documentos (~Google)
 - Buscadores de datos en documentos
 - Resúmenes automáticos
 - Question-Answering
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...

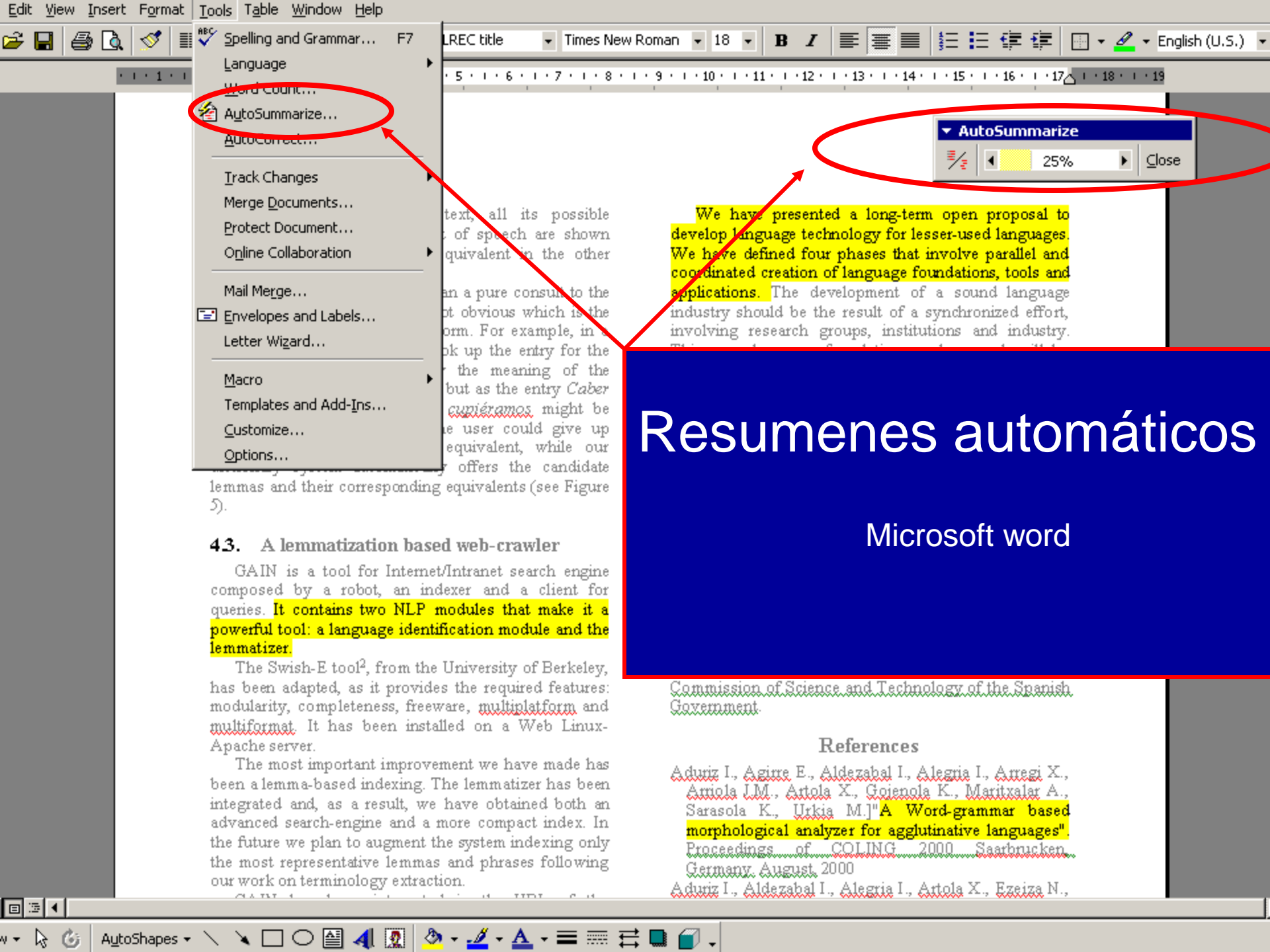


Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
- Tratamiento de grandes masas de texto
 - Buscadores de documentos (~Google)
 - Buscadores de datos en documentos
 - Resúmenes automáticos
 - Question-Answering
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...





- Spelling and Grammar... F7
- Language
- Word Count...
- AutoSummarize...**
- AutoCorrect...
- Track Changes
- Merge Documents...
- Protect Document...
- Online Collaboration
- Mail Merge...
- Envelopes and Labels...
- Letter Wizard...
- Macro
- Templates and Add-Ins...
- Customize...
- Options...

AutoSummarize

25%

Close

Resumenes automáticos

Microsoft word

4.3. A lemmatization based web-crawler

GAIN is a tool for Internet/Intranet search engine composed by a robot, an indexer and a client for queries. It contains two NLP modules that make it a powerful tool: a language identification module and the lemmatizer.

The Swish-E tool², from the University of Berkeley, has been adapted, as it provides the required features: modularity, completeness, freeware, multiplatform and multiformat. It has been installed on a Web Linux-Apache server.

The most important improvement we have made has been a lemma-based indexing. The lemmatizer has been integrated and, as a result, we have obtained both an advanced search-engine and a more compact index. In the future we plan to augment the system indexing only the most representative lemmas and phrases following our work on terminology extraction.

We have presented a long-term open proposal to develop language technology for lesser-used languages. We have defined four phases that involve parallel and coordinated creation of language foundations, tools and applications. The development of a sound language industry should be the result of a synchronized effort, involving research groups, institutions and industry.

Commission of Science and Technology of the Spanish Government.

