

# Apertium

Traducció automàtica lliure per a tothom

Marc Riera Irigoyen  
@marcrierai

9 de maig de 2018  
Sessió Tradumàtica IV

<https://apertium-tradumatica2018.netlify.com>

## **La traducció automàtica**

Definició

Tipus

## **Què és Apertium?**

Característiques

Objectius

Contribuir-hi

Casos d'èxit

## **Funcionament d'Apertium**

Introducció

Mòduls principals

Altres mòduls

## **Dades lingüístiques**

Pràctica

## **Raons per a fer servir Apertium**

## La traducció automàtica

Definició

Tipus

## Què és Apertium?

Característiques

Objectius

Contribuir-hi

Casos d'èxit

## Funcionament d'Apertium

Introducció

Mòduls principals

Altres mòduls

## Dades lingüístiques

Pràctica

## Raons per a fer servir Apertium

# Què és la traducció automàtica?

## Què és la traducció automàtica?

*La traducció automàtica (TA) és el procés de traducció, mitjançant un sistema informàtic (compost per ordinadors i programes), de textos informatitzats escrits en la llengua origen a textos informatitzats escrits en la llengua meta. (Ginestí i Forcada, 2009)*

# Tipus de traducció automàtica

Generalment podem parlar de dos tipus de sistemes:

- ▶ Traducció automàtica basada en regles
- ▶ Traducció automàtica basada en corpus
  - Traducció automàtica estadística
  - Traducció automàtica neuronal

També hi ha tecnologies híbrides.

## La traducció automàtica

Definició

Tipus

## Què és Apertium?

Característiques

Objectius

Contribuir-hi

Casos d'èxit

## Funcionament d'Apertium

Introducció

Mòduls principals

Altres mòduls

## Dades lingüístiques

Pràctica

## Raons per a fer servir Apertium

# Característiques d'Apertium

- ▶ Plataforma de traducció automàtica lliure
- ▶ Funcionament basat en regles
- ▶ Transferència superficial
- ▶ Llengües generalment properes



# Objectius d'Apertium

- ▶ Oferir traducció automàtica per a **qualsevol mena** de llengües
  - Llengües minoritàries i minoritzades
  - Llengües de poc interès comercial
  - Llengües amb pocs recursos lingüístics
- ▶ Ser una alternativa lliure a la traducció automàtica comercial
- ▶ Implicar a tota mena de gent en la traducció automàtica
- ▶ Fomentar l'ús de la traducció automàtica
  - Assimilació
  - Disseminació

# Qui contribueix a Apertium?

- ▶ Projecte gratuït i lliure: no és una empresa
- ▶ Voluntaris de tota mena
  - Informàtics
  - Lingüistes
- ▶ Finançament indirecte
  - Premis
  - Beques de Google per a estudiants (GSoC i GCI)

# Casos d'èxit

- ▶ Wikipedia Content Translation
- ▶ Prompsit
- ▶ Levante-EMV
- ▶ La Voz de Galicia
- ▶ Salt



WIKIPEDIA  
The Free Encyclopedia

# Trädgårdsföreningen

español

[ver página](#)

La Sociedad de Horticultura de Gotemburgo, en sueco Trädgårdsföreningen es un jardín botánico situado en una zona céntrica de la ciudad sueca de Gotemburgo, no lejos de la Estación Central y del parque Brunnsparken.

El jardín propiamente dicho posee una extensión de pocas hectáreas, y está contiguo y adosado al **Jardín Botánico de Götemburgo** y sin solución de continuidad a la gran zona verde del parque Slottsskogen, que entonces sumaría una extensión de 175 **ha** (unos 430 acres), estando incluido el **Arboretum** y la reserva natural.



Puerta de entrada al Trädgårdsföreningen de Gotemburgo.

## Trädgårdsföreningen

català

La Societat d'Horticultura de Gotemburgo, en suec Trädgårdsföreningen és un jardí botànic situat en una zona central de la ciutat sueca de Gotemburg, no lluny de l'Estació Central i del parc Brunnsparken.

El jardí pròpiament dit posseeix una extensió de poques hectàrees, i està contigu i adossat al *Jardí Botànic de Gotemburgo* i sense solució de continuïtat a la gran zona verda del parc *Slottsskogen*, que llavors sumaria una extensió de 175 ha, estant inclòs el *Arboretum* i la reserva natural.

Publicar traducción

 Buscar una palabra

Traducción automática

De Apertium, del \$2 al \$3

→ Restaurant

☐ Usar texto de origen

✕ Borrar traducción

## La traducció automàtica

Definició

Tipus

## Què és Apertium?

Característiques

Objectius

Contribuir-hi

Casos d'èxit

## Funcionament d'Apertium

Introducció

Mòduls principals

Altres mòduls

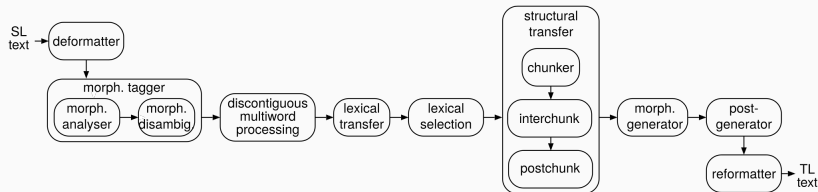
## Dades lingüístiques

Pràctica

## Raons per a fer servir Apertium

# Funcionament general

Apertium és una cadena de mòduls independents, cadascun amb una funció específica.



