# Traducció automàtica de codi obert: Apertium, una oportunitat per a llengües menors

Mikel L. Forcada[1,2]

[1]Departament de Llenguatges i Sistemes Informàtics, Universitat d'Alacant, E-03071 Alacant

[2]Prompsit Language Engineering, S.L., E-03690 St. Vicent del Raspeig

Càtedra LinguaMÓN-UOC – Barcelona, 14 de juliol de 2008

# Índex

# Índex

# Llengües menors i parells de llengües menors/1

Què és una llengua *menor*? S'usen moltes denominacions alternatives (en "ordre de Google"):

- *minority* languages (minoritàries)
- *lesser-used* languages (menys usades)
- *minor* languages (menors)
- *small* or *smaller* languages (petites o més petites)
- *lesser* languages (menors)
- *under-resourced*, *resource-poor* or *less-resourced* languages (amb pocs recursos)
- etc.

# Llengües menors i parells de llengües menors/2

Què és una llengua menor?

- Té un nombre reduït de parlants [alfabetitzats].
- És lluny de la normalitat (s'usa més a casa que a l'escola o a l'administració, està socialment discriminada, reprimida políticament, etc.).
- No té un sistema estable d'escriptura, una ortografia fixa, o una variant estàndard.
- Té una presència limitada a Internet.
- Hi manquen lingüistes experts.
- Disposa de pocs recursos llegibles per ordinador: diccionaris, corpus, etc.

# Llengües menors i parells de llengües menors/3

Els efectes de les tecnologies de traducció sobre una llengua menor succeeixen a través de parells de llengües.
Per exemple:

- les llengües menors *A* i *B* són llengües relacionades (és fàcil construir programes per a traduir entre elles)
- *C* és una llengua *important*.
- hi ha programes de traducció de *C* a *A*

Com a resultat, serà més fàcil tenir programes de traducció de *C* a *B*

# Programari lliure o de codi obert

El programari de codi obert s'anomena també *programari lliure* per aquestes quatre llibertats:

0. "La llibertat d'usar el programa per a qualsevol propòsit."
1. "La llibertat d'estudiar com funciona el programa, i adaptar-lo a les teues necessitats."
2. "La llibertat de distribuir còpies i ajudar així el teu veí"
3. "La llibertat de millorar el programa i de fer públiques les millores als altres, de manera que tota la comunitat se'n beneficie."

Perquè les condicions 1 i 3 es complisquen, s'ha de tenir accés al codi font (tal com l'ha escrit el programador). Per això se'n diu *programari de codi obert*.

# Programari de traducció automàtica/1

- La traducció automàtica (TA) és especial: depén fortament de l'existència de dades. Hi ha tres components en qualsevol sistema de TA:[1]
  - El *motor* (el programa pròpiament dit)
  - Les *dades lingüístiques* (diccionaris, regles)
  - Les *eines* necessàries per a mantenir aquestes dades i convertir-les al format usat pel *motor*

[1] TA "basada en regles"; la TA "basada en corpus" té requisits anàlegs

# Programari de TA/2 : TA comercial

- Els sistemes comercials usen tecnologies *privatives* o *de propietat* (*proprietary*) que no es revelen (hom les percep com un avantatge competitiu fonamental)
- Només s'hi permet la modificació parcial (*personalització*) de les dades lingüístiques

# Programari de TA/3: TA de codi obert

Perquè la TA siga de codi obert, tant

- el motor,
- les dades,
- com les ferramentes

han de ser de codi obert.

# Programari de TA/4: TA que no és ni comercial ni de codi obert

Però hi ha més possibilitats:

- Sistemes que es poden usar lliurement per Internet (alguns ni tan sols es comercialitzen).
- El motor i les eines poden ser programes de *codi tancat* ben documentats, i ser usats amb dades lingüístiques obertes.

