

La plataforma de traducció automàtica de codi obert Apertium: aplicació al català

Mikel L. Forcada, Gema Ramírez-Sánchez

Grup Transducens, Departament de Llenguatges i Sistemes Informàtics. Universitat d'Alacant

Prompsit Language Engineering, S.L.

mlf@dlsi.ua.es & gramirez@prompsit.com

1. Què és Apertium?

- Apertium és una plataforma de traducció automàtica lliure (o de codi obert). Proporciona:
 - un motor de traducció independent de les llengües
 - eines per a gestionar les dades lingüístiques necessàries per a construcció de sistemes de traducció automàtica
 - dades lingüístiques per a un nombre creixent de parells de llengües
- El programari i les dades d'Apertium usen la *General Public License* de GNU. Són, per tant:
 - lliures
 - amb *copyleft* (es crea un *comú* de coneixements i recursos)

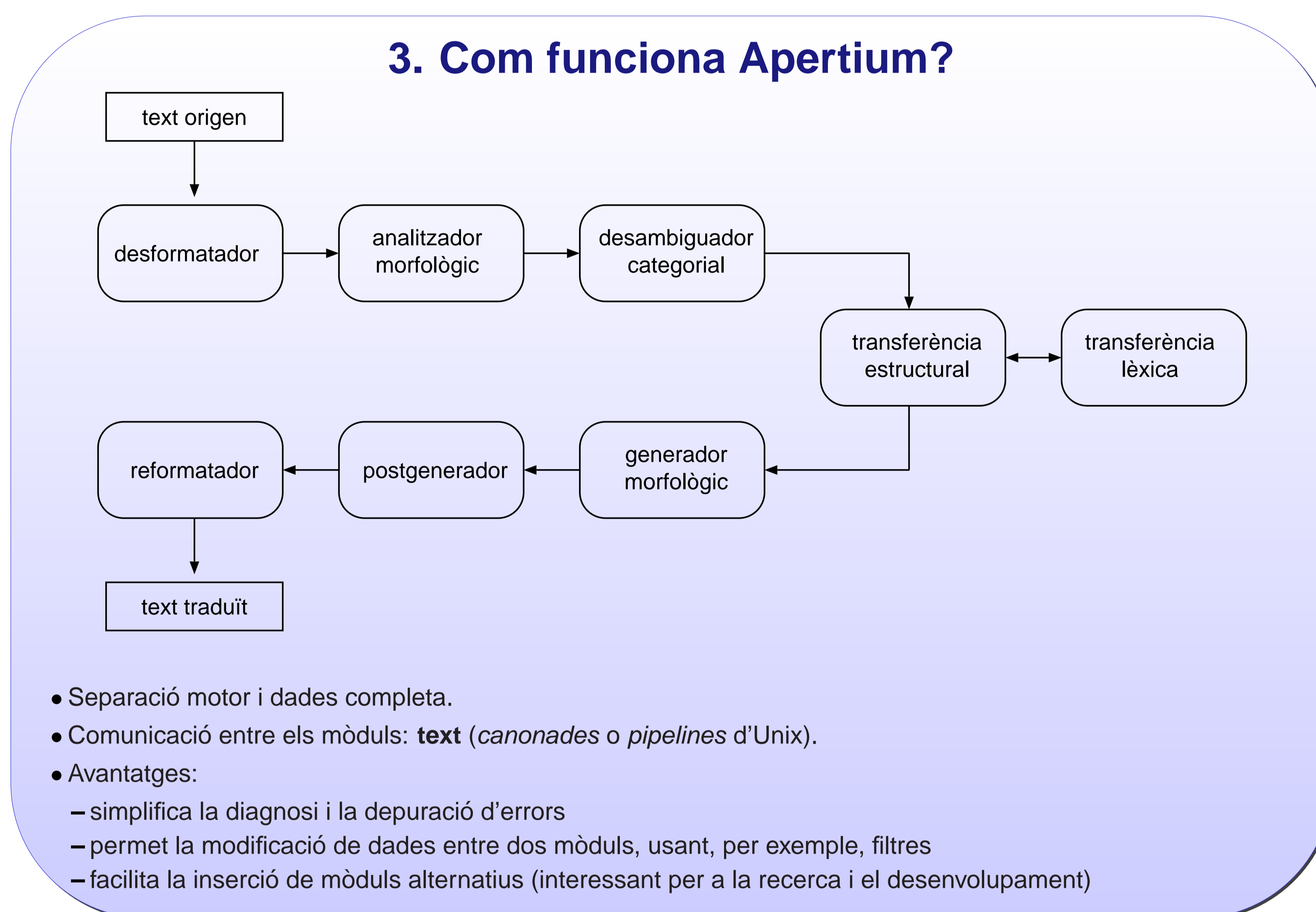
2. Apertium, compromés amb el català

- Molts parells de llengües d'Apertium són amb el català:

Parell	Versió	Data	Parell	Versió	Data
ca→eo	0.9	20/02/2008	en→ca	0.6	19/03/2008
es→ca	1.1.0	01/02/2009	fr→ca	1.0	05/10/2007
oc→ca	1.0.5	21/07/2008	pt→ca	0.8.1	04/07/2009

- Això proporciona: diccionaris morfològics, bilingües i de postgeneració, regles de transferència estructural, desambiguadors categorials, lematitzadors, etc.
- Les dades lingüístiques estan en formats basats en XML:
 - facilita la interoperabilitat, la transformació en d'altres formats i el manteniment

3. Com funciona Apertium?



4. Apertium és una comunitat activa

A més dels desenvolupadors originals (finançats), s'ha format una comunitat al voltant del projecte:

- Hi ha **90 desenvolupadors** inscrits en sourceforge.net/projects/apertium/, molts de fora del grup original; el codi s'actualitza molt freqüentment (centenars d'actualitzacions cada mes).
- Un **wiki** mantingut col·lectivament documenta els components d'Apertium, mostra l'estat actual de desenvolupament i dona consells per als desenvolupadors de dades lingüístiques o de programes.
- Molts desenvolupadors es troben en el **canal IRC #apertium** (de irc.freenode.net).
- Hi ha també una **llista de correu** apertium-stuff@lists.sf.net
- Els paquets estables estan disponibles en **Debian GNU/Linux** (i per tant, en **Ubuntu**).

6. Qui ho ha pagat, això?

- Ministeri d'Indústria, Turisme i Comerç (també: Ministeri d'Educació i Ciència i Ministeri de Ciència i Tecnologia) d'Espanya
- Secretaria de Comunicacions i Societat de la Informació de la Generalitat de Catalunya
- El Ministeri d'Assumptes Exteriors de Romania
- La Universitat d'Alacant
- Empreses: Prompsit Language Engineering, ABC Enciklopedioj, imaxin|software, Eleka Ingeniaritza Linguistikoa, Eolaistrii Technologies, etc.

5. S'hi generen eines noves constantment

Exemples d'eines i codi desenvolupat per la comunitat d'Apertium:

- la interfície gràfica d'ús `apertium-tolk`, i l'eina de diagnòstic `apertium-view`
- *plugins* per a OpenOffice.org, per al missatger Pidgin (abans Gaim), per al gestor de continguts Wordpress, etc.
- Una versió dels diccionaris bilingües per a mòbils i PDA (`tinylex`)
- Una aplicació per a subtítols (`apertium-subtitles`)
- Versions preliminars per a Windows

7. M'hi puc apuntar?

És clar que sí! Busquem constantment:

- Programadors
- Lingüistes i traductors
- Avaluadors

I ara, particularment:

- **Google** ha seleccionat Apertium per al seu *Google Summer of Code 2009*: busquem estudiants programadors.