# Funcionament general

- ▶ Els mòduls són universals, no depenen de cap idioma
  - Dades lingüístiques accessibles per a tothom
  - No cal modificar els mòduls per crear un parell nou
- ▶ Els mòduls es comuniquen mitjançant text pla
- ▶ Cada parell pot fer servir més o menys mòduls

# Mòduls principals

- ▶ Analitzador
- ▶ Desambiguador
- ▶ Mòdul de transferència lèxica
- ▶ Mòdul de transferència estructural
- ▶ Generador



# Analitzador

L'analitzador cerca les paraules del text d'origen en un diccionari monolingüe i n'extreu:

- ▶ La forma interna
- ▶ La informació morfològica

the → the<det><def><sp>

blue → blue<adj><sint>

house → house<n><sg>

# Desambiguador

El desambiguador fa servir un model estadístic per triar la millor opció quan una paraula té més d'una anàlisi.

```
^the/the<det><def><sp>$  
^blue/blue<adj><sint>/blue<n><sg>$  
^house/house<n><sg>/house<vblex><inf>$  
  
^the<det><def><sp>$  
^blue<adj><sint>$  
^house<n><sg>$
```

## Mòdul de transferència lèxica

El mòdul de transferència lèxica cerca els equivalents en la llengua d'arribada de les paraules del text original mitjançant un diccionari bilingüe.

`^the<det><def><sp>$`

`^blue<adj><sint>$`

`^house<n><sg>$`

`^el<det><def><GD><ND>$`

`^blau<adj><GD><ND>$`

`^casa<n><f><sg>$`

# Mòdul de transferència estructural

El mòdul de transferència estructural aplica canvis d'ordre o modifica les etiquetes de les paraules a partir de regles.

`^el<det><def><GD><ND>$`

`^blau<adj><GD><ND>$`

`^casa<n><f><sg>$`

`^el<det><def><f><sg>$`

`^casa<n><f><sg>$`

`^blau<adj><f><sg>$`

# Generador

El generador cerca les paraules de la llengua d'arribada en un diccionari monolingüe i genera la forma final. Fa la feina de l'analitzador a l'inrevés.

el<det><def><f><sg> → la

casa<n><f><sg> → casa

blau<adj><f><sg> → blava

## Altres mòduls

- ▶ Gramàtica de restriccions (*Constraint Grammar*)
- ▶ Mòdul de paraules separables
- ▶ Mòdul de selecció lèxica

## La traducció automàtica

Definició

Tipus

## Què és Apertium?

Característiques

Objectius

Contribuir-hi

Casos d'èxit

## Funcionament d'Apertium

Introducció

Mòduls principals

Altres mòduls

## Dades lingüístiques

Pràctica

## Raons per a fer servir Apertium

# Dades lingüístiques

Les dades lingüístiques fan que Apertium conegui la morfologia i els contrastos de les llengües. El format més habitual és l'XML i els fitxers principals són els següents:

- ▶ Diccionaris monolingües
- ▶ Diccionaris bilingües
- ▶ Regles de transferència estructural



# Pràctica

Pràctica per a posar en pràctica els coneixements i aprendre a editar els diccionaris d'Apertium:

<https://apertium-tradumatica2018.netlify.com/practice>

## La traducció automàtica

Definició

Tipus

## Què és Apertium?

Característiques

Objectius

Contribuir-hi

Casos d'èxit

## Funcionament d'Apertium

Introducció

Mòduls principals

Altres mòduls

## Dades lingüístiques

Pràctica

## Raons per a fer servir Apertium

# Raons per a fer servir Apertium

- ▶ És programari lliure
- ▶ Es pot integrar fàcilment en eines TAO
- ▶ Amb alguns parells funciona molt bé
- ▶ Sabem com funciona
  - Podem predir les traduccions resultants
  - Podem solucionar nosaltres mateixos els problemes
- ▶ Respecta la privacitat i la confidencialitat

I sobretot, perquè té una comunitat fantàstica de col·laboradors!

## Una mica d'humor per acabar...

Carlota Corredera, en la cuerda floja.

*Carlota Corredissa, en la corda fluixa.*

Los abrigos estrella de este invierno.

*Els abrics estavella d'aquest hivern.*

¿Dibujó Da Vinci a la Mona Lisa desnuda?

*Va dibuixar Da Vinci a la Mico Llisa nua?*

El rei Lluís XIV de França.

*El rey Luciera XIV de Francia.*

Pilar de la Foradada, municipi del Baix Segura.

*Pilar del Socavón, municipio del Baix Segura.*

Gràcies!