

<http://uocpapers.uoc.edu>

## article

# Com es treballa en una aula virtual multilingüe? Un entorn d'aprenentatge multilingüe amb Moodle i Apertium

Antoni Oliver  
Cristina Borrell

Data de presentació: gener de 2008

Data d'acceptació: febrer de 2008

Data de publicació: abril de 2008

## Resum

La UOC, en el marc de la càtedra de multilingüisme UOC-Linguamón, ha desenvolupat un entorn virtual d'aprenentatge amb un sistema de traducció automàtica integrat. Gràcies a aquest projecte, que funciona amb les aplicacions de programari lliure Moodle i Apertium, es pot disposar d'un entorn d'aprenentatge multilingüe en català, castellà, anglès i francès. Aquest entorn s'alliberarà amb una llicència de programari lliure. El traductor inclou un detector de llengua i un mòdul de preedició que corregeix els missatges entrants sense cap mena d'intervenció de l'usuari. Això ajuda a millorar la qualitat de les traduccions resultants. Tot el sistema s'ha adaptat a la tipologia textual dels missatges electrònics amb un estudi previ dels missatges dels fòrums de la UOC.

## Paraules clau

Moodle, Apertium, traductor automàtic, multilingüe, entorn, programari lliure

## Abstract

*The UOC, within the framework of the Linguamón-UOC Chair in Multilingualism, has developed a virtual learning environment with an integrated machine translation system. Thanks to this project, which works with free software applications Moodle and Apertium, a multilingual learning environment can be provided in Catalan, English, French and Spanish. This environment is to be released under a free software licence. The translator includes a language detector and a pre-editing module that corrects incoming messages, without requiring any intervention from the user. This helps improve the quality of the resulting translations. The whole system has been adapted to the type of text seen in email messages following the prior study of the messages seen on the UOC's forums.*

## Keywords

*Moodle, Apertium, machine translator, multilingual, environment, free software*

## Introducció

En aquest article presentem un projecte que té com a finalitat integrar un sistema de traducció automàtica en un entorn virtual d'aprenentatge. S'ha escollit un sistema de traducció automàtica (Apertium) i un entorn virtual d'aprenentatge (Moodle) que es distribueixen amb una llicència de programari lliure. Tot el desenvolupament fet en el projecte també s'alliberarà amb una llicència de programari lliure. Aquest fet és molt important, ja que el producte final podrà ser utilitzat per qualsevol persona o institució d'una manera totalment lliure i gratuïta.

Les llengües incloses al projecte són el català, el castellà, l'anglès i el francès, però es podrà adaptar fàcilment a qualsevol

altra llengua mentre es disposi d'un sistema de traducció automàtica per a les llengües implicades.

L'article s'organitza de la manera següent: en primer lloc, expliquem breument què són les aplicacions Moodle i Apertium; després, descrivim els aspectes tècnics que s'han desenvolupat en aquest projecte per a integrar aquestes dues aplicacions, en concret la creació d'un corrector ortogràfic automàtic; a continuació, comentem els aspectes lingüístics que s'han estudiat per a alimentar els correctors ortogràfics, i, finalment, presentem el prototipus i les conclusions.

El projecte s'ha dut a terme amb el finançament de la Càtedra de Multilingüisme UOC-Linguamón.

<http://uocpapers.uoc.edu>

Com es treballa en una aula virtual multilingüe?

## Què és Moodle?

Moodle (<http://moodle.org>) és un sistema de gestió de cursos en línia (CMS, *course management system*) que es distribueix gratuïtament com a programari lliure amb la llicència pública GNU.<sup>1</sup> La paraula *Moodle* era al principi un acrònim de *Modular Object-Oriented Dynamic Learning Environment* (entorn d'aprenentatge dinàmic orientat a objectes i modular). Aquest sistema permet crear de manera senzilla cursos que es poden fer totalment en línia. Com que el sistema és molt fàcil d'instal·lar i mantenir, no té gaires requisits de maquinari ni de programari i la creació de cursos és molt senzilla, Moodle s'ha convertit en un sistema CMS realment popular.



