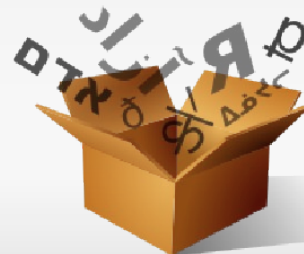


# La plataforma lliure de traducció automàtica Apertium

I Jornada de difusió de programari informàtic en valencià

Universitat d'Alacant  
21 de maig de 2010

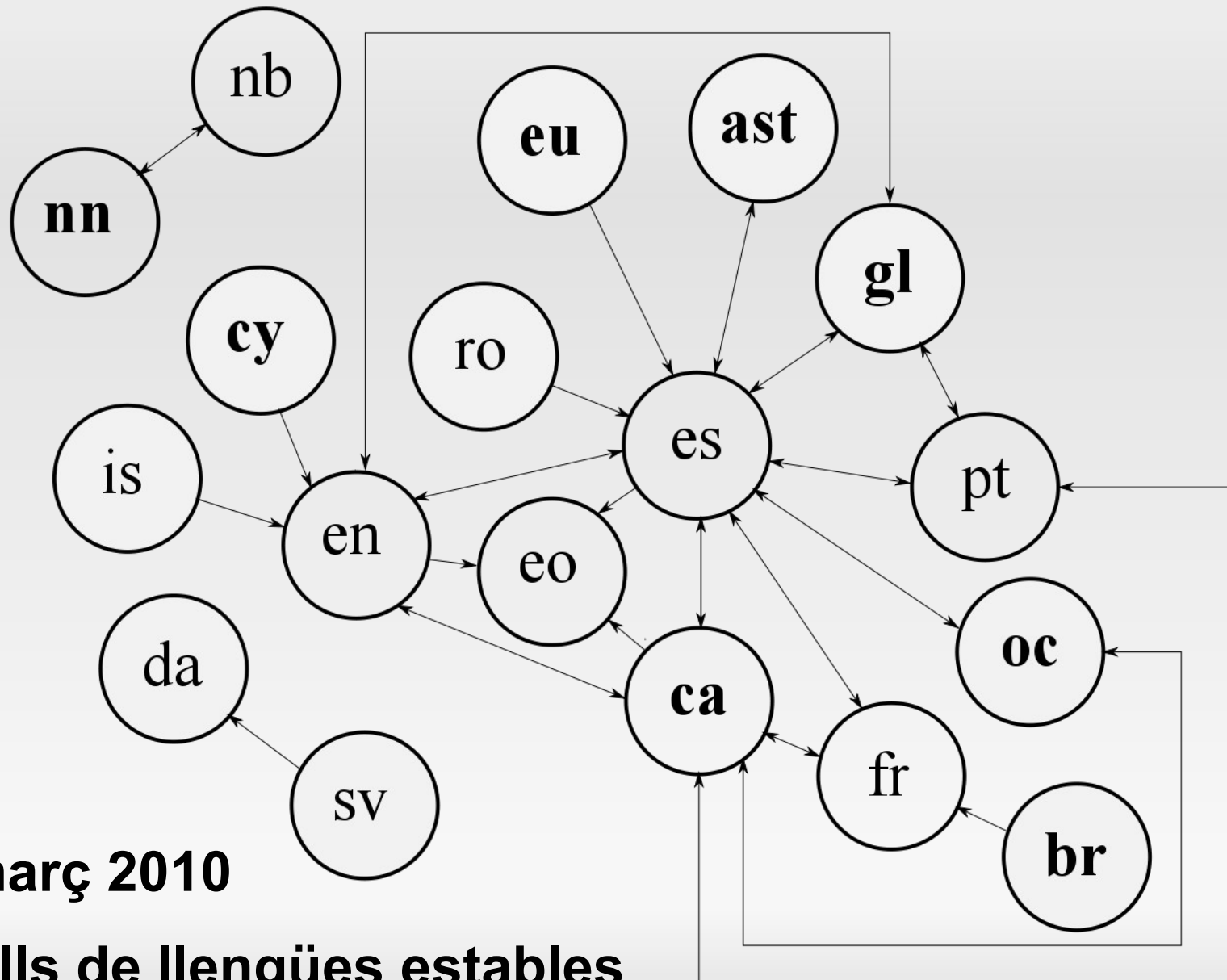
Gema Ramírez-Sánchez  
[gramirez@prompsit.com](mailto:gramirez@prompsit.com)  
[www.prompsit.com](http://www.prompsit.com)