References

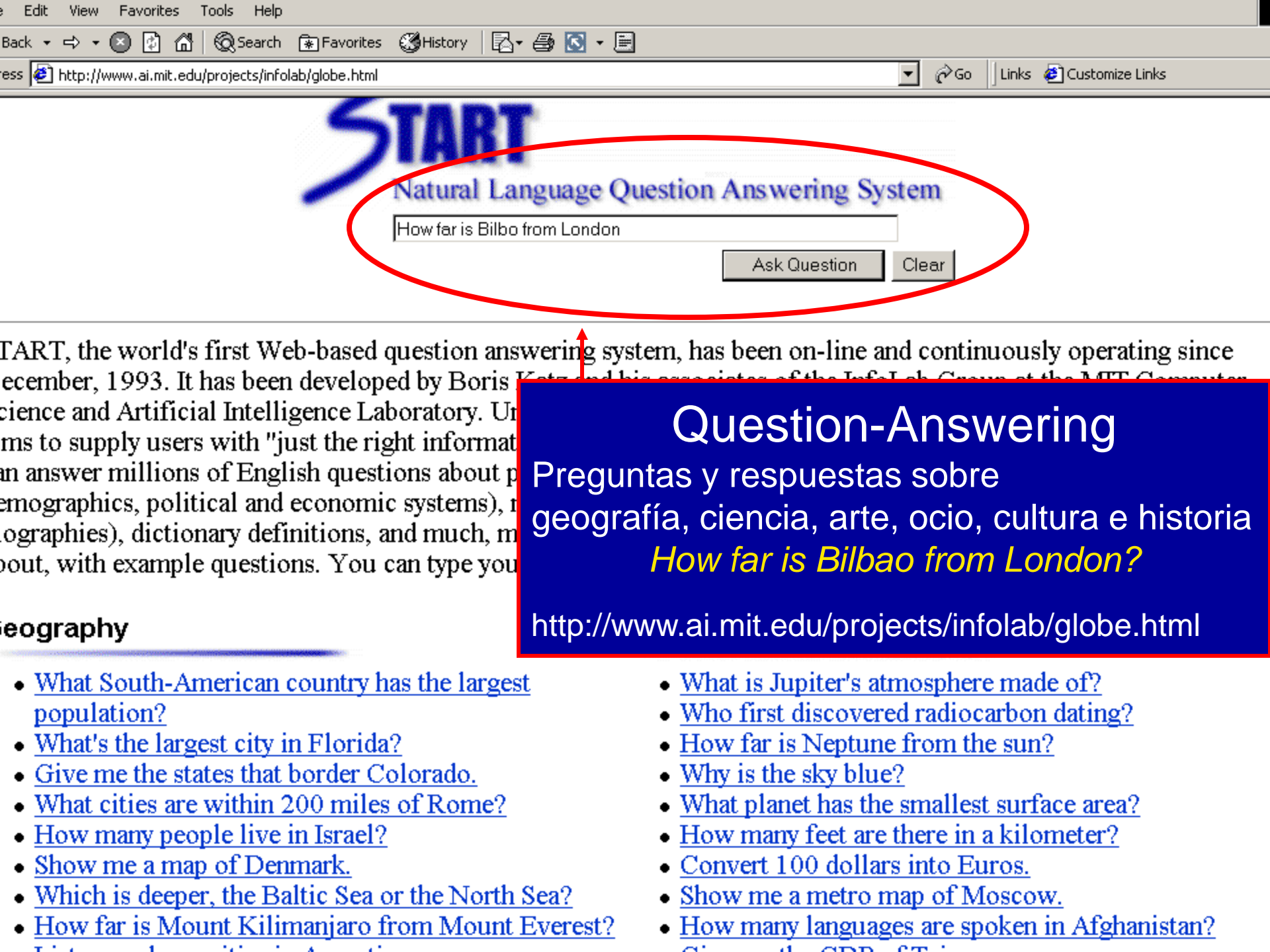
- Aduriz I., Agirre E., Aldezabal I., Alegria I., Arregi X., Arriola JM., Artola X., Goienola K., Maritxalar A., Sarasola K., Urkia M.] "A Word-grammar based morphological analyzer for agglutinative languages". Proceedings of COLING 2000 Saarbrücken, Germany, August, 2000
- Aduriz I., Aldezabal I., Alegria I., Artola X., Ezeiza N.,

Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
- Tratamiento de grandes masas de texto
 - Buscadores de documentos (~Google)
 - Buscadores de datos en documentos
 - Resúmenes automáticos
 - Question-Answering
- Tratamiento de voz
- Enseñanza de segundas lenguas
- ...





START

Natural Language Question Answering System

Ask Question Clear

START, the world's first Web-based question answering system, has been on-line and continuously operating since December, 1993. It has been developed by Boris Katz and his associates of the InfoLab Group at the MIT Computer Science and Artificial Intelligence Laboratory. Our system aims to supply users with "just the right information" (an answer millions of English questions about population demographics, political and economic systems), (geographies), dictionary definitions, and much, much more. Here are some example questions. You can type your own question in the box above.

Geography

- [What South-American country has the largest population?](#)
- [What's the largest city in Florida?](#)
- [Give me the states that border Colorado.](#)
- [What cities are within 200 miles of Rome?](#)
- [How many people live in Israel?](#)
- [Show me a map of Denmark.](#)
- [Which is deeper, the Baltic Sea or the North Sea?](#)
- [How far is Mount Kilimanjaro from Mount Everest?](#)

- [What is Jupiter's atmosphere made of?](#)
- [Who first discovered radiocarbon dating?](#)
- [How far is Neptune from the sun?](#)
- [Why is the sky blue?](#)
- [What planet has the smallest surface area?](#)
- [How many feet are there in a kilometer?](#)
- [Convert 100 dollars into Euros.](#)
- [Show me a metro map of Moscow.](#)
- [How many languages are spoken in Afghanistan?](#)

Question-Answering
 Preguntas y respuestas sobre geografía, ciencia, arte, ocio, cultura e historia
How far is Bilbao from London?
<http://www.ai.mit.edu/projects/infolab/globe.html>



START's reply

How far is Bilbao from London

The distance between Bilbao, Spain and London, England is 602 miles (971 kilometers).