# Índex

# Efectes de la disponibilitat de TA sobre les llengües menors

La disponibilitat de TA per a una llengua menor pot

- incrementar-ne la “normalitat”
- millorar-ne els nivells d’alfabetització
- tenir un efecte en l’estandardització
- augmentar-ne la“visibilitat”

# Incrementar la “normalitat”

La TA pot contribuir a la normalitat d’una llengua menor:

- traducció de *materials educatius* d’una llengua important a una menor
- traducció de *notícies* d’una llengua important per a crear mitjans de comunicació en la llengua minoritària
- les *lleis*, normes i informacions governamentals es podrien traduir a la llengua menor més fàcilment
- les empreses ho tindrien més fàcil per a traure al mercat *nous productes* en la llengua menor (“localització”)

[Ací s’assumeix que la *postedició* de la TA en brut és factible]

# Millorar els nivells d’alfabetització

- La disponibilitat de text en la llengua menor (obtingut a través de traducció automàtica i elaboració posterior) pot motivar l’alfabetització en la llengua minoritària

# Efectes sobre l’estandardització

- L’existència d’un sistema de TA d’èxit pot promoure
  - un sistema particular d’escriptura (p.e. alfabet romà sense diacrítics per a l’amazic)
  - una ortografia determinada (*kreyòl asisyen* [= crioll haitià])
  - un dialecte concret com a estàndard (variant aranesa de l’occità?)

  si es genera tecnologia lingüística per a ells.

# Augmentar la “visibilitat”

- La disponibilitat de TA des de la llengua menor a llengües importants pot ajudar a la difusió de material escrit originalment en la llengua menor:
  - per exemple, TA de llocs web (“al vol” o seguida de postedició)

# Índex

# Sistemes de TA comercials i llengües menors: oportunitats limitades

- Les companyies de TA solen tenir com a objectiu les llengües més importants del món (existeixen excepcions, com el català, però. . . és realment el català una llengua menor?)
- És molt difícil adaptar sistemes comercials tancats a llengües menors

# Índex

# Oportunitats de la TA de codi obert

- L'ús de sistemes de TA de codi obert proporciona oportunitats *addicionals*, a més dels efectes positius *genèrics* que acabe d'esmentar:
  - Augmenta la perícia (*expertise*) i els recursos lingüístics
  - Augmenta la independència

# Augment de la perícia i dels recursos lingüístics

- La construcció de sistemes de TA de codi obert per a una llengua menor comporta el creixement de la perícia i dels recursos lingüístics per a la llengua menor, a través de
  - la *reflexió* sobre la llengua menor
  - la *elicitació* (explicitació) del coneixement lingüístic (unilingüe o bilingüe) sobre la llengua menor
  - la *codificació* subsegüent d'aquest coneixement
- L'escenari de codi obert posa de manera natural el coneixement i els recursos a la disposició de la comunitat.

# Cas 1: Creació des de zero de dades per a un motor de TA existent

- És un escenari molt desfavorable. Necessitem:
  - Un *motor* de TA de lliure disposició (obert o no).
  - *Eines* de lliure disposició (obertes o no) per a gestionar les dades lingüístiques
  - *Documentació completa* sobre com construir dades lingüístiques per usar-les amb el motor i les eines
- S'han de prendre moltes decisions lingüístiques. La *síndrome del full en blanc* pot paralitzar el projecte.
- Si se supera, la perícia adquirida i les dades obertes resultants poden ser millorades o usades per a altres fins: efecte positiu en la llengua menor.

# Cas 2: Creació de dades per a un motor de TA existent a partir de dades lingüístiques existents

- Si es disposa de dades obertes per a altre parell de llengües similar o emparentat, la *síndrome del full en blanc* es redueix dramàticament.
- Es podria, per exemple:
  - usar el mateix conjunt de categories lèxiques i indicadors de flexió
  - construir regles de flexió basant-se en les ja existents.

# Cas 3: Adaptació d'un motor i eines de TA de codi obert per a un parell de llengües nou

- Si el motor i les eines són oberts, hom els pot modificar o adaptar per a abordar característiques no previstes del nou parell de llengües:
  - jocs de caràcters (sistema d'escriptura),
  - necessitat d'una anàlisi més profunda, etc.
- Més difícil que crear dades noves
- Però els programadors no necessiten tenir un control total de la llengua menor (és possible una gestió més abstracta dels aspectes lingüístics)

La rescriptura del codi aportaria nous coneixements i recursos a la comunitat de la llengua menor.

# Augment de la independència

- Disposar d'un motor, d'eines i de dades lingüístiques obertes fa que els usuaris d'una llengua menor siguen menys dependents d'un únic proveïdor comercial de programari tancat
- Açò té un efecte anàleg, no sols sobre la TA, sinó també sobre altres tecnologies lingüístiques.