# Què és Apertium?

- Apertium és una plataforma lliure/de codi obert per a la traducció automàtica
- Proporciona:
  - un motor de traducció independent de les llengües
  - eines per a crear i fer funcionar traductors
  - dades lingüístiques per a un nombre creixent de parells de llengües
- Motor i dades GPL = comú de coneixements  
= comú de recursos

# Parells de llengües d'Apertium



**Data: març 2010**

**22 parells de llengües estables**

# Apertium i el valencià

- Una de les llengües millor representades:

PARELL	VERSIÓ   DATA
es←→ca (ca@valencia)	1.2.0   15/10/2009
en←→ca	0.8.9   15/01/2010
eo←→ca	0.9.1   04/07/2009
fr←→ca	1.0.2   12/03/2009
oc(oc@aran)←→ca	1.0.5   21/07/2008
pt←→ca	0.8.1   04/03/2009

# Orígens d'Apertium

- Antecedents
  - Internostrum: es $\longleftrightarrow$ ca(ca@valencia)
  - Traductor Universia: es-pt@brasil
  - Anys: 1998-[2002-2004]-2006
  - Grup Transducens (Departament de Llenguatges i Sistemes Informàtics de la UA)
  - Tecnologies anàlogues: sistemes basats en regles, arquitectura modular, tècniques de estats finits, etc.

# Orígens d'Apertium

- InterNOSTRUM (iN)
  - Finaçat: per CAM-UA - Ministeri Ciència Tecnologia
  - Disponible:
    - CAM
    - Gestor de la UA
    - Traductor de Softcatalà
    - [www.internostum.com](http://www.internostum.com)
  - Model de llengua: variant central.
  - Adaptat al model dels "Criteris lingüístics" de les universitats valencianes (gràcies al Servei de Promoció del Valencià de la UA)

# Orígens d'Apertium: (iN)

Fitxer Edita Visualitza Historial Adreces d'interès Eines Ajuda


← → ↻ × 🏠 <http://www.interostrum.com/tradquadtext.php> ☆ Google 🔍

🌟 📁 Més visitades ▾ 🔄 Prompsit 🔄 Apertium 📄 SourceForge.net R... 🇪🇸 RAE 🇪🇸 Apertium alpha-te... 🗣️ Tek Translation :Lo... 🇪🇸 TLF

**Menú principal**

- Versió estàndard
- Versió grans volums
- Sol·liciteu accés complet
- Introduïu mots nous
- Informació tècnica
- Contacteu amb nosaltres
- Enllaceu-nos
- Preguntes més freqüents
- Analitzadors morfològics

[Versió del 10/07/2006]



**Sistema de traducció automàtica castellà-català**

Text traduït - Versió estàndard

➔ **Reviseu el text traduït**  
Comproveu també l'ortografia del text original

Hola buenos días, bienvenidos a la I Jornada de difusión de software informático en valenciano de la UA.

➔ **Continueu provant la versió actual d'interNOSTRUM**  
Introduïu o enganxeu un text ANSI (fins a 16384 caràcters) en el quadre de text, seleccioneu el sentit de la traducció, el tipus de sortida que voleu generar i premeu "Tradueix".

**Sentit**

**Novetat!**

Dóna prioritat a les **formes valencianes**

**Tipus de sortida**




Hola bon dia, benvinguts a la I Jornada de difusió de programari informàtic en valencià de la UA.

**Traduir...**

- ➔ **Text**
- Documents i websites
- Traduir i navegar
- XAT!


**PRINCIPAL**

Finançat per



TIC-2000-1599-C02-02

Desenvolupat pel



Departament de Llenguatges i Sistemes Informàtics

# Apertium: www.apertium.org

Fitxer Edita Visualitza Historial Adreces d'interès Eines Ajuda

← → ↻ × 🏠  ☆ Google 🔍

🌟 📁 Més visites ▾ 🔄 Prompsit 🐦 Apertium 📦 SourceForge.net R... 🇪🇸 RAE 🇪🇸 Apertium alpha-te... 🗣️ Tek Translation :Lo... 🌐 TLF >>

05-03-2010: Icelandic--English 0.1.0 released, [read more...](#)

## Apertium

Plataforma lliure/de codi font obert per a la traducció automàtica

[proveu-lo](#) | [què es Apertium?](#) | [descàrregues](#) | [documentació](#) | [contacte](#) 🇪🇸 català (ca) ▾

### Traducció de textos

Apertium ofereix tant traducció de textos, com a [traducció de documents](#) i [navegar i traduir](#). També es pot [buscar als diccionaris](#) en línia i provar les [versions en desenvolupament](#) dels nostres traductors.

Direction:  ▾

Marqueu les paraules desconegudes

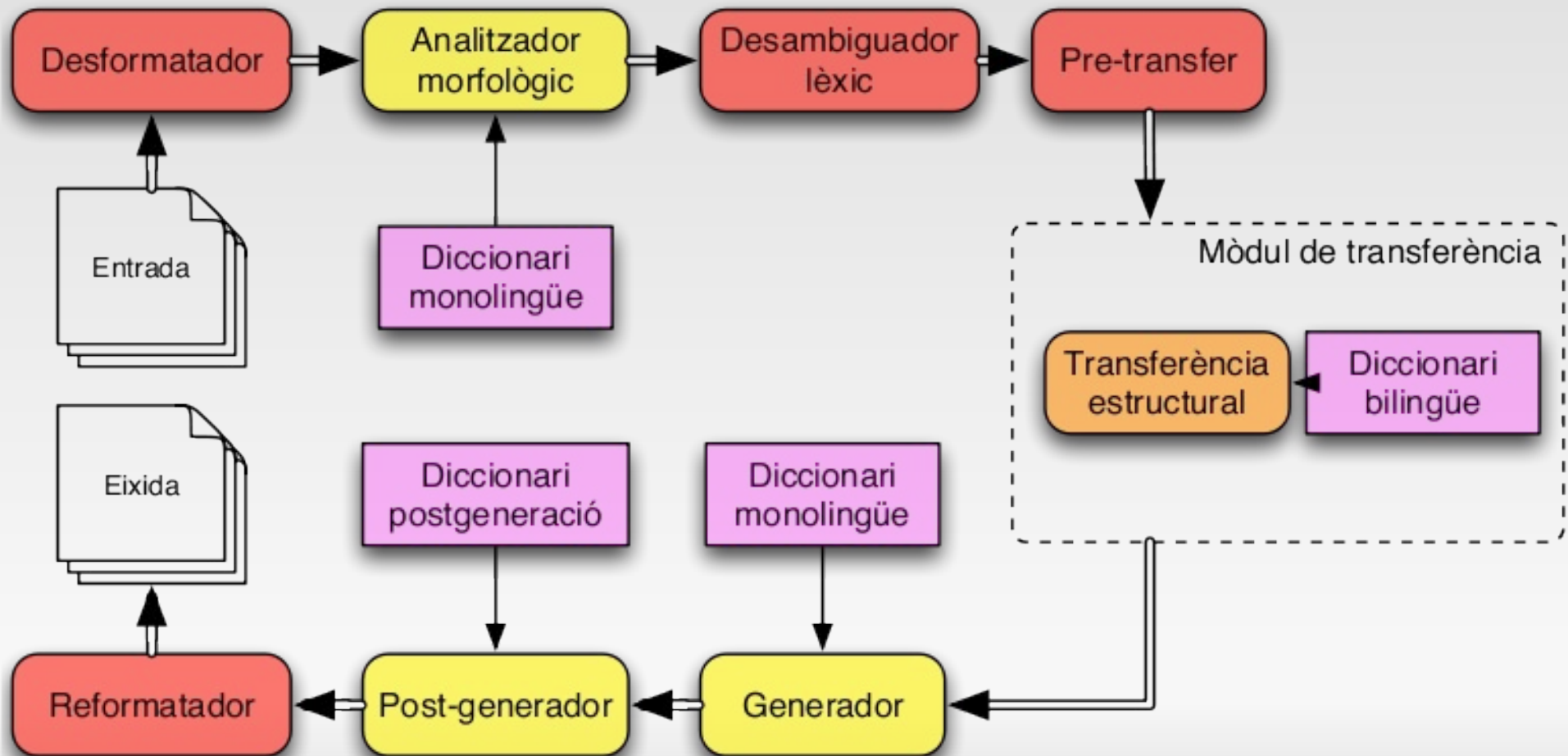
<http://www.apertium.org/?id=surfandtranslate> Disabled

# Traducció automàtica: bàsics

- Traducció automàtica = LO → sistema de TA → LM
- Usos:
  - assimilació (comprendre)
  - disseminació (difondre)
- Tipus de sistemes
  - **basats en regles: Apertium**
  - basats en corpus:
    - estadístics
    - basats en exemples

# Com funciona Apertium?

- Sistema de transferència sintàctica parcial:



# Mòduls: analitzador

- Exemple: "Tuvo que mover la señal"
- Eixida de l'analitzador:

^Tuvo que/Tener<vbmod><ifi><p3><sg># que\$

^mover/mover<vblex><inf>\$

^la/el<det><def><f><sg>/lo<prn><pro><p3><f><sg>\$

^señal/señal<n><f><sg>\$

# Mòduls: desambiguador

- Exemple: "Tuvo que mover la señal"
- Eixida del desambiguador:

^Tener<vbmod><ifi><p3><sg># que\$

^mover<vblex><inf>\$

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

^señal<n><f><sg>\$

# Mòduls: transferència

- Exemple: "Tuvo que mover la señal"
- Eixida del mòdul de transferència:

^**Anar**<vaux><p3><sg>\$

^**haver**<vbmod><inf># **de**\$

^**moure**<vblex><inf>\$

^**el**<det><def><m><sg>\$

^**senyal**<n><m><sg>\$

# Mòduls: generació

- Exemple: "Tuvo que mover la señal"
- Eixida del generador:

**Va haver ~de moure ~el senyal**

# Milliores fetes a Apertium

- Separació completa de codi i dades
- Definició d'estàndards per a la codificació de dades (XML)
- Millora de la velocitat de compilació i execució
- Millora de la gestió de clítics i unitats multiparaula
- Ampliació de vocabulari i regles de transferència
- Ampliació de funcionalitats
- TOT ES GPL!!!

# Parell apertium-es-ca@valencia -I

- Analitza (sentit ca@valencia→es) i genera (sentit es→ ca@valencia)
- Model de llengua: universitats valencianes.
- Lèxic: 38.000 lemes (15.677 noms + 5.622 adj. + 4.096 verbs + 4.503 adverbis + 8.006 noms propis, etc.)
- Regles de transferència: 100 regles per sentit
- Adaptació finançada pel Servei de Promoció del Valencià de la UA i desenvolupada per Prompsit

# Parell apertium-es-ca@valencia -II

- Continuament en desenvolupament (UA, UOC, Prompsit, desenvolupadors particulars)
- Darrera versió estable: octubre 2009
  - disponible al web d'Apertium: [www.apertium.org](http://www.apertium.org)
- Versió més actualitzada: març 2010
  - disponible al gestor de la UA
  - disponible al web de Prompsit: [www.prompsit.com](http://www.prompsit.com)
- Apostar per Apertium = benefici present i futur

# Recursos per al valencià

- Apertium proporciona:
  - diccionaris monolingües i bilingües en XML
  - analitzadors morfològics
  - desambiguadors categorials
  - lematitzadors
  - recull de regles contrastives
  - traductors ràpids i configurables per a molts tipus de formats: txt, rtf, odt, doc, odp, ppt, odx, xls, xml, html, etc.