Source: START KB

- [Go back to the START dialog window.](#)

Question-Answering
How far is Bilbao from London?
602 Miles, 971 km
<http://www.ai.mit.edu/projects/infolab/globe.html>

Natural Language Interfaces / Lingubots

- [ELIZA online](#)
([Open Mind](#))
Link added on 2003-04-03
- [WEBCHAT. a geographical database with a natural language interface](#)
(Technical University of Madrid)
Link added on 2003-04-03
- [Natural language interfaces to geographical information \(Globe Server\) and information about the MIT AI Lab](#)
(MIT InfoLab)
Link added on 2003-04-03
- [Ask Jeeves A natural language interface to the WWW](#)
(Ask Jeeves, Inc.; Berkeley, CA)
Link added on 2003-04-03
- [BusTUC The Bus Route Oracle - a natural language interface to local bus departures](#)
(Technical University of Trondheim, Norway)
Link added on 2003-04-03
- [ELF A natural language interface to the N](#)
(ELF Software Co., New York)
Link added on 2003-04-03
- [ELF A natural language interface to a Ho](#)
(ELF Software Co., New York)
Link added on 2003-04-03
- [1001 Questions. An AI system that learn](#)
(Openmind)
Link added on 2003-04-03

Question-Answering
Demos de otros 7 sistemas

Interactive online CL Demos

<http://www.ifi.unizh.ch/CL/InteractiveCLtools/index.php>

Motivación

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
- Tratamiento de grandes masas de texto
- Tratamiento de voz
 - Sistemas que leen textos
 - Sistemas reconocedores del habla
- Enseñanza de segundas lenguas
- ...



AhoTTS Testu-ahots bihurgailuaren demoa

Sartu esaldi bat edo gehiago, eta sakatu *Bihurtu* testu ahots bihurketa entzuteko:

Kaixo, euskararako testu-ahots bihurgailua naiz.

Bihurgailua konfiguratz

Fujisaki-ren intonazio eredua [v]
Tonu normala [v]

Bihurtu Ezabatu

TTS(Text to Speech)
Sistemas que leen textos

Aholab (Ingenieros Bilbao)
http://bips.bi.ehu.es/tts/tts_es.html

Hemen dauzkazu adibide batzuk, klikatu esaldi batean, eta gero *Bihurtu*.

- Kaixo, euskararako testu-ahots bihurgailua naiz.
- Hegoak ebaki banizkion, nirea izango zen.
- Aita-semea tabernan daude, ama-alaba iokoan

Aplicaciones en uso

- Ayuda a la edición y comprensión de textos
- Traducción automática
- Tratamiento de grandes masas de texto
- Tratamiento de voz
 - Sistemas que leen textos,
 - Sistemas reconocedores del habla,
- Enseñanza de segundas lenguas
- ...



- Information For
- Home
 - Business
 - Enterprise
 - Developers
- Resources
- Events
 - Webinars
 - White Papers
 - Press
 - Investors
 - Careers



ScanSoft, Inc. is the leading supplier of speech and imaging solutions that are used to automate a wide range of manual processes - increasing productivity, reducing costs and improving customer service.

Business Solutions

Save time and

Tratamiento de voz
Diferentes aplicaciones

Scansoft
www.scansoft.com/

- Localized Websites
- Select one

Producing PDF Converter 2 PROFESSIONAL

Instantly Create and Convert PDF Files

Learn More

Productivity Applications



- Dragon NaturallySpeaking
- IBM ViaVoice
- OmniForm
- OmniPage
- PaperPort
- PDF Converter - Create!
- Imaging/OCR

Embedded Speech



SpeechWorks Solutions

- Speech Recognition
- Text-To-Speech
- Speaker Verification
- Automotive
- Games

Network Speech



SpeechWorks Solutions

- Speech Recognition
- Text-To-Speech
- Speaker Verification
- Applications & Tools
- Professional Services
- Training

Language Technology World

Search Products [Help / Search globally](#)

Name Section
Organization All Fields

- Information & Knowledge**
- Information Sources
- Technologies
- Abbreviations
- Players & Teams**
- People
- Projects
- Organizations
- Systems & Resources**
- R & D - Systems
- Products
- Repositories
- Communication & Events**
- News
- Conferences
- About LT World

[You can enter additional p](#)

monotonic, paraconsistent, first-order term logic, with compos
contraction, an acting component, including a well-defined sy
rules that allow for acting in support of reasoning and reason
consisting of an English morphological analyzer and synthesiz
Grammar interpreter/compiler.

Universal Translator Deluxe
Homepage: <http://www.languageforce.com>
Authors: LanguageForce, language force
Organization: LanguageForce
E-mail: sales@wforce.com
Supported Languages: independent
Platforms: Windows 95/98, Windows NT
Section/s: Written Language Translation
Abstract: an omni-directional translation system
Description: Instantly translate web pages, email, documents

Xerox Web@ssistant
Homepage: <http://www.mkms.xerox.com>
Authors: Xerox Multilingual Knowledge Management Solution
Organization: Xerox Multilingual Knowledge Management Sol
E-mail: xts-support@mkms.xerox.com
Supported Languages: English, French, German
Platforms: independent
Section/s: Written Language Translation, Written Language Understanding
Abstract: Xerox Web@ssistant allows you to get a precise understanding of multi-word expressions in their context with only one click.
Description: Xerox Web@ssistant is a comprehension aid module part of Xerox Terminology Suite (XTS) which is a flexible and scalable suite of software that helps you create, validate, translate, update and manage all aspects of a terminology database and make the knowledge available to everyone who needs it throughout the extended enterprise. Xerox Web@ssistant allows you to get a precise understanding of multi-word expressions in their context with only one click. Xerox Web@ssistant has a broad language coverage and its dictionary content is customisable to use your bespoke terminology via Xerox TermOrganizer and its Xerox Dictionary Builder feature.

source languages: english, french, german target languages:all unicode languages

one-month free evaluation licence available on <http://www.mkms.xerox.com>: <http://www.mkms.xerox.com>

Información
en general
www.lt-world.org



- 1 info
- 2
- 3 queries
- 4 submit
- 5 faq
- 6
- 7
- 8
- 9
- 10
- partners

Sections of the Natural Language Software Registry

Sections

- [Annotation Tools \(30\)](#)
- [Evaluation Tools \(8\)](#)
- [Language Resources \(55\)](#)
- [Multimedia \(9\)](#)
- [Multimodality \(11\)](#)
- [NLP Development Aid \(67\)](#)
- [Spoken Language \(42\)](#)

Written Language (171)

- ◆ [Alignment Tools \(8\)](#)
- ◆ [Corpus Analysis \(25\)](#)
- ◆ [Deep Generation \(9\)](#)
- ◆ [Deep Syntactic Analysis \(31\)](#)
- ◆ [Document Image Analysis \(1\)](#)
- ◆ [Dynamic Hyperlinking \(2\)](#)
- ◆ [Grammar & Style Checkers \(14\)](#)

Spell Checkers

entries 6-9 of 9



Orthogرافix

proofing tools for Microsoft Word and Adobe InDesign (for German, Finnish, Danish, Norwegian, Swahili)

TWOL

Two-level morphological analysis tools

WMTrans Products

text processing software for various NLP applications

XUXEN. Spelling Corrector for Basque

spelling corrector integrated in MS-Office. Morphology based. Treatment of typical errors.