# Índex

# Reptes

La creació d'un sistema de traducció s'enfronta, entre d'altres, als següents reptes:

- L'estandardització de la llengua menor
- La neutralització d'actituds *tecnofòbiques*
- L'organització d'una comunitat de desenvolupadors
- L'elicitació del coneixement lingüístic
- El manteniment de la simplicitat del coneixement lingüístic necessari.
- L'estandardització i documentació dels formats de les dades lingüístiques
- La modularitat de programes i dades.

# Estandardització de la llengua menor

La traducció automàtica pot accelerar l'estandardització d'una llengua menor, però açò té el seu costat negatiu:

- la manca d'un sistema d'escriptura, d'una ortografia o d'un dialecte de referència estàndards és un seriós repte per als desenvolupadors ("síndrome del pioner").

# Neutralització de les actituds tecnofòbiques

- Per a tenir èxit cal conjugar l'*activisme* en pro de la llengua menor amb un nivell adequat de *formació* en tecnologies de la informació.
- S'hi oposen les actituds *tecnofòbiques*: els erudits de la llengua solen desconfiar de les tecnologies per causa de
  - un visió idealitzada de la llengua i la comunicació
  - poca estima pels usos informals o no literaris
  - donar massa èmfasi a *joies* (estructures o paraules especials) poc probables i resistents a l'automatització en lloc de als *maons* (estructures i paraules quotidianes) molt probables i automatitzables.

Aquestes adversitats "socioacadèmiques" es donen (jo mateix les he patides).

# Organització d'una comunitat de desenvolupadors/1

[Assumim que només estem desenvolupant dades lingüístiques]

- El codi obert fa possible que la comunitat d'una llengua menor desenvolupe de manera col·laborativa sistemes de TA per a ella.
- Moltes llengües allunyades de la normalitat tenen grups d'activistes amb habilitats lingüístiques i de traducció.
- Però el temps oferit voluntàriament i aquestes habilitats són necessàries però no suficients.

# Organització d'una comunitat de desenvolupadors/2

Cal una certa organització. Idealment:

- Un *equip coordinador* que domine motor i eines, amb:
  - *coordinadors de programació* (instal·lació, manteniment, modificacions al codi del programa)
  - *coordinadors lingüístics* (creació i modificació de dades lingüístiques)
- Un *servidor web* per al projecte
  - per a distribuir l'última versió del sistema
  - on es puga usar en línia
  - a través del qual els voluntaris puguen contribuir dades lingüístiques
- Un grup de voluntaris ben formats, certificats per l'equip coordinador.

# Elicitació del coneixement lingüístic

- El coneixement lingüístic existent s'ha de fer explícit (*elicitació*) per a poder-ho aportar al sistema.
- L'elicitació del coneixement *lèxic* és possible a través d'interfícies (formularis) web ben dissenyades que permeten
  - proporcionar els lemes de les paraules origen i meta
  - seleccionar el paradigma de flexió de les paraules origen i meta
  - establir l'àmbit d'una equivalència lèxica (bidireccional, d'esquerra a dreta o de dreta a esquerra).
- L'elicitació d'altres tipus de coneixement (p. ex., regles de transferència estructural) és més difícil (i és objecte d'intensa investigació).

# Simplicitat del coneixement lingüístic necessari

El nivell de coneixements lingüístics necessaris per a començar a construir un sistema de TA hauria de ser el mínim possible (p. ex., conceptes i habilitats gramaticals bàsiques de batxillerat).

- Açò és bastant fàcil en sistemes de *transfèrencia superficial* com els que s'usen entre llengües emparentades.
- Però és molt difícil (si no impossible) en sistemes de *transferència profunda* (sintàctica o semàntica).

Una documentació ben escrita pot ser molt útil.

# Estandardització i documentació dels formats de dades lingüístiques

- Una documentació adequada del format de les dades lingüístiques és crucial.
- La solució és usar XML. Per què?
  - En XML cada element de les dades està *explícitament* etiquetat amb una marca que té un nom descriptiu amb un significat clar.
  - L'estructura de les dades pot ser validada automàticament amb DTDs (definicions de tipus de document) o similars (esquemes).
  - Existeixen moltes tecnologies per a XML (que converteixen des de XML i a XML: *interoperabilitat*).