Figura 1. Aspecte de l'entorn d'aprenentatge Moodle.

## Què és Apertium?

Apertium és un sistema de traducció automàtica de programari lliure que es distribueix amb la llicència GNU General Public License (GPL). De fet, més que un sistema de traducció automàtica és una *plataforma de traducció automàtica*. Això el que vol dir és que a partir d'aquest sistema es poden desenvolupar nous parells de llengües o bé millorar els parells de llengües ja existents. És a dir, a més de distribuir tot el que és necessari per a poder fer servir el sistema de traducció automàtica també es distribueixen totes les eines necessàries per a poder crear nous sistemes. Així, doncs, la plataforma Apertium proporciona:

- un motor de traducció que és comú per a tots els parells de llengües disponibles,
- eines per a gestionar les dades lingüístiques necessàries per a construir un sistema de traducció automàtica per a un determinat parell de llengües,
- dades lingüístiques per a diversos parells de llengües.

Apertium fa servir una metodologia de traducció automàtica que s'anomena *de transferència superficial*. Aquesta metodologia el que fa és una anàlisi sintàctica de l'oració en la llengua de partida i transferir aquesta anàlisi sintàctica a la llengua d'arribada per a poder generar l'oració d'arribada. Aquesta anàlisi sintàctica no és profunda, sinó superficial. Això vol dir que l'arbre sintàctic que es genera no és complet i poden faltar relacions entre alguns constituents. Aquesta metodologia de traducció automàtica és concebuda inicialment per a parells de llengües emparentades (com poden ser català-castellà, català-francès, castellà-portuguès, etc.). De totes maneres, s'ha expandit la metodologia per a poder tractar parells de llengües més divergents (català-anglès, per exemple).

En Apertium el processament del text d'entrada es fa en diverses etapes:

- desformatatge: es tracten diversos formats d'entrada (com text, RTF, HTML, etc.) i se separa el text que s'ha de traduir de la informació de format;
- anàlisi morfològica: a cada paraula del text que s'ha de traduir s'hi assignen totes les informacions morfològiques possibles;
- desambiguació categorial: es desambigua la informació morfològica assignada a l'etapa anterior, de manera que per a cada paraula només quedi una possible lectura;
- anàlisi sintàctica superficial: es construeix un arbre d'anàlisi superficial de la frase de partida;
- transferència sintàctica superficial: es transfereix l'arbre d'anàlisi superficial de la frase de partida a un arbre d'anàlisi superficial equivalent en la llengua d'arribada;
- transferència lèxica: es busca la traducció en la llengua d'arribada de les paraules de la frase de partida, de totes les possibles traduccions el sistema escollirà la més adequada;
- generació morfològica: es generen les formes adequades de les paraules traduïdes (en plural o singular, el temps i la persona per als verbs, etc.);
- reformatatge: es recupera la informació de format per a recuperar el format del document original.

La plana web d'Apertium és <http://xixona.dlsi.ua.es/apertium-www/>. En aquesta plana web es pot consultar diversa informació sobre el projecte. Les baixades es poden fer des d'aquesta plana o des de Source Forge: <http://sourceforge.net/projects/apertium/>.

1. La llicència pública general de GNU és una llicència de programari lliure molt estesa. Podeu trobar-ne més informació a <http://www.gnu.org/licenses/licenses.ca.html>.

## Aspectes tècnics

En projectes anteriors (Climent *et al.*, 2005, i Moré *et al.*, 2005) s'han estudiat a fons les característiques dels missatges enviats a fòrums acadèmics i s'ha observat que en general es produeix una quantitat de faltes d'ortografia elevada. També presenten algunes característiques pròpies (com abreviatures, ús d'emoticones, etc.) que fan que el rendiment dels sistemes de traducció automàtica per a aquest tipus de missatges no sigui gaire satisfactori. La connexió de Moodle amb Apertium no es fa directament, és a dir, els missatges que apareixen a Moodle no es tradueixen directament amb Apertium.