**Catalogo general de productos
NLSR:
Directorio de programas para PLN
registry.dfki.de**

Motivación

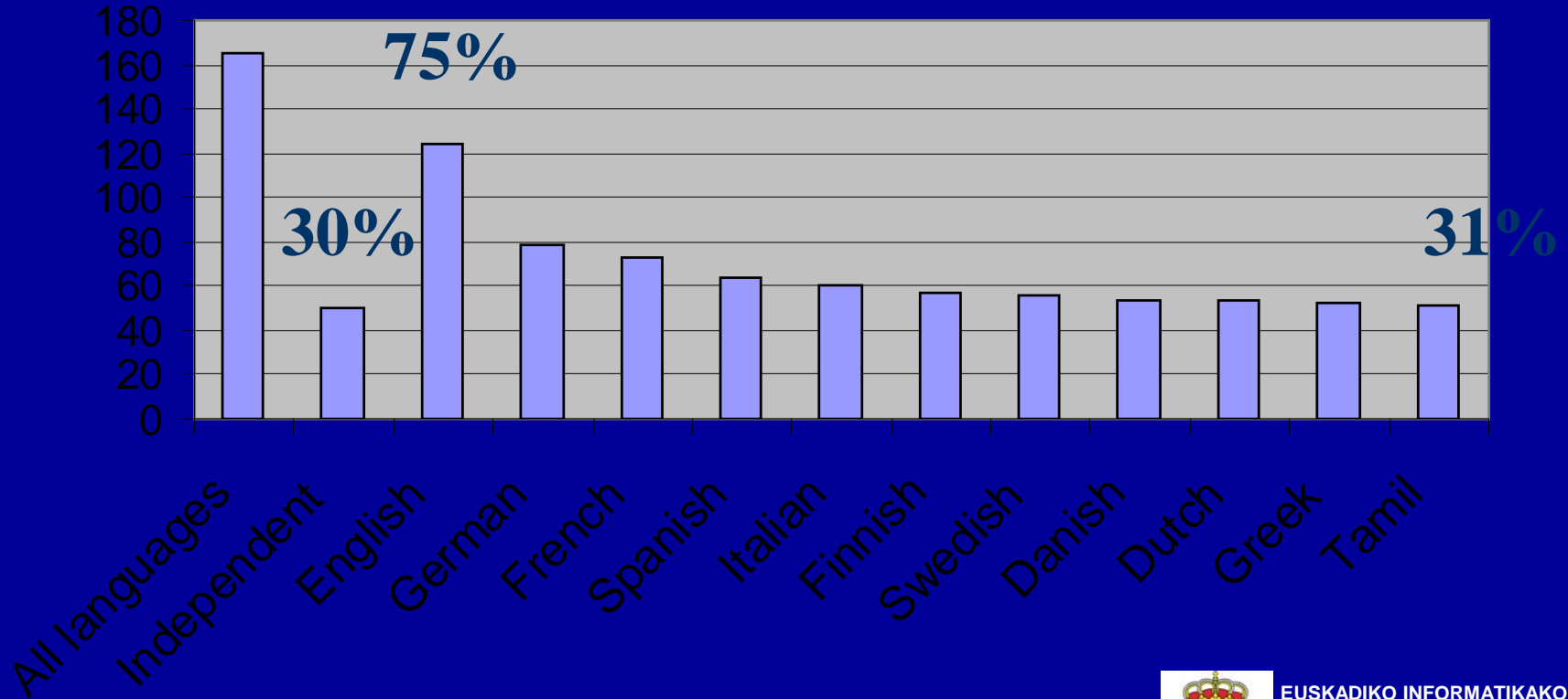
Pero la inmensa mayoría de los sistemas en vigor sólo son aplicables para el inglés

⇒ Las demás lenguas han de realizar un gran esfuerzo para no quedarse descolgadas en esta carrera.



Motivación

NLSR: Directorio de programas para PLN



SOFTKAT catalogo de soft en euskara

www.ueu.org/softkat

(5) OFIMÁTICA

- Procesadores de textos, contabilidad ...

(16) Tiempo Libre

- Musica,juegos...

(28) Aplicaciones relacionadas con la lengua

- Itzultzaileak, zuzentzaileak, hiztegiak...

(8) INTERNET

- Navegadores, correo electrónico...

(12) HERRAMIENTAS GENERALES

- Sistemas operativos, bases de datosy buscadores para ...

(36) ENSEÑANZA Y APLICACIONES DIDACTICAS

- Matemática, ciencias...



Euskararako hizkuntza-aplikazioak

- Ediziorako laguntzak:
 - Xuxen: zuzentzaile ortografikoa
 - Elhuyar hiztegia. Officeko plug-ina.
- Hizketaren tratamendua
 - BIZKAIFON (Bizkaieraren Fonoteka)
 - AhoTTS Testu-Ahots Bihurgailua
- Euskara ikasteko metodoak:
 - BOGA
 - HEZINET
 - Bai & Bye
- Lematizatzailea, informazioa bilatzeko tresna
 - Euslem
- Datu-base dokumentala
 - Kapsula
- Baliabide lexikalak: hiztegiak, esamoldeak, ...
 - 16 produktu



EUROPA. Un mercado abierto: TECNOLOGÍA DE LA LENGUA

- Motivación. Aplicaciones
- Objetivos
- Estrategia
- Promotores y usuarios
- Conclusiones



Objetivos generales (a medio plazo)

Oportunidades ofrecidas por la ingeniería lingüística:

Gestión de contenido

- Buscadores de información multimedia (Partes de tiempo, bolsa, deportes, noticias, petición de películas o imágenes)
- Televisión digital
- Sistemas multimedia

Reconocimiento y síntesis de voz

- Domótica
- Ayudas para discapacitados
- Sistemas de diálogo

Producción de documentación y tecnología de la traducción

Sistema de enseñanza (e-learning)



Objetivos generales (2)

