

poli [Traductor]

Guia d'ús del poli[Traductor]

<http://politrador.upv.es/>

17.09.2013



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Índex de continguts

Introducció.....	3
1. Observacions generals.....	3
1.1. Motor de traducció.....	4
1.2. Traducció automàtica.....	4
1.3. Codi obert.....	4
1.4. Recomanacions generals.....	4
2. Opcions disponibles.....	4
2.1. Opció Text.....	4
2.2. Opció Documents.....	6
2.3. Opció Pàgines web.....	8
2.4. Opció TMX.....	9
2.5. Opció Android.....	10

Introducció

El poliTraductor (<http://politrador.upv.es/>) és un portal de traducció basat en Apertium (<http://www.apertium.org>), un motor de traducció automàtica de codi obert desenvolupat, entre altres, per l'empresa Prompsit (<http://www.prompsit.com>), amb la qual ha col·laborat la Universitat Politècnica de València (<http://www.upv.es>). Aquest motor permet fer traduccions del valencià al castellà i viceversa. Les opcions de traducció són diverses, i es poden traduir textos escrits directament a la finestra de la interfície del traductor, documents en múltiples formats, pàgines web, etc.

La versió actual d'aquest motor de traducció es va presentar a la UPV el setembre del 2011. Des d'aleshores, hem rebut suggeriments, idees i correccions per part de la comunitat universitària i altres usuaris que ens han ajudat a millorar-lo.

Cal remarcar que el document que surt del motor de traducció en qualsevol de les llengües que ofereix és un esborrany i que cal, imprescindiblement, fer una revisió acurada i extensa del document final.

Esperem que aquesta eina lingüística siga útil i eficaç per a les tasques de l'activitat docent i acadèmica de la comunitat universitària i del públic en general.

1. Observacions generals

Apertium és una plataforma de traducció automàtica de codi obert. Això vol dir que tots els components de la plataforma poden ser utilitzats, copiats, estudiats, modificats i redistribuïts per qualsevol usuari. En aquest apartat introduïm molt breument quin és el funcionament general del motor, les consideracions que cal tenir en compte a l'hora de treballar en traducció automàtica i algunes recomanacions generals per a un bon ús i funcionament de l'eina.

1.1. Motor de traducció

La plataforma de traducció automàtica de codi obert **Apertium** funciona amb un motor de traducció amb diversos parells de llengües, com ara català(val)-castellà, català/castellà-francès, català-portuguès, etc.

Apertium fa servir un motor de traducció de transferència sintàctica que processa el text d'entrada en etapes, com en una cadena de muntatge: desformatatge, anàlisi morfològica, desambiguació de categories morfològiques, transferència estructural, transferència lèxica, generació morfològica i reformatatge.

1.2. Traducció automàtica

La traducció automàtica consisteix en l'obtenció d'un text en una **llengua de destinació** equivalent i que preserve el contingut d'un text en una **llengua d'origen**.

La ciència actual no és capaç d'expressar de manera formal, mitjançant regles, tots els mecanismes que són subjacents al llenguatge natural ni als processos mentals involucrats en la comunicació. Aquesta impossibilitat per a obtenir una caracterització precisa és un dels principals arguments que es poden donar per a explicar la complexitat d'escriure un programa d'ordinador que tradueix textos.

La traducció automàtica continua sent un problema difícil de resoldre. El problema es redueix considerablement quan les llengües implicades en la traducció estan emparentades, com ara el català i l'espanyol o el català i el francès. En aquest cas, les afinitats morfològiques, sintàctiques i semàntiques simplifiquen el disseny d'aquests sistemes i permeten arribar fàcilment a traduccions en les quals només una part relativament petita del text és incorrecta. En qualsevol cas, si es vol difondre el text, s'ha de fer una revisió de la traducció lliurada pel sistema de traducció automàtica.