El primer pas del procés de traducció dels missatges és la detecció automàtica de la llengua en què està escrit el missatge que s'ha de traduir. En el nostre sistema els missatges poden estar escrits en català, castellà, anglès o francès. Caldrà saber, doncs, en quina llengua està escrit un missatge. Per a evitar la intervenció de l'usuari en tot el procés, aquesta informació no es demana, sinó que es determina mitjançant un algorisme de detecció automàtica de llengua. L'algorisme que hem utilitzat és el TextCat de Gertjan van Noord (<http://www.let.rug.nl/~vannoord/TextCat/>) basat en Cavnar (Cavnar *et al.*, 1994).

Un cop determinada la llengua del missatge, i com a pas previ a la traducció, els missatges passen per un corrector ortogràfic. El corrector ortogràfic s'ha dissenyat perquè sigui totalment automàtic i no requereixi cap intervenció de l'usuari. Això fa que només corregeixi les faltes que siguin clarament errors. Una opció alternativa seria fer ús d'un corrector ortogràfic normal que requereixi la intervenció de l'usuari. Aquesta opció es va descartar perquè pot provocar el rebuig de l'usuari, que es trobaria amb la necessitat de fer un pas més. En tot cas, l'usuari pot fer ús d'altres correctors ortogràfics, com per exemple el que ofereix el navegador Mozilla. El mòdul de correcció ortogràfica que hem desenvolupat funciona en tres etapes:

- Una primera etapa fa una sèrie de canvis tipogràfics en el text. Aquest mòdul funciona amb expressions regulars de l'estil `s/\'/g`, que fa que es canviïn els apòstrofs expressats com a accents per apòstrofs normals. Una altra transformació que fa, per exemple, és el canvi del punt baix de les eles geminades pel punt volat.
- Una segona etapa s'encarrega de fer una sèrie de canvis *ad hoc* que són especificats per un fitxer que es pot adaptar a les pròpies necessitats. Aquests canvis *ad hoc* serveixen per a expressar errors típics i les seves solucions, per a resoldre els casos en què la tria de l'opció correcta en la tercera etapa sigui difícil o impossible de trobar. Per exemple, podria expressar canviar *\*susto* per *ensurt*.
- La tercera etapa de la correcció ortogràfica funciona mitjançant un diccionari de formes o formari i un o més

diccionaris donats per l'usuari. Un diccionari de formes o formari és una llista de «totes» les paraules de la llengua en totes les seves formes. Bé, quan diem «totes» volem dir moltes, el major nombre possible, i quan diem «en totes les seves formes» ens referim a totes les seves possibles flexions (singular, plural, els diferents temps verbals i persones i nombres, etc.). Aquest formari es pot completar amb un o més diccionaris personalitzats. El corrector verifica si totes les paraules del text que s'ha de corregir són en aquesta llista i si no hi són intenta trobar la paraula de la llista que tingui més possibilitats de ser la paraula correcta. Que una paraula no sigui a la llista no vol dir necessàriament que es tracti d'una falta d'ortografia. Un cas típic d'aquesta situació són els noms propis, especialment els estrangers, que no estan inclosos al formari. Per evitar aquest fet, el corrector que hem dissenyat només corregeix les paraules que estiguin escrites amb minúscules. Això es pot considerar una limitació del sistema, però cal recordar que el nostre corrector no té cap intervenció de l'usuari i que només pot corregir allò que sigui molt segur. Un cop té una paraula escrita amb minúscula no continguda al formari ni als diccionaris de l'usuari, l'algorisme ha de ser capaç de trobar les paraules més semblants a aquesta i decidir quina té més possibilitats de ser la correcció. Aquesta cerca, a més, ha de ser molt ràpida i s'ha de fer sobre una llista de formes molt extensa. Per fer això fa els següents passos:

- Elimina un caràcter en cada una de les posicions possibles i verifica si el resultat és al formari. Això serviria per a un error de l'estil *dessig* escrit en català en comptes de *desig*; en aquest cas el sistema provaria *essig*, *dssig*, *desig*, *desig*, *dessg*, *dessi*, i trobaria l'opció correcta: *desig*.
- Afegeix un caràcter a totes les posicions de la paraula i verifica si el resultat és al formari. Això ho prova per a tots els caràcters possibles de la llengua. Si continuem amb l'exemple de *dessig* el sistema faria *adessig*, *bdessig*, *cdessig*, ..., *daessig*, *dbessig*, etc. En aquest cas no trobaria cap paraula.
- Canvia cada caràcter de la paraula per tots els caràcters possibles de la llengua i verifica si el resultat és al formari. Continuant amb el mateix exemple tindriem *aessig*, *bessig*, *cessig*, ..., *dassig*, *dbssig*, *dcssig*, etc. En aquest cas trobaria dues paraules de la llengua possibles: *dessi* i *dessin*, totes dues formes del verb *dar*.
- Permuta totes les lletres consecutives de la paraula. L'algorisme faria *edssig*, *dsesig*, *dessig*, *desisig*, *desgi*. Tampoc en aquest cas trobaria cap paraula de la llengua.

El corrector ara ha de decidir quina de les tres possibles formes correctes (*desig*, *dessis* i *dessin*) és la que té més possibilitats de ser la correcció de *dessig*. Per a fer això s'ha desenvolupat una estadística d'errors més freqüents, que ens indicarà que el canvi *ss:s* (de *dessig* per *desig*) és més freqüent que el canvi *g:s* (de *dessig* per *dessis*) i que el canvi *g:n* (de *dessig* per *dessin*).

Un cop corregit el missatge ja es pot traduir automàticament i presentar la versió original i les seves traduccions. La versió original que es presenta és la versió sense corregir ortogràficament, tot i que les traduccions que es presenten són generades a partir de la versió corregida. A la figura 2 podem veure esquemàticament els tres passos del procés de traducció automàtica dels missatges.

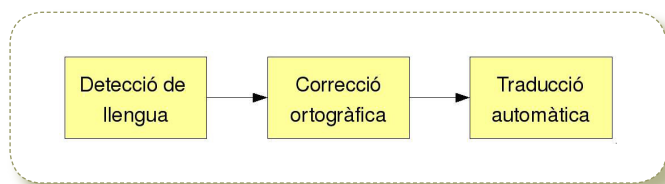


Figura 2. Passos del procés de traducció automàtica

## Aspectes lingüístics

En aquest apartat expliquem com s'han confeccionat els recursos lingüístics que alimenten el corrector ortogràfic automàtic. El corrector necessita els recursos següents:

- formaris per a cada una de les llengües,
- llistes d'*ad-hocs*,
- estadístiques d'errors més freqüents.

Les llistes d'*ad-hocs* i les estadístiques d'errors més freqüents s'han fet a partir d'un corpus de missatges. En aquest apartat expliquem d'on hem obtingut els formaris per a cada una de les llengües, com s'han confeccionat els corpus de missatges per a català i castellà i com són les llistes d'*ad-hocs* i d'estadístiques d'errors.

## Formaris

Un formari és una llista de paraules d'una determinada llengua flexionades en totes les formes que tenen. Aquestes llistes poden arribar a tenir mides considerables, especialment per a llengües força flexives, com el català, el castellà o el francès. En aquest apartat expliquem d'on hem obtingut els formaris per a les llengües de treball i la mida d'aquests formaris.

- Català: s'ha fet servir el diccionari morfològic que ha estat desenvolupat pel grup Glicom de la Universitat Pompeu Fabra per a la creació d'El Corrector, corrector ortogràfic i gramatical del català que s'alliberarà com a programari lliure. D'aquest diccionari s'han aprofitat només les formes i s'ha desestimat la resta d'informació. El formari del català té un total de 634.000 entrades.
- Castellà: s'ha fet servir el diccionari del corrector Ispell i té 653.000 entrades.
- Els formaris de l'anglès i del francès encara no s'han confeccionat, però molt probablement també es faran a partir dels diccionaris del corrector Ispell.