⇒ desarrollo de tecnología exportable y con posibilidad de internacionalización

- El mercado de productos para el inglés es muy grande, pero esos productos tienen una insuficiente difusión para otros idiomas
- Nosotros en Europa, en Euskal Herria vivimos en una sociedad plurilingüe
- El euskara es una lengua muy diferente que permite validar la portabilidad de un producto a otros idiomas



Objetivos generales (3)

⇒ Creación de una comunidad I+D en Ingeniería Lingüística conectada con el entorno internacional a través de alianzas estratégicas con centros de referencia y proyectos

– Personal

- 2002: 120-150 2006: 400-500

– Empresas/Agentes

- 2002: 35 2006: 50

Que comparta :

- Algoritmos y programas
- Metodologías
- Tecnología



EUROPA. Un mercado abierto: TECNOLOGÍA DE LA LENGUA

- Motivación. Aplicaciones
- Objetivos
- Estrategia
- Promotores y usuarios
- Conclusiones

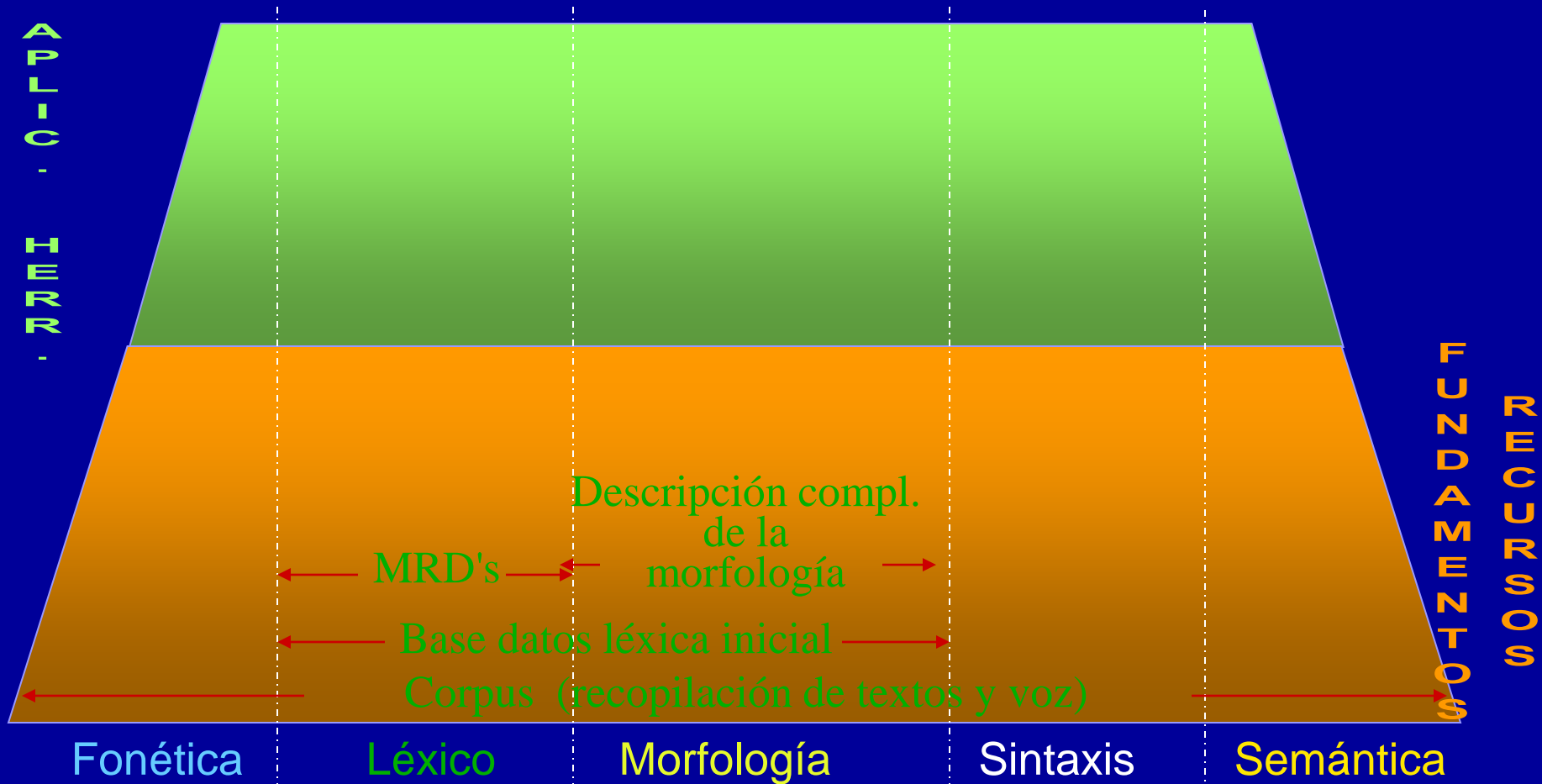


Estrategia:

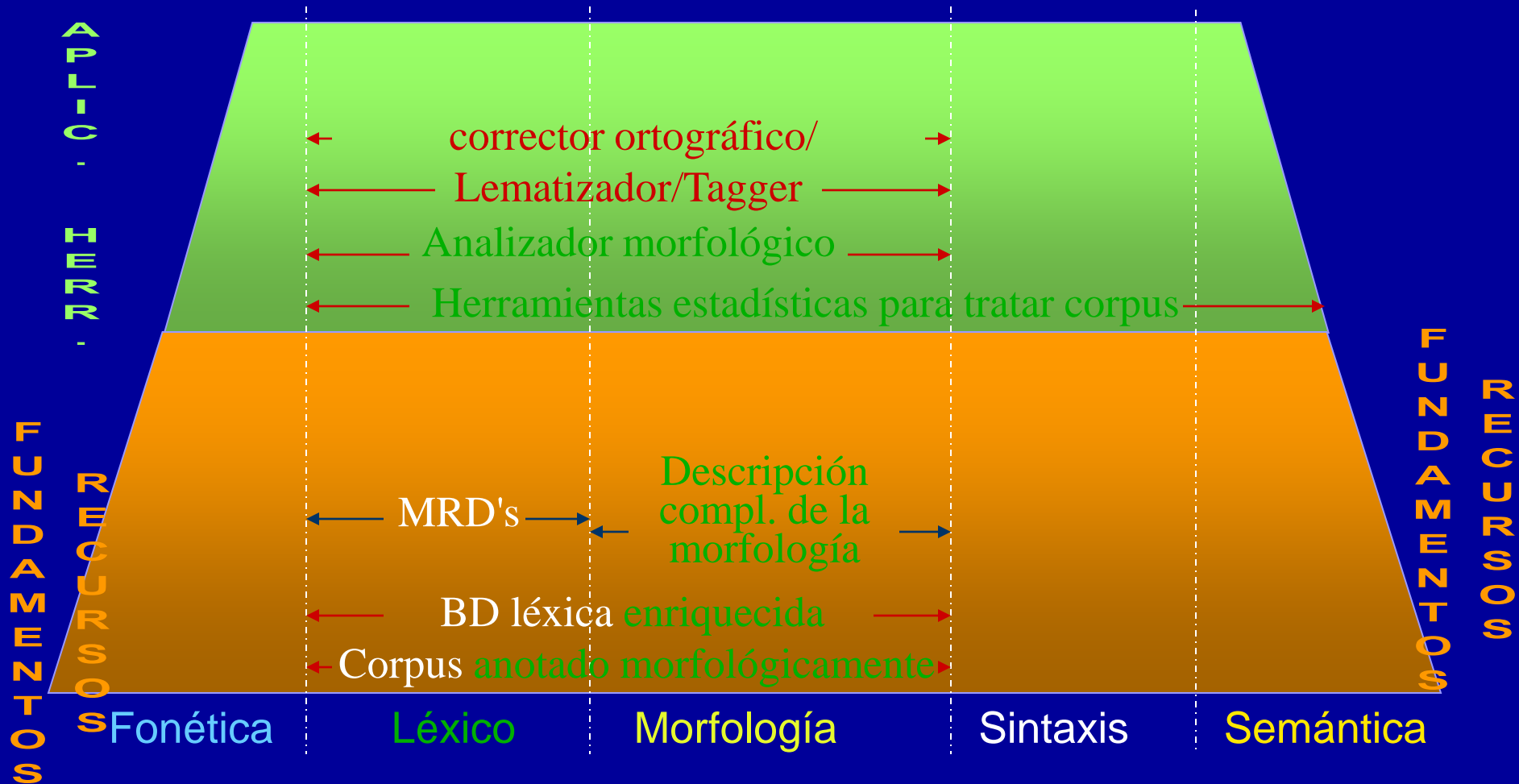
Primero investigación básica y aplicada
y luego desarrollo de aplicaciones



Fase I: Establecimiento de fundamentos lingüísticos (1992)