# Modularitat

- Un dels avantatges del codi obert és la possibilitat de reutilitzar el codi i les dades lingüístiques per a crear nous sistemes de TA o noves aplicacions de tecnologia de la llengua.
- Per a això, és necessària la *modularitat*.
- Un motor modular *indueix* modularitat en les dades que usa.
- Per exemple, tenir un analitzador morfològic independent i un diccionari morfològic independent
  - Facilita la creació d'un sistema de TA per a altra llengua meta
  - Es pot usar per a crear un cercador intel·ligent (que cerca paraules independentment de la flexió).

# Índex

# Rerefons

Apertium està basat en les tecnologies creades pel grup Transducens de la Universitat d'Alacant durant el desenvolupament de dos sistemes existents:

- **interNOSTRUM** (`interNOSTRUM.com`, `es`↔`ca`)
- **Tradutor Universia** (`tradutor.universia.net`, `es`↔`pt`)

Aquestes tecnologies, inicialment dissenyades per a parells de llengües relacionades, han estat esteses per a tractar parells de llengües que no estiguen tan relacionades.

# Fonaments /1

Per a generar traduccions que siguen:

- raonablement intel·ligibles i
- fàcils de corregir

entre llengües relacionades com l'espanyol (`es`) i el català (`ca`) o el portugués (`pt`), etc., només cal millorar la traducció *mot per mot* amb:

- processament lèxic robust (incloent-hi unitats lèxiques multi-mot)
- desambiguació lèxica categorial (*part-of-speech tagging*)
- processament estructural local basat en regles simples i ben formulades per a transformacions estructurals freqüents (reordenació, concordança)

# Fonaments /2

Per a parells de llengües més difícils, no tan relacionats:

- Hauria de ser possible estendre aquest model senzill.
- Hauria de ser possible generalitzar-ne els conceptes de manera que la complexitat es mantinga tan baixa com siga possible.

# Fonaments /3

- Hauria de ser possible generar un sistema complet de traducció automàtica a partir de dades lingüístiques (diccionaris monolingües i bilingües, regles gramaticals), especificades de manera declarativa.
- Aquesta informació hauria d'estar en un format interoperable $\Rightarrow$ XML. Aquests són els tipus necessaris de dades:
  - regles (independents de la llengua) per a tractar formats de text
  - especificació del desambiguador lèxic categorial
  - diccionaris morfològics i bilingües i diccionaris de regles de transformació ortogràfica
  - regles de transferència estructural

# Fonaments /4

- Hauria de ser possible tenir un motor de traducció únic (independent de la llengua) que llegiria dades específiques per a cada parell de llengües ("separació d'algorismes i dades").
- Les dades lingüístiques del parell de llengües haurien de ser preprocessades de manera que el sistema siga ràpid ($>$10,000 mots per segon) i compacte; per exemple, les transformacions lèxiques es farien amb transductors d'estats finits (TEFs).

# Fonaments /5

Raons per al desenvolupament d'Apertium en *codi obert*:

- Donar a tothom accés lliure i il·limitat a les millors tecnologies possibles de traducció automàtica.
- Establir una plataforma modular, documentada i oberta per a la traducció automàtica de transferència superficial i per a altres tasques de processament automàtic de la llengua.
- Afavorir l'intercanvi i la reutilització de les dades lingüístiques existents.
- Facilitar la integració amb altres tecnologies de codi obert.

# Fonaments /6

Més raons per al desenvolupament d'Apertium en *codi obert*

- Beneficiar-se del desenvolupament col·laboratiu
  - del motor de traducció i de les eines
  - de dades per a parells de llengües existents o nous