# Altres sistemes per a es-ca:

- automatictrans (At):
  - <http://www.automatictrans.es/>
- Lucy software (Lt):
  - [http://www.lucysoftware.de/cms/front\\_content.php?idcat=45&changelang=4](http://www.lucysoftware.de/cms/front_content.php?idcat=45&changelang=4)
- Google translator (Gt):
  - <http://translate.google.es/>

# Diferències Apertium - iN/At/Lt/Gt

- 1a diferència:

**ús lliure**

**vs.**

**codi obert/lliure**

- **iN, At, Lt, Gt** : només l'ús en línia es lliure (i restringit)
- **Apertium**:
- llibertat descàrrega, ús i execució
- llibertat modificació
- llibertat distribució
- llibertat de millora i publicació

# Diferències Apertium - iN/At/Lt/Gt

- 2a diferència:

**varietat de formats**

**vs.**

**pocs formats**

- **iN, Apertium, Gt:** text, documents i webs (Apertium i Gt molta varietat de formats)
- **Lt:** text (curt) i webs (curtes)
- **At:** text (50 paraules)

# Diferències Apertium - iN/At/Lt/Gt

- 3a diferència:

**rapidíssim**

**vs.**

**rapid**

**vs.**

**lent**

- 1. Google translator**
- 2. Apertium**
- 3. interNOSTRUM**
- 4. Lucy Translator**
- 5. Automatictrans**

**Atenció al maquinari!!**

# Diferències Apertium - iN/At/Lt/Gt

- 4a diferència:

**ca i ca@valencia**

**vs.**

**ca (una variant o variants barrejades)**

- **iN i Apertium:** possibilitat d'admetre i generar + d'una variant
- **At, Lt:** només genera una variant
- **Gt:** admet i genera + d'una variant, però totes alhora!!!

# Diferències Apertium - iN/At/Lt/Gt

- 5a diferència:

**coneixement  
lingüístic**

**vs.**

**models estadístics**

**Atenció: original!!**

- **iN, Apertium, At, Lt:** traduccions sistematitzades = errors sistemàtics. L'original és opcional.
- **Gt:** traduccions fluïdes = traduccions que poden enganyar. L'original és fonamental.

# Un exemple. Google vs. Apertium


Fitxer Edita Visualitza Historial Adreces d'interès Eines Ajuda

← → ↻ × 🏠 [http://translate.google.com/translate?js=y&prev=\\_t&hl=ca&ie=UTF-8&layout=2&eof...](http://translate.google.com/translate?js=y&prev=_t&hl=ca&ie=UTF-8&layout=2&eof...) enciclopedia catalana

Més visitades Prompsit Apertium SourceForge.net R... RAE Apertium alpha-te... Tek Translation :Lo... TLF

## Google traductor

[Ajuda](#) | [Entra](#)

**Text castellà original:** 

[Una novela y nueve vídeos para pasar miedo](#)

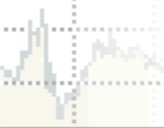





[Suggeriu una traducció millor](#)

**Una novel-nou vídeos per passar por**  
Arriba a Espanya 'Skeleton Creek', una història interactiva de terror de l'autor nord-americà Patrick Carman

seus serveis professionala

### Un tribunal dóna vaga de British Ai

L'aerolínia havia aconsegu els 20 dies d'aturades pe

	Ibex 35	-3.20% 	Eurostoxx 50	-3.91% 	Cac 40	-4.26% 	Do
	Eurotop 100	-3.61% 	S & p 500	-3.32% 	Nikkei 225	-1.54% 	Na

[http://translate.googleusercontent.com/translate\\_c?hl=ca&ie=UTF-8&sl=es&tl=ca&u=http://www.elpais.com/articulo/cultura/novela/videos/pasar/mie...](http://translate.googleusercontent.com/translate_c?hl=ca&ie=UTF-8&sl=es&tl=ca&u=http://www.elpais.com/articulo/cultura/novela/videos/pasar/mie...) Disabled

# Un exemple. Google vs. Apertium

Fitxer Edita Visualitza Historial Adreces d'interès Eines Ajuda

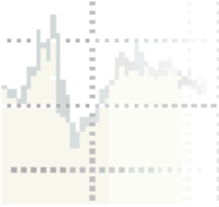
← → ↻ × 🏠

🌟 📁 Més visitades ▾ 🎯 Prompsit 🐦 Apertium 📦 SourceForge.net R... 🏠 RAE 🏠 Apertium alpha-te... 🗣️ Tek Translation :Lo... 🌐 TLF

## Una novel·la i nou vídeos per passar por

Arriba a Espanya 'Skeleton Creek', una història interactiva de terror de l'autor nord-americà Patrick Carman

Una novela y nueve vídeos para pasar miedo

	Ibex 35	-2.84% 📉	Eurostoxx 50	-3.19% 📉	Cac 40
	Eurotop 100	-3.08% 📉	S&p 500	-2.72% 📉	Nikkei 225

[> VER COTIZACIONES](#)

publicidad

idealista.com

http://www.prompsit.com/tradurl.php?marcar=0&direccion=es-ca&inurl=http://www.elpais.com/articulo/cultura/novela/videos/pasar/miedo/elpepucul/... Disabled

# Qui desenvolupa Apertium?

- Apertium és una comunitat activa:
  - 105 desenvolupadors inscrits (particulars, empreses, universitats)
  - el codi s'actualitza molt freqüentment
  - wiki mantingut col·lectivament
  - canal IRC per a desenvolupadors
  - llista de correu i gestor d'errades
  - paquets estables disponibles en Debian GNU/Linux (i per tant, en Ubuntu)

# Qui finança Apertium?

- Finançament públic i privat (exemples):
  - Ministeri d'Indústria, Turisme i Comerç
  - Secretaria de Comunicacions i Societat de la Informació de la Generalitat de Catalunya
  - Universitat d'Alacant, Universitat Oberta de Catalunya
  - Ministeri d'Assumptes Exteriors de Romania, Ofis ar Brezhoneg
  - Empreses: Prompsit Language Engineering, ABC Enciklopedioj, imaxin|software, Eleka Ingeniaritza Linguistikoa, Eolaistriu Technologies, etc.

# Quines aplicacions té Apertium?

- Traductors:
  - caixa de text, documents, webs
- Altres aplicacions:
  - la interfície gràfica d'ús (*apertium-tolk*) i l'eina de diagnòstic (*apertium-view*)
  - plugins per a OpenOffice.org, Firefox, per al missatger Pidgin, per a Wordpress, etc.
  - diccionaris bilingües per a mòbils i PDA (*tinylex*)
  - aplicació per a subtítols (*apertium-subtitles*)
  - serveis web (*ScaleMT*, *apertium-service*)



# Enllaços importants:

- Apertium: <http://www.apertium.org>
- Wiki: <http://wiki.apertium.org>
- Descàrregues:  
<http://sourceforge.net/projects/apertium/>
- Canal irc #apertium a Freenode
- Llista de correu: [apertium-stuff@lists.sourceforge.net](mailto:apertium-stuff@lists.sourceforge.net)
- Apertium al Google Summer of Code 2010:  
[http://wiki.apertium.org/wiki/Google\\_Summer\\_of\\_Code](http://wiki.apertium.org/wiki/Google_Summer_of_Code)

# La plataforma lliure de traducció automàtica Apertium

## Gràcies i proveu Apertium!

I Jornada de difusió de programari informàtic en valencià.

Universitat d'Alacant  
21 de maig de 2010

Gema Ramírez-Sánchez  
[gramirez@prompsit.com](mailto:gramirez@prompsit.com)  
[www.prompsit.com](http://www.prompsit.com)

(aquesta xarrada es GPL, demaneu-me còpia!)