Fase II: Primeras herramientas básicas y aplicaciones



Fase IV: Multilingüismo y aplicaciones generales

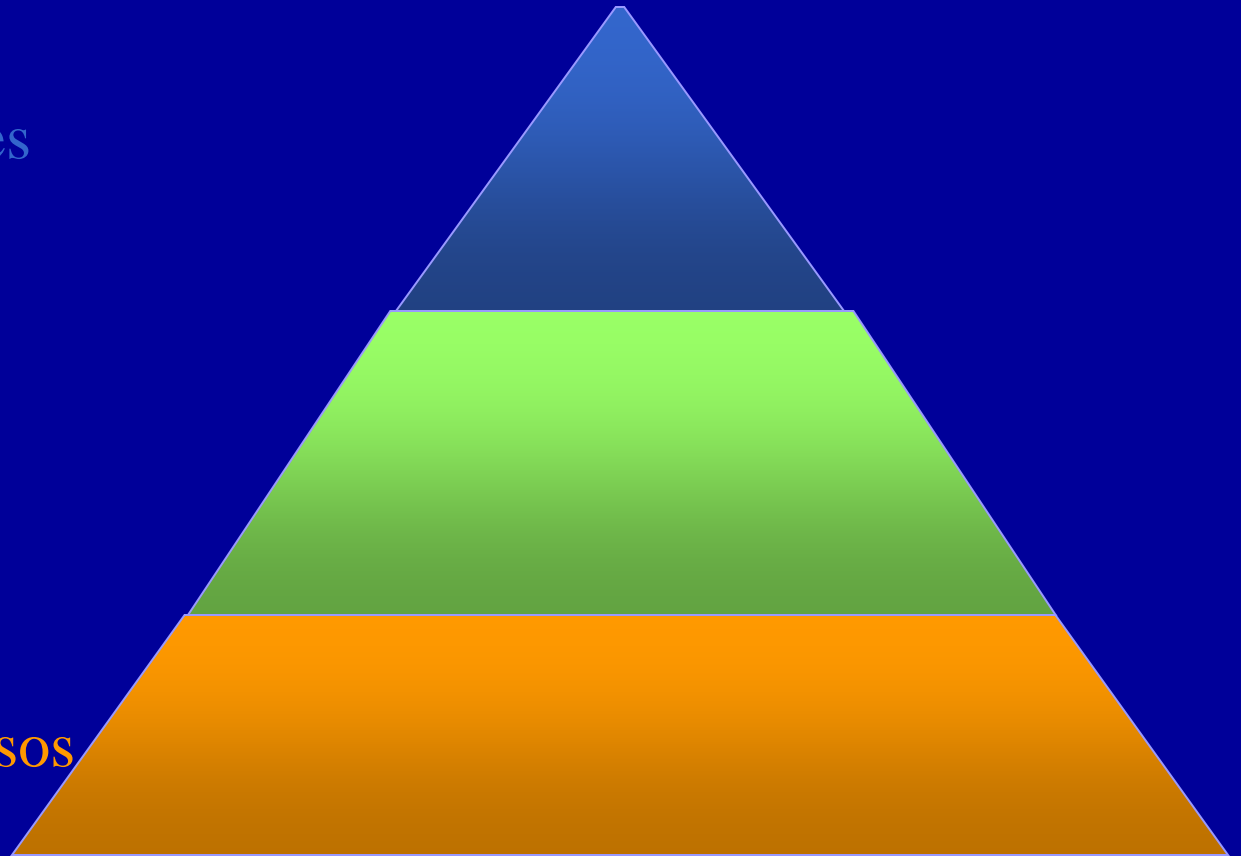


Fase V: Explotación industrial

Productos comerciales

Herramientas

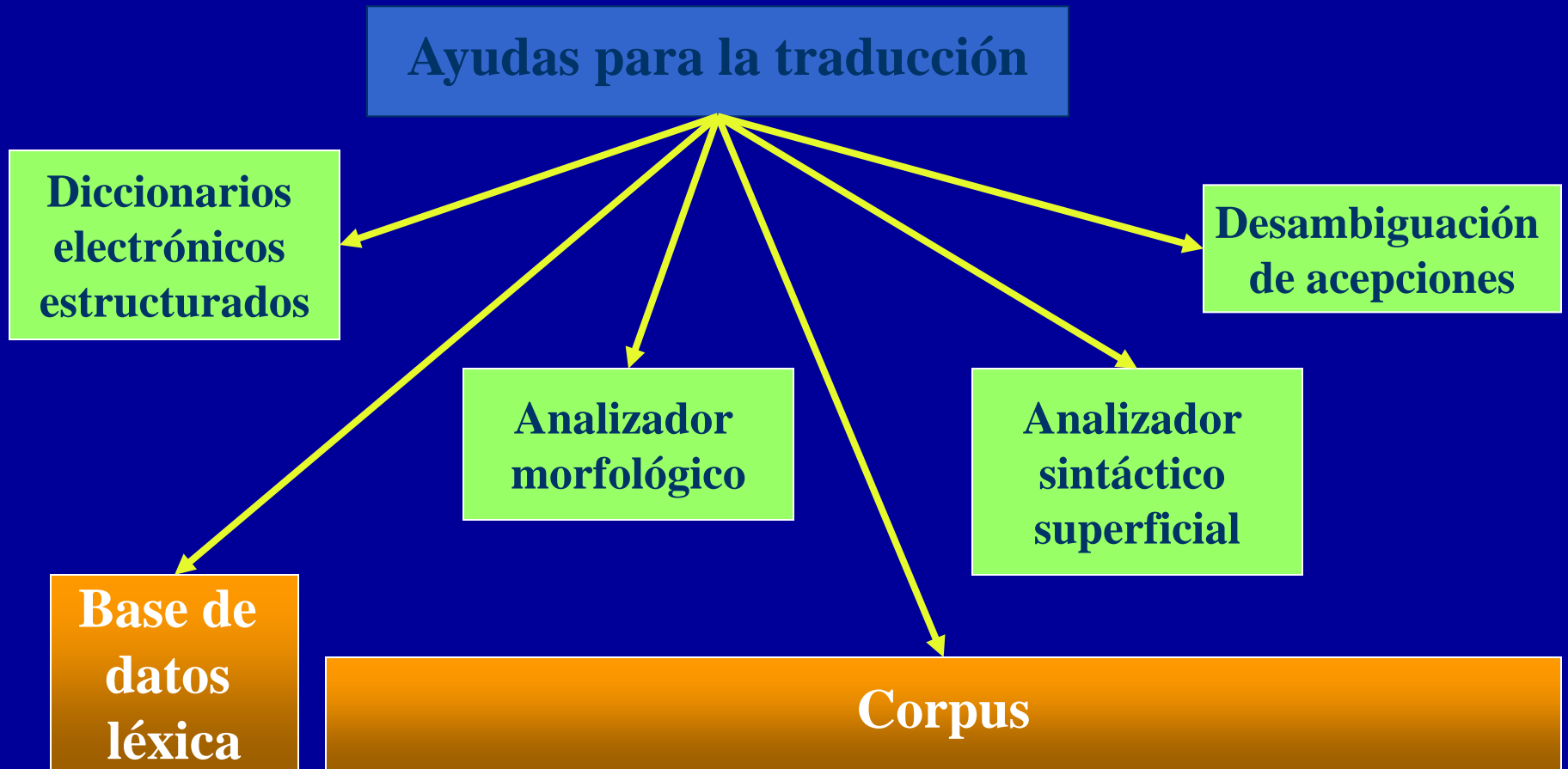
Fundamentos y recursos



EUSKADIKO INFORMATIKAKO
INGENIARIEN ELKARGO OFIZIALA

COLEGIO OFICIAL DE INGENIEROS
EN INFORMÁTICA DEL PAÍS VASCO

Necesidad de herramientas y fundamentos



Propuesta

Objetivos concretos

- I+D+I
 - Creación de recursos lingüísticos
 - Desarrollo de herramientas base para futuras aplicaciones
 - Técnicas para la integración de la tecnología a la vida cotidiana
- Formación
- Infraestructura
- Colaboración internacional
- Difusión
- Observatorio tecnológico



Objetivos concretos. I+D+I

Recursos de corpus escrito

Type of Corpus	Basque (Megawords)	English	Spanish	Basque 2006
Corpus I. Collection of raw text without any tagging mark. (light XML)	5 Mw (EEBS)	(BNC) 100 Mword	200 Mw (CRAE)	100 Mw
Corpus IIa. Word-forms are tagged with their part of speech and lemma.	5 Mw		0	100 Mw
Corpus IIb. Word-forms are tagged with their morphosyntax analysis (hand corrected)	50 Kw		0	2 Mw
Corpus III. Syntactically tagged text.	300w	Penn TreeBank (1Mw)	0	1 Mw
Corpus IV. Semantically tagged text after word-sense disambiguation.	3000 w (Senseval)	Semcor 200 Kw		1 Mw
Corpus Va. Multilingual corpora (light XML)				100 Mw
Corpus Vb. Multilingual and paralelized corpora		CraterMultext 5 Mw		5 Mw
Corpus A. Speech	1Mw	100Mw		100 Mw

Objetivos concretos . I+D+I

Herramientas de desarrollo

- An environment for tool integration.
- Lemmatizer/tagger.
- Syntactic tagger.
- Semantic tagger. (Word-sense disambiguation.)
- Morphological analyzer/generator.
- Syntactic analyzer
- Syntactic generator.
- Speech processing at sentence level.
- Speech processing at word level.
- Lexicographer workbench
- Statistical tools for the treatment of corpus



Objetivos concretos. I+D+I

Integración de la tecnología informática

- Interfaces y multimedia
- Definición de APIs para aplicaciones
- Técnicas de integración de herramientas en web
- Técnicas de indexación
- Lenguajes de marcado estándar
- Métodos estadísticos y aprendizaje automático
- Autómatas de estados finitos
- Formalismos basados en reglas
- Programación orientada a componentes