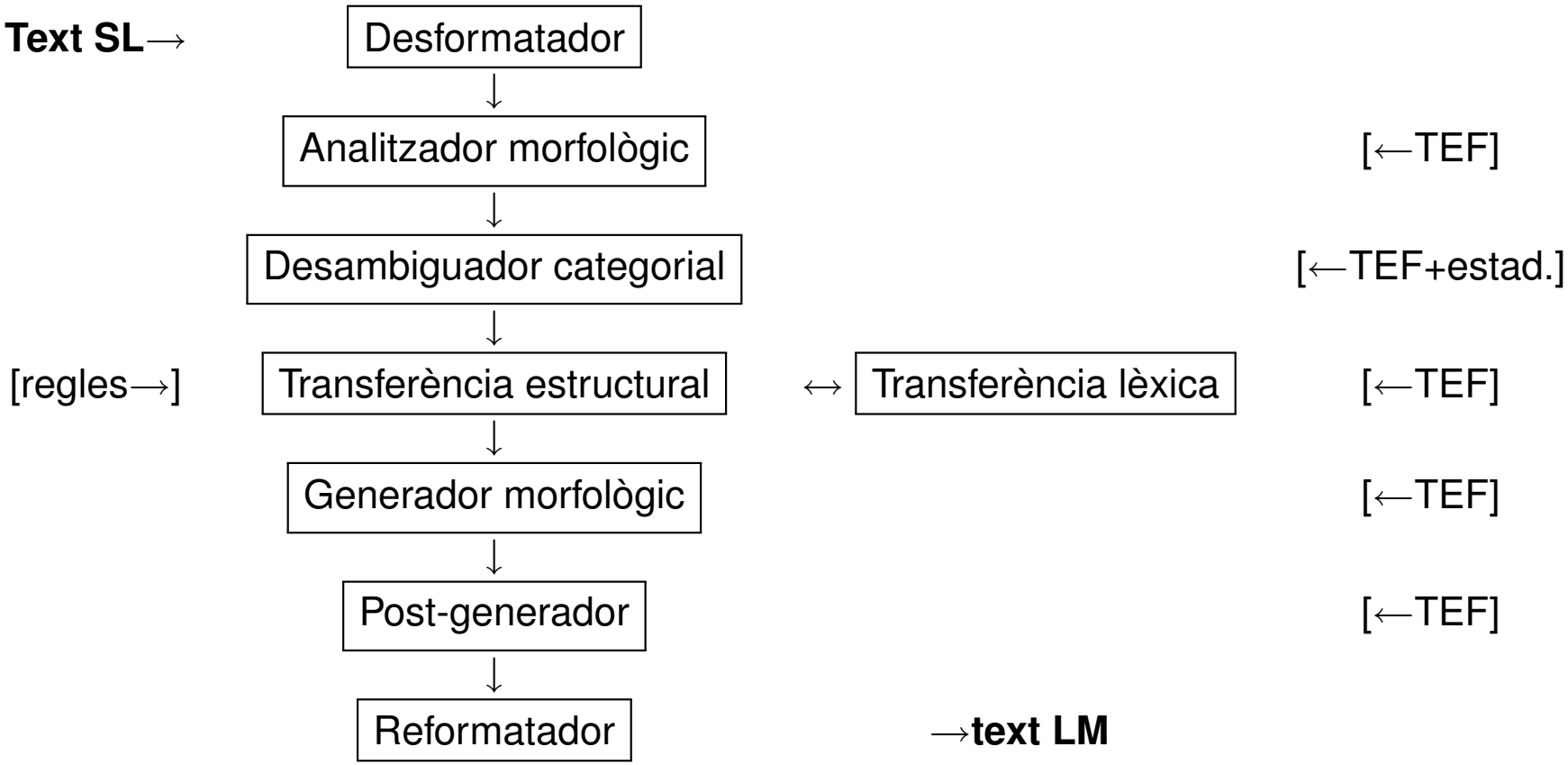
  per part de la indústria, de les universitats o d'organitzacions de suport de llengües menors.
- Promoure el canvi de model de negoci en traducció automàtica, del model *basat en llicències* (obsolescent) a un model *basat en serveis*.
- Garantir radicalment la reproducibilitat de la recerca en TA.
- Perquè no té sentit usar diners públics per a desenvolupar programari no lliure i de codi tancat.

# La plataforma Apertium

Apertium és una plataforma de traducció automàtica de codi obert (`http://www.apertium.org`) que proporciona:

1. Un **motor** de traducció automàtica, basat en transferència sintàctica superficial, amb:
   - gestió de formats de text
   - processament lèxic basat en estats finits
   - transferència sintàctica superficial basada en reconeixement de patrons basat en estats finits
2. **Dades lingüístiques** en formats XML ben especificats per un nombre creixent de parells de llengües
3. **Compiladors** per a passar les dades a la forma compacta i ràpida usada pel motor i programari per a aprendre regles de desambiguació o de transferència estructural.