1.3. Codi obert

El motor de traducció de la plataforma Apertium és de codi obert. Això vol dir que pot ser utilitzat, copiats, estudiats, modificats i redistribuïts amb l'única restricció que el codi font ha d'estar sempre disponible. Per descomptat, Apertium va acompanyat d'una llicència compatible amb aquests principis, la llicència pública general (GPL, General Public License). El motor de traducció Apertium, les eines auxiliars, la documentació corresponent i la majoria de les dades lingüístiques desenvolupades fins avui per Apertium es poden baixar des del web <http://apertium.sourceforge.net>.

1.4. Recomanacions generals

Per a obtenir una bona traducció és recomanable dur a terme els passos següents:

- Escriure o copiar i enganxar textos i documents que siguen ortogràficament i sintàcticament ben escrits.
- Redactar textos amb oracions simples, el sistema de traducció funcionarà millor que amb oracions molt complexes.
- Comprovar les errades sistemàtiques per a unificar la correcció del text.
- Revisar sempre el text resultant abans de publicar-lo o difondre'l, atès que el resultat de la traducció automàtica es considera un esborrany.

2. Opcions disponibles

En aquest apartat descrivim les cinc opcions de traducció i les funcions que actualment hi ha disponibles en el Politractor.

2.1. Opció Text

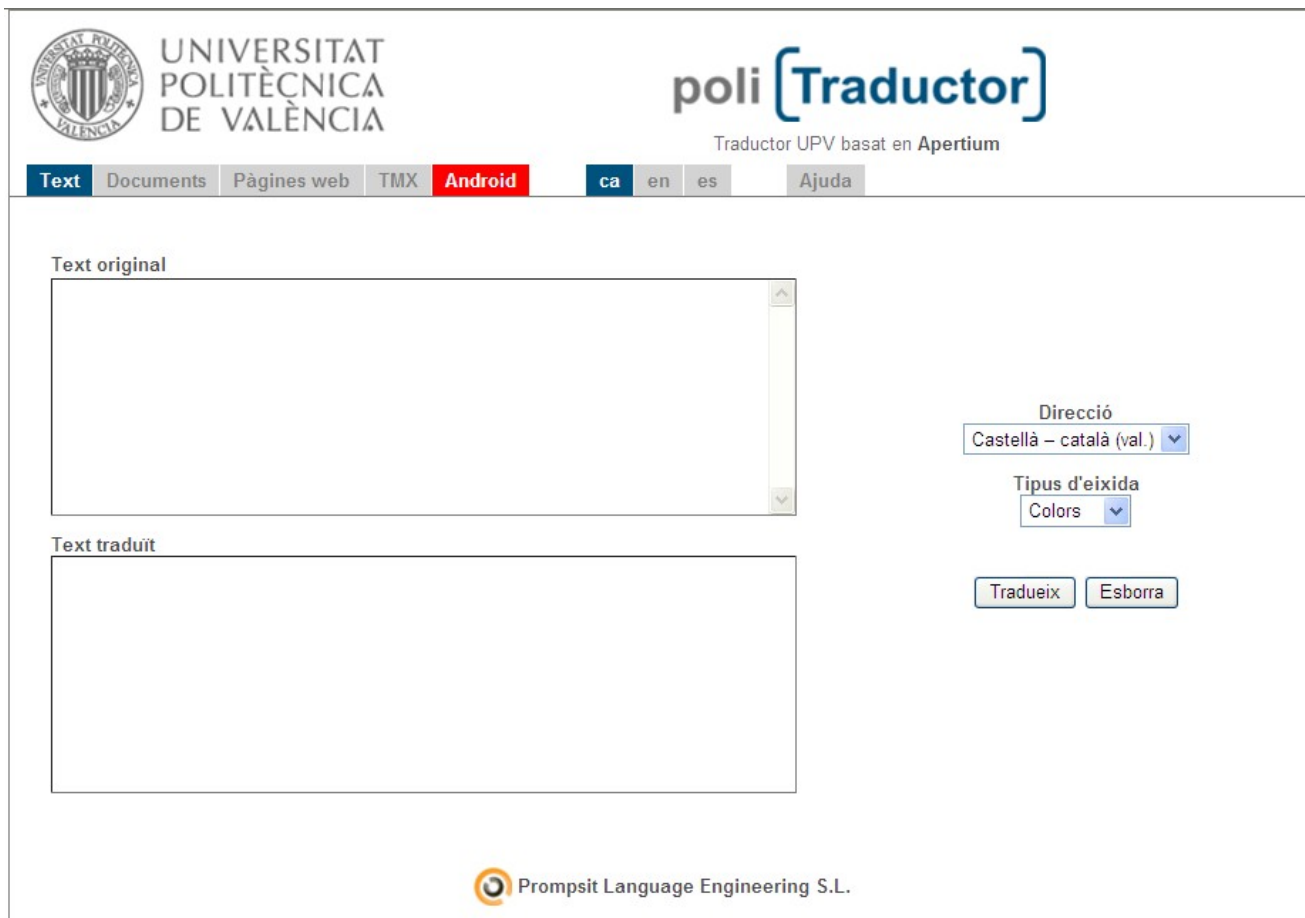


Figura 1. Captura de pantalla de traducció de textos curts

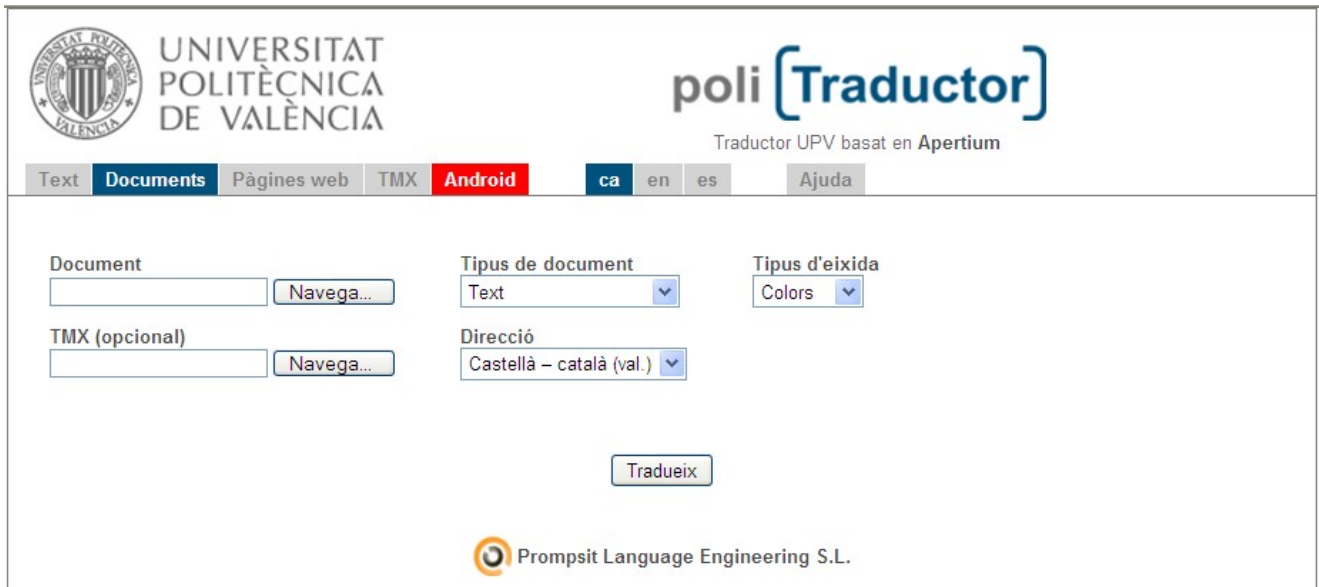
La primera opció és la de **traducció de textos curts** (figura 1), que es recomana quan l'objectiu és la traducció d'oracions o paraules soltes. Cal seguir aquestes accions:

- Escriure o copiar i enganxar el text a la caixa de **text original** que apareix a la pantalla.
- Seleccionar la **direcció** de la traducció que voleu obtenir: llengua origen → llengua destinació. Podeu seleccionar una de les direccions següents:
 - castellà → català (val)
 - castellà → català
 - català → castellà
 - anglès → català (val)
 - anglès → català
 - català → anglès
 - castellà → anglès
 - anglès → castellà
- Triar el **tipus d'eixida** que voleu veure:
 - a) Text net** mostra el text traduït pel sistema sense cap marca.
 - b) Colors** mostra les paraules desconegudes amb marques (dos asteriscos de color vermell) o diverses accepcions d'una paraula (separades per una barra vertical verda i entre claus).

