# Matxin-Informatika:
## versión del traductor Matxin adaptada al dominio de la informática

Iñaki Alegria, Unai Cabezón, Gorka Labaka, Aingeru Mayor, Kepa Sarasola
IXA Taldea. Euskal Herriko Unibertsitatea
aingeru.mayor@ehu.es

**Objetivo**
Mejorar Matxin, traductor automático es-eu basado en reglas, para el dominio de la informática

**Contexto**
Proyecto OpenMT-2
*http://ixa.si.ehu.es/openmt2*

## Adaptación del léxico al dominio

### • a partir de corpus

**Recopilación de corpus paralelo**
- Del dominio de la informática
- Creado en la localización de Sw

| | |
|---|---|
| Segmentos | 138.000 |
| Palabras es | 600M |
| Palabras eu | 440M |

**Tratamiento del corpus**
- Analizado, lematizado y procesado con Giza++
- Para cada lema (es) se extraen:
  - sus posibles traducciones (eu)
  - y su probabilidad

**Uso**
- Reordenación de equivalencias en **444** entradas del lexicón

*dirección: ~~norabide(rumbo)~~ → helbide(ubicación)*

### • a partir de recursos diccionariales

**Búsqueda en diccionarios**
- Accesibles en la red
- Entradas marcadas del dominio de la Informática o similares

**Uso**
- Inclusión de **1623** entradas nuevas en el lexicón (sobre todo multi-palabra)

*base de datos, lenguaje de programación...*

- Modificación de la primera equivalencia léxica en **184** entradas del lexicón

*rutina: ~~ohitura(hábito)~~ → errutina(procedimiento)*

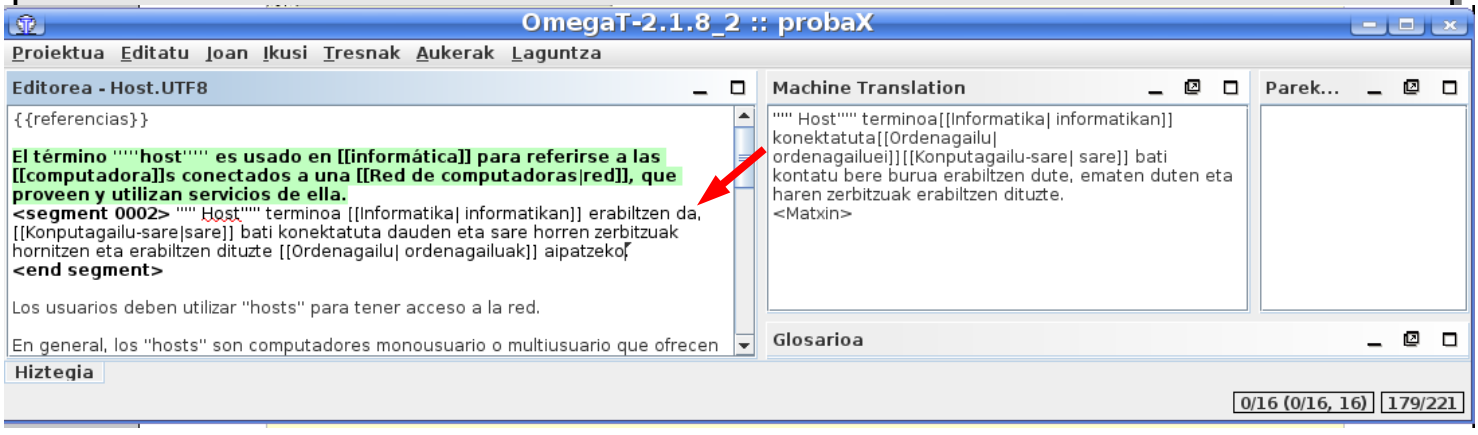
es → matxin informatika → eu

## Desarrollo de un corpus de postedición para su uso en postedición estadística

**Objetivo**
- Creación de un corpus de traducciones es-eu manualmente posteditadas de al menos 100.000 palabras.

**Metodología**
- Traducción posteditada de 50 artículos largos de Wikipedia
- En colaboración con la comunidad eu.wikipedia
- Se usa Matxin-Informatika

**Interfaz**
- Uso de una Interfaz basada en OmegaT para facilitar el trabajo de postedición
  - Se importa un artículo de es.Wikipedia
  - Traducción automática con Matxin-Inf
  - Corrección manual posteditando
  - Se sube la traducción a eu.Wikipedia


OmegaT-2.1.8_2 :: probaX
Proiektua Editatu Joan Ikusi Tresnak Aukerak Laguntza
Editorea - Host.UTF8
{{referencias}}
El término '''''host''''' es usado en [[informática]] para referirse a las [[computadora]]s conectados a una [[Red de computadoras|red]], que proveen y utilizan servicios de ella.
<segment 0002> '''''Host''''' terminoa [[Informatika| informatikan]] erabiltzen da, [[Konputagailu-sare|sare]] bati konektatuta dauden eta sare horren zerbitzuak hornitzen eta erabiltzen dituzte [[Ordenagailu| ordenagailuak]] aipatzeko
<end segment>
Los usuarios deben utilizar "hosts" para tener acceso a la red.
En general, los "hosts" son computadores monousuario o multiusuario que ofrecen
Machine Translation
''''' Host''''' terminoa[[Informatika| informatikan]] konektatuta[[Ordenagailu| ordenagailuei]][[Konputagailu-sare| sare]] bati kontatu bere burua erabiltzen dute, ematen duten eta haren zerbitzuak erabiltzen dituzte.
<Matxin>
Parek...
Glosarioa
Hiztegia
0/16 (0/16, 16)
179/221