# El motor d'Apertium/1



# El motor d'Apertium/2

Comunicació entre els mòduls: **text** (*canonades* o *pipelines* d'Unix).
Avantatges:

- Simplifica la diagnosi i la depuració d'errors
- Permet la modificació de dades entre dos mòduls, usant, per exemple, filtres
- Facilita la inserció de mòduls alternatius (interessant per a la recerca i el desenvolupament)

# Desformatador

- Separa el text de la informació de format.
- Actualment disponible per a text pla ISO-8859 o UTF-8, HTML, RTF, ODF, `.sxw` d'OpenOffice.org, etc.
- Basat en tècniques d'estats finits.
- La majoria d'aquests filtres es generen (usant un full d'estil XSLT) a partir d'un fitxer XML d'especificació de desformatadors.

# Analitzador morfològic

- segmenta el text en llengua origen (LO) en *formes superficials* (FSs),
- assigna a cada FS una o més *formes lèxiques* (FLs), cada una amb:
  - lema
  - categoria lèxica (part de l'oració)
  - informació de flexió morfològica
- processa contraccions i unitats lèxiques multi-mot que poden ser invariables (`es`: *con cargo a*, *de suerte que*) o variables (`es`: *echaría de menos* → *echar de menos*).
- llig transductors d'estats finits *compilats* a partir d'un diccionari morfològic en XML.

# Desambiguador lèxic categorial

- tria una de les FLs corresponents a cada FS ambigua (n'hi ha un 30%) segons el context
- usa models de Markov ocults i restriccions escrites a mà
- s'entrena usant corpus representatius per a la llengua origen (desambiguats manualment o no) o, més recentment, usant models estadístics de la llengua meta
- està controlat per un arxiu XML

# Transferència estructural /1

- Mòdul basat en tècniques d'estats finits (reconeixedors d'estats finits).
- El fitxer de regles de transferència XML es preprocessa perquè siga interpretat més ràpidament.
- Les regles tenen la forma *patró–acció*.
- Detecta patrons de FLs que es processen d'esquerra a dreta i elegint el patró concordant més llarg.
- Executa les accions associades a cada patró en el fitxer de regles per a generar el patró de LFs corresponent en la llengua meta.

# Transferència estructural /2

Per a parells de llengües "més difícils", hi ha disponible una transferència estructural en tres etapes:

- Es detecten, processen i marquen patrons de FLs (*xuncs* o *chunks*)
- Es detecten i processen patrons de *xuncs*: aquest processament *inter-xunc* permet transformacions sintàctiques d'abast més llarg
- Els *xuncs* d'eixida son reprocessats si és necessari i les FLs resultants s'envien a l'eixida.

# Mòdul de transferència lèxica

- llig cada FL de la LO i genera la FL corresponent en LM
- llig transductors d'estats finits *compilats* a partir de diccionaris bilingües en XML
- és invocat pel mòdul de transferència estructural

# Generador morfològic

- Genera, flexionant adequadament cada FL en LM, la FS corresponent
- Llig transductors d'estats finits *compilats* a partir de diccionaris morfològics en XML

# Post-generador

- Realitza transformacions ortogràfiques com ara contraccions (`ca`: *de* +*els* $\longrightarrow$ *dels*; `en`: *can* + *not* $\longrightarrow$ *cannot*), o inserció d'apòstrofs (`ca`: *de* + *amics* $\longrightarrow$ *d'amics*), etc.
- Es basa en transductors d'estats finits *compilats* a partir de diccionaris de regles de post-generació.

# Reformatador

- Reintegra la informació de format (text pla ISO-8859 or UTF-8, HTML, RTF, ODT, `.sxw` d'OpenOffice.org, etc.) en el text traduït.
- S'usa també per a modificar els URLs dels enllaços per a *navegar i traduir*.
- Es basa en tècniques d'estats finits.
- Es genera (usant un full d'estil XML) a partir d'un fitxer XML d'especificació de formats de text.

# Dades lingüístiques

El projecte Apertium acull el desenvolupament de dades per a un gran nombre de parells de llengües:

- Entre els parells *estables* hi ha: es↔ca, es↔gl, es↔pt, pt↔ca, pt↔gl, en↔ca, en↔es, es↔fr, ca↔oc, ro→es, es→eo, ca→eo.
- També hi ha un nombre creixent de parells de llengües en desenvolupament.