Objetivos concretos

Formación

- Master en “Tecnologías de la lengua” (UEU y UPV-EHU)
<http://www.ueu.org/titulazioak/Hiztek2003>

2001-02: 300 horas
20 alumnos
2002-02: 500 horas
18 alumnos
2003-04: 500 horas
19 alumnos



- Programa de doctorado en
“Estudio y tratamiento del lenguaje” (UPV-EHU y UEU)
2004-2006



EUSKADIKO INFORMATIKAKO
INGENIARIEN ELKARGO OFIZIALA

COLEGIO OFICIAL DE INGENIEROS
EN INFORMÁTICA DEL PAÍS VASCO

EUROPA. Un mercado abierto: TECNOLOGÍA DE LA LENGUA

- Motivación. Aplicaciones
- Objetivos
- Estrategia
- Promotores y usuarios
- Conclusiones



Participantes potenciales

Informe para la promoción de la Ingeniería lingüística
(Gobierno Vasco, 2000)

- *ASP, Ametzagaiña AIE, Aurten Bai Fundazioa, BAI & BY, ELHUYAR, EHUKo Ahots Taldea, EHUKo Ixa Taldea, EHUKo Zientzien Fakultatea, Euskaltzaindia, Eusko Ikaskuntza, Eusko Jaurlaritzako Kultura Sailaren Hizkuntza Politikarako Sailordetza, Eusko Jaurlaritzako Hezkuntza, Unibertsitate eta Ikerketa Saila, Eusko Jaurlaritzako Industria, Merkataritza eta Turismo Saila, GEINSA, HABE, Ihardun Multimedia, Interlinea 2000, KAIXO, LKS S. Coop., Telefonica, UZEI eta Zabaltzen*
- + *Empresas de nueva creación :*
VICOMTech, Diana, CodeSyntax, Eleka



Promotores

hizkingo21

HIZKUNTZ INGENIARITZA XXI. MENDEAREN ATEAN

- **Aholab**: Escuela de Ingenieros UPV-EHU.
Tratamiento de voz
- **IXA**: Facultad de Informática UPV-EHU.
Tratamiento de textos (morfología, sintaxia, semantika, corpus, itzulpen automatikoa, IE-IR, ...)
- **Vicomtech**: Centro de investigación aplicada
(EiTB + Franhauser)
Imágenes interactivas por ordenador y multimedia digital
- **Elhuyar Fundazioa**: Centro intermedio de investigación.
Lexikografia, terminologia, hiztegiak, hizkuntz planak, zientzia eta teknologiaren zabalkuntza, multimedia-produktu eta zerbitzuak.
- **Robotiker**: Centro tecnológico.
Tecnologías de la información y de las telecomunicaciones.



EUSKADIKO INFORMATIKAKO
INGENIARIEN ELKARGO OFIZIALA

COLEGIO OFICIAL DE INGENIEROS
EN INFORMÁTICA DEL PAÍS VASCO

Usuarios potenciales: *Editoriales*

- *Editorial Desclee de Brouwer S.A., Grupo Delta, Zabaltzen banatzailea, Auñamendi argitaldaria, Editorial Donostiarra, Sendoa, Ostoa S.A., Erein S.A., Lur argitaletxea, Editorial Planeta S.A., Euskal Kulturgintza S.A., Sendoki S.A., Ediciones Saldaña S.A., Aralar liburuak S.A., Alberdania S.L., Donostiako Komunikabideak E.M., Ediciones Txingudi S.L., Euskalgaiak Abarka S.L., Basandere argitaletxea S.L., Udako Euskal Unibertsitatea U.E.U., Miatzen S.A. R.L., Elhuyar Kultur Elkarteak, Harlouxet, Susa, Tarttalo, Elkarlan .*



Usuarios potenciales: *Instituciones*

- *EIZIE, HAEE/IVAP, Diputaciones, Eusko Jaurlaritza, ayuntamientos,*
- *Grupos en euskara: Ikastolen Elkarteak, Goienea, Oarso Komunikabideak Fundazioa, Ttipi Ttapa, Topagunea,, Bertsozaleen Elkarteak.*



Usuarios potenciales: *Medios de comunicación*

- *Berria, Grupo Correo, Gara, Deia, Diario El País, El Mundo, Diario As, Diario Marca, revistas locales*
- *Radio...*
- *Televisión...*
- *...*



Usuarios potenciales: *Bancos y cajas de ahorro*

- *Kutxa, BBK, Vital Kutxa, Euskadiko Kutxa, Caixa...*
- *BBVA, Banco Guipuzcoano, ...*



Conclusiones (1)

- En la actualidad ya hay múltiples servicios lingüísticos que facilitan el trabajo al usuario humano
 - Pero la inmensa mayoría de los sistemas en vigor sólo son aplicables para el inglés
- ⇒ Las demás lenguas han de realizar un gran esfuerzo para no quedarse descolgadas en esta carrera.



Conclusiones (2)

- El mercado de productos para el inglés es muy grande, pero esos productos tienen una insuficiente difusión para otros idiomas
- Nosotros en Europa, en Euskal Herria vivimos en una sociedad plurilingüe
- El euskara es una lengua muy diferente que permite validar la portabilidad de un producto a otros idiomas
- Experiencia de 17 años

⇒ Podemos crear una industria puntera que desarrolle tecnología exportable y con posibilidad de internacionalización



Conclusiones (3)

Metodología

- *Es preciso seguir una **estrategia a medio plazo** para la investigación y desarrollo de la Ingeniería Lingüística*
- *Cada recurso lingüístico, cada herramienta, cada aplicación ha de ser diseñada para que pueda ser **reutilizable** en múltiples productos.*



ElIEO Hitzaldiak – Conferencias COIIE 2004

Informatikako Ingeniariei Erronkak. / Retos de los Ingenieros en Informática.”

Europa. La ingeniería Informática tiene un mercado abierto:

TECNOLOGÍA DE LA LENGUA

Ponente: Kepa Sarasola

Cargo Profesor de la UPV-EHU

Miembro del grupode investigación IXA



EUSKADIKO INFORMATIKAKO
INGENIARIEN ELKARGO OFIZIALA

COLEGIO OFICIAL DE INGENIEROS
EN INFORMÁTICA DEL PAÍS VASCO