Com podem observar, les mides són considerables. Per a poder fer cerques molt ràpides en aquestes llistes, s'han convertit en bases de dades Berkeley. Les bases de dades Berkeley estan organitzades com a parells clau-valor i permeten fer una cerca molt ràpida. Les bases de dades que representen el formari fan servir la clau per a emmagatzemar la forma i el valor simplement s'instància (adquireix un valor qualsevol, en aquest cas 1), de manera que per a verificar si una forma és al formari només cal verificar si hi ha una entrada amb aquesta clau.

## Confecció dels corpus

Per a fer aquesta aproximació hem agafat una mostra dels missatges dels fòrums consultables al Campus Virtual de la UOC. La intenció era obtenir dos corpus: un per al català amb els missatges que hem trobat en català i un altre en castellà per als missatges en aquesta llengua, per tal de poder estudiar quin tipus de llenguatge s'empra en aquests casos. Dins de l'apartat *Comunitat > Fòrums*, ens hem dedicat a buidar els missatges classificats com a «Acadèmics». Això ha estat possible per al català i per al castellà, però no per a l'anglès ni el francès, perquè no disposàvem de les dades necessàries (no hi ha missatges als fòrums estrictament en aquestes llengües).

Hem intentat escollir els fòrums per la diversitat que ofereixien, de manera que compreguessin totes les temàtiques i tots els usuaris possibles, per tal de tenir un lèxic com més extens millor. Ens hem quedat amb els 25 missatges més recents de cada fòrum al moment de la consulta (al principi d'octubre de 2007). En total hem analitzat 290 missatges de 17 fòrums diferents, amb usuaris d'edats compreses entre 20 anys i majors de 45 anys, amb participació tant d'alumnes com de consultors.

A partir dels missatges, pretenem ser capaços de preveure els errors més freqüents en aquest tipus de registre i presentar solucions immediates. Els missatges van ser copiats literalment i desats en documents Word per tal de ser tractats posteriorment

<http://uocpapers.uoc.edu>

Com es treballa en una aula virtual multilingüe?

i analitzar els errors que es cometien habitualment quan s'escriu en aquest registre.

## Llistes d'*ad-hocs*

A partir dels corpus constituïts, hem analitzat els errors i les característiques d'aquesta tipologia textual, per tal de sistematitzar en la mesura que sigui possible la correcció automàtica d'errors. Aquest procés consisteix a identificar cada falta, categoritzar-la segons l'origen i buscar tècniques de correcció adequades i específiques a cada tipus d'equivocació. Això ens ha permès elaborar llistes d'errors freqüents o llistes d'*ad-hocs*: relació de paraules que normalment són susceptibles de ser mal escrites (i que consten com a errors en el corpus), juntament amb la correcció. Per a aquest traductor automàtic en concret, les llistes d'*ad-hocs* s'han fet a partir de dues fonts d'informació: principalment a partir dels errors trobats al corpus elaborat amb aquesta finalitat, però també de llistes de correcció dels serveis lingüístics de les universitats catalanes.

Així, doncs, hem constituït dues llistes d'*ad-hocs*: una per al català i una altra per al castellà. Com ja hem dit abans, per a l'anglès i el francès això no ha estat possible perquè no hem trobat les dades necessàries i, per tant, no hem pogut analitzar el «comportament estàndard» a l'hora d'escriure missatges al fòrum.

Cada llista té la paraula en què hem trobat l'error, la seva forma en femení (o masculí si escau) i les formes en plural, seguint el patró de l'error i corregides en cada cas:

volcanòleg	vulcanòleg
volcanòlegs	vulcanòlegs
volcanòloga	vulcanòloga
volcanòlogues	vulcanòlogues

En els casos en què la paraula mal escrita era un verb, hem inclòs a la llista d'*ad-hocs* les formes següents: l'infinitiu, la primera persona del present d'indicatiu, la primera persona del pretèrit imperfecte i, en els casos pertinents, la forma que apareixia mal escrita al corpus (fos el temps verbal que fos):

transtornar	trastornar
transtornava	trastornava
transtorno	trastorno

La llista d'*ad-hocs* també conté una relació estadística dels canvis més freqüents; a saber, per exemple, quantes vegades he

utilitzat *b* quan en realitat havia d'utilitzar *v*. D'això en direm *ponderació*, i l'estadística ens servirà per a decantar-nos per una forma o una altra en els casos en què hi hagi ambigüitat. Per exemple, cada vegada que ens indiqui que hi ha un error a la paraula *\*això*, el corrector sabrà que ha de substituir-la per *això*, perquè el canvi *o:ò* és un dels més freqüents del corpus. Per al castellà, la dinàmica que hem seguit ha estat exactament la mateixa. En total, la llista d'errors en català té 2.369 ítems incorrectes, dels quals 771 són formes verbals. Per al castellà, la llista d'*ad-hocs* té 445 ítems incorrectes, dels quals 142 són formes verbals.

## Estadístiques d'errors més freqüents

Hem comptat i classificat tots els errors del corpus segons els criteris establerts anteriorment i, com a dada més significativa (i d'acord amb les previsions), podem dir que la majoria d'errors són d'accentuació, i això és vàlid tant per al català com per al castellà. No és fàcil de dir si és una errada intencionada o no, perquè si bé és cert que el fet de no accentuar un text forma part del nou registre usat en els missatges electrònics, no es manté la coherència. Amb això volem dir que en un missatge tan aviat s'accentua com no s'accentua, per la qual cosa no podem considerar que sigui un error intencionat, sinó més aviat ortogràfic. En català, per exemple, ens hem trobat casos com aquest:

*\*informatica de sistemes o de gestió*

En el cas concret del català, si bé la majoria d'errors d'accentuació són causats per la manca d'accentuació (accentuació negativa), molts dels errors d'accentuació positiva (accentuar paraules incorrectament) denoten una clara influència del castellà (*podria*).

Podem apreciar que la quantitat d'accents «negatius» és molt important i que, per tant, és probable que comportin un grau d'errors molt elevat. Arribats en aquest punt, cal que fem la diferència entre els errors i els «no-errors», errors que són paraules incorrectes en el context en què són utilitzades. En un traductor automàtic, aquest tipus d'errors comporten molta controvèrsia. En aquest sentit, podem considerar el cas en català d'*angles* com a bon exemple per a il·lustrar els possibles problemes per al traductor. Si no hi ha cap tipus d'edició prèvia a la correcció automàtica, el traductor, per exemple, ens convertirà aquesta paraula en *ángulos* quan traduïm al castellà, si bé en cap cas no es fa referència a la figura geomètrica, sinó a la llengua: *anglés*.

Algunes paraules tenien més d'un error, per això hi ha més errors que paraules analitzades. Els casos d'error múltiple són més difícils de corregir, perquè no encaixen amb la tendència general establerta («el canvi sempre ha de ser mínim»).



<http://uocpapers.uoc.edu>

Com es treballa en una aula virtual multilingüe?

A tall de resum, la majoria dels ítems considerats a la llista d'errors són causats per una accentuació incorrecta, que en el cas del català és deguda a una forta influència del castellà i a la inseguretat dels usuaris a l'hora d'escriure en llengua catalana, i que en general es justifica per la manca d'atenció pròpia del registre escrit dels missatges de correu electrònic. Després de l'estudi, sí que podem afirmar, però, que es fan més faltes en català que en castellà.

## Prototipus

Actualment es desenvolupa el prototipus d'integració del corrector, Apertium, i Moodle. Un dels aspectes fonamentals d'aquesta integració és la forma de presentar les diferents versions d'un mateix missatge. La primera idea que ve al cap és presentar els missatges en totes les llengües (l'original i les diverses traduccions) una darrere de l'altra. Aquesta opció pot ser vàlida per a entorns bilingües, però si l'entorn és multilingüe no és una solució adequada, ja que de seguida les pantalles quedaran plenes de text i serà difícil de trobar el missatge en la llengua desitjada.

En el nostre prototipus s'ha desenvolupat un enfocament diferent. Cada usuari pot definir