# Finançament

Finançat per:

- 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 Inforamció de la Generalitat de Catalunya
- El Ministeri d'Assumptes Exteriors de Romania
- La Universitat d'Alacant
- Empreses: Prompsit Language Engineering, ABC Enciklopedioj, imaxin|software, etc.

# La comunitat d'Apertium/1

No és la situació ideal de desenvolupament comunitari, però s'hi acosta bastant. A més dels desenvolupadors originals (finançats), s'ha format una comunitat al voltant del projecte (instigada fonamentalment per Francis Tyers).

- Hi ha més de 60 desenvolupadors inscrits en `sourceforge.net/projects/apertium/`, molts de fora del grup original; el codi s'actualitza molt freqüentment (centenars d'actualitzacions SVN cada mes).
- Un *wiki* mantingut col·lectivament mostra l'estat actual de desenvolupament i dóna consells per als desenvolupadors de dades lingüístiques o de programes.

# La comunitat d'Apertium/2

- Exemples d'eines i codi desenvolupat externament:
  - la interfície gràfica d'ús `apertium-tolk`, i l'eina de diagnòstic `apertium-view`
  - *plugins* per a OpenOffice.org o per al missatger Pidgin (abans Gaim)
  - Una versió dels diccionaris bilingües per a mòbils (`tinylex`)
  - Versions preliminars per a Windows
- Molts desenvolupadors es troben en el canal IRC `#apertium` (de `freenode.net`).
- Els paquets estables estan disponibles en Debian GNU/Linux (i per tant, en Ubuntu).

# Exemple: Apertium i l'occità/1

- El desenvolupament de TA per a l'occità va començar en 2006 amb el parell aranés–català (`oc@aran`↔`ca`), finançat per la STSI de la Generalitat de Catalunya (Universitat d'Alacant i Universitat Pompeu Fabra): connectava
  - una llengua *mitjana* (`ca`, ≃6.000.000 parlants) i
  - una variant estandarditzada *molt menuda* (`oc@aran`, ≃6.000 parlants d'una llengua més gran, `oc`, ≃1.000.000 parlants?)

# Exemple: Apertium i l'occità/2

- El 2007 Prompsit i Taller Digital són contractats per la Generalitat de Catalunya per a construir els traductors oficials `oc↔ca` i `oc↔es`, també per a l'occità general.
- Problema: estandardització de l'occità general (*occitan larg*): iniciativa pionera.
- Es crea una comissió d'experts lingüístics de *quasi* tot Occitània (2 experts per *regió*) amb participació d'una experta d'Apertium
- El model de llengua elegit (no sense llargues discussions) està basat en el llenguadocià.

# Exemple: Apertium i l'occità/3

Amb un sistema que té el 95% de cobertura i una taxa d'error del 10% per a l'`oc@aran↔ca` i del 25% d'error per a l'`oc↔ca` (*occitan larg*) (millorable!)

- La quantitat de text en occità en la web pot augmentar (visibilitat)
- L'existència de traducció automàtica de qualitat pot promoure la difusió de les variants de l'occità tractades.
- La comunitat occitana general (la majoria a França) pot crear un traductor `oc–fr` a partir de les dades `oc–ca` o `oc–es` i `fr–ca` o `fr–es` existents.
- Fa disponibles públiques dades obertes d'occità, útils per a crear aplicacions de tecnologia lingüística.

# Índex

# Conclusions temptatives

- La TA pot tenir un efecte positiu en llengües menors (normalitat, “visibilitat”, alfabetització, estandardització)
- La TA de *codi obert* pot tenir efectes *específics addicionals* (augment de la perícia lingüística, aportació de recursos reutilitzables, reducció de la dependència tecnològica).
- El desenvolupament de TA per a una llengua menor s’enfronta a bastants reptes (falta d’estandardització, actituds tecnofòbiques, elicitació del coneixement lingüístic, necessitat de formats estàndards, modularitat).
- Apertium ofereix a les llengües menors una plataforma de codi obert (amb solucions per als dos últims problemes).

Per descomptat, estaré encantat de debatre sobre tot això...

# Agraïments

Gràcies

- a totes les agències i empreses citades abans,
- a tots els desenvolupadors d'Apertium,
- i als organitzadors d'aquesta reunió per haver-me convidat.