- Prémer el botó **Tradueix**.

La traducció resultant apareixerà en una caixa de **text traduït** segons les opcions que s'hagen triat.

2.2. Opció Documents



The screenshot shows the 'poli Traductor' web interface. At the top left is the logo of the Universitat Politècnica de València. The main title is 'poli Traductor' with the subtitle 'Traductor UPV basat en Apertium'. Below the title is a navigation bar with tabs for 'Text', 'Documents', 'Pàgines web', 'TMX', 'Android', 'ca', 'en', 'es', and 'Ajuda'. The 'Documents' tab is selected. The interface includes several input fields and dropdown menus: 'Document' with a 'Navega...' button, 'Tipus de document' with a dropdown menu set to 'Text', 'Tipus d'eixida' with a dropdown menu set to 'Colors', 'TMX (opcional)' with a 'Navega...' button, and 'Direcció' with a dropdown menu set to 'Castellà - català (val.)'. A 'Tradueix' button is located below these fields. At the bottom, there is a logo for 'Prompsit Language Engineering S.L.'.

Figura 2. Captura de pantalla de traducció de documents

La segona opció és la traducció de **documents** (figura 2). Es poden traduir documents en diversos formats. Cal seguir aquestes accions:

- Cercar amb **Examina/Navega... (o Selecciona el fitxer)** el document que voleu traduir.
- Seleccionar el **tipus de document**. És imprescindible que es dese el document que heu de traduir en el format que heu triat. A la taula 1 es presenta un quadre amb els formats possibles:

Nom del format	Descripció
Text	Text pla codificat en ISO-8859-1/15 o UTF-8
HTML	Hypertext Markup Language
RTF	Rich Text Format, un dels formats del processador de textos Word de Microsoft Office
ODT	Open Document Text. format del processador de textos d'OpenOffice.org
ODS	Open Document Spreadsheet. format del full de càlcul d'OpenOffice.org
ODP	Open Document Presentation. format de l'aplicació Impress d'OpenOffice.org
SXW	Format del processador de textos StarOffice compatible amb OpenOffice.org 1.0
DOC	Document de Word de Microsoft Office 97 i posterior
XLS	Document d'Excel de Microsoft Office 97 i posterior
PPT	Document de PowerPoint de Microsoft Office 97 i posterior
DOCX	Document de Word de Microsoft Office 2007 i posterior
PPTX	Document de PowerPoint de Microsoft Office 2007 i posterior
XLSX	Document d'Excel de Microsoft Office 2007 i posterior
Document Word XML	Document de Word en format XML de Microsoft Office 2003
LaTeX	Documents amb format de marcatge TeX amb accents tipus "\grave{a}"
LaTeX amb accents	Documents amb format de marcatge TeX amb accents tipus "à"

Taula 1. Formats de documents possibles i significat

- Seleccionar la **direcció** de la traducció que voleu obtenir: llengua origen → llengua destinació. Podeu seleccionar una de les direccions següents:
 - castellà → català (val)
 - castellà → català
 - català → castellà
 - anglès → català (val)
 - anglès → català
 - català → anglès
 - castellà → anglès
 - anglès → castellà
- Triar el **tipus d'eixida** que voleu veure, igual que a l'opció de **traducció de textos curts**:
 - a) Text net** mostra el text traduït pel sistema sense cap marca
 - b) Colors** mostra les paraules desconegudes amb marques (dos asteriscos de color vermell) o diverses accepcions d'una paraula (separades per una barra vertical verda i entre claus).
- Prémer el botó **Tradueix**.

Com es pot veure a la figura 3, el document que conté el resultat de la traducció automàtica apareix en una finestra perquè el pugueu baixar amb un nom anàleg al del document original, amb la notació del sentit de traducció triat incorporada. També apareix el nombre de paraules del document original que s'han traduït.



Figura 3. Resultat de la traducció de documents

2.3. Opció Pàgines web



Figura 4. Captura de pantalla de la traducció de pàgines web

Amb aquesta opció es poden traduir pàgines d'Internet (figura 4). S'han de fer les accions següents:

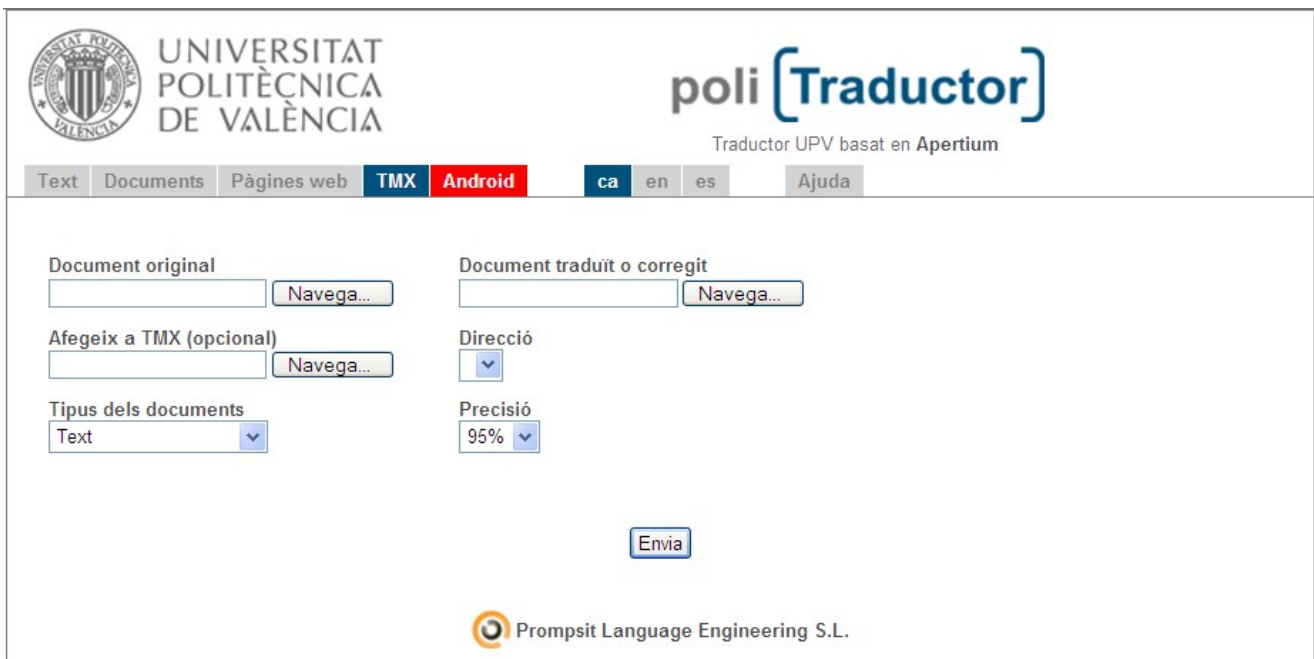
- Escriure o copiar i enganxar la URL en la caixa de text. Cal comprovar que la URL és correcta i no és una redirecció a una altra pàgina. Per a comprovar-ho, heu d'enganxar la URL completa que hi ha a la barra del navegador quan entreu a la pàgina que voleu traduir.
- Seleccionar correctament el sentit de la traducció que voleu al menú del cantó superior de la dreta (llengua origen → llengua destinació).
- Triar el **tipus de sortida** que voleu veure:
 - a) **Text net** mostra el text traduït pel sistema sense cap marca

b) Colors mostra les paraules desconegudes amb marques (dos asteriscos de color vermell) o diverses accepcions d'una paraula (separades per una barra vertical verda i entre claus).

- Triar la codificació de caràcters de la URL (UTF-8 o ISO-8859-15) que voleu traduir. Prémer el botó **Tradueix**.

La traducció resultant apareixerà en una finestra de navegador nova segons les opcions que hàgeu triat i amb un enllaç, a la part superior, que portarà a la pàgina original.

2.4. Opció **TMX**



La imatge mostra la interfície web de poli Traductor, un traductor basat en Apertium de la Universitat Politècnica de València. La interfície està dissenyada amb un menú de navegació superior que inclou 'Text', 'Documents', 'Pàgines web', 'TMX' (seleccionat), 'Android', 'ca', 'en', 'es' i 'Ajuda'. El títol principal és 'poli Traductor' amb el subtítol 'Traductor UPV basat en Apertium'. La interfície principal té dos camps d'entrada per a 'Document original' i 'Document traduït o corregit', amb botons 'Navega...' per a cada un. També hi ha un camp 'Afegeix a TMX (opcional)' amb un botó 'Navega...'. A sota, hi ha un menú desplegable 'Tipus dels documents' amb 'Text' seleccionat, i un menú 'Precisió' amb '95%' seleccionat. Un botó 'Envia' està centrat a la part inferior. Al peu de pàgina, hi ha el logotip i el nom de 'Prompsit Language Engineering S.L.'.

Figura 5. Captura de pantalla de creació de TMX

L'opció **TMX** permet generar un fitxer TMX per a crear una memòria de traducció a partir de dos textos: un text original i un altre amb la traducció corresponent. La memòria de traducció resultant consisteix en una sèrie de frases alineades: cada frase original anirà associada a la traducció corresponent. Cal seguir aquestes accions:

- Cercar amb **Examina/Navega...** el document original que voleu traduir.
- Cercar amb **Examina/Navega...** el document traduït.
- Triar amb el menú desplegable el tipus de documents que voleu tractar. Compte! Els dos documents han de ser del mateix format.

- | | |
|--------|--------|
| → Text | → PPTX |
| → RTF | → DOC |
| → ODT | → DOCX |
| → HTML | → XLS |
| → ODS | → XLSX |
| → PPT | |

<p>Document original</p> <input type="text" value="ió'WOK_Apertium.doc"/> <input type="button" value="Navega..."/>	<p>Document traduït o corregit</p> <input type="text" value="'OK_Apertium_ES.doc"/> <input type="button" value="Navega..."/>
<p>Afegeix a TMX (opcional)</p> <input type="text"/> <input type="button" value="Navega..."/>	<p>Direcció</p> <input type="text" value="Català - Castellà"/> ▼
<p>Tipus dels documents</p> <input type="text" value="DOC"/> ▼	<p>Precisió</p> <input type="text" value="95%"/> ▼
<input type="button" value="Envia"/>	

Figura 6. Detall de creació de TMX a partir de dos documents DOC

- De manera opcional, podeu afegir noves frases alineades a un TMX preexistent. És recomanable desfer el nou TMX amb un nom diferent per a evitar afegir frases que no hagen quedat ben alineades.
- Seleccionar correctament la **direcció** de la traducció (llengua origen → llengua destinació).
- Prémer el botó **Envia** per a enviar la petició de construcció de la memòria TMX.

2.5. Opció Android

Aquesta darrera opció us permet descarregar l'aplicació del poliTraductor en dispositius mòbils amb Android (aplicació amb extensió .apk).

L'aplicació té els avantatges següents:

- No cal Internet per a fer-la servir, només cal descarregar l'aplicació.
- Tradueix textos curts i adreces web (URL)
- Es pot usar el teclat o la veu per a interaccionar.
- Permet compartir fàcilment el text traduït.
- Té controls accessibles per a dispositius tàctils (botons per a triar la direcció de traducció, la generació o no de marques, etc.).
- Té missatges d'ajuda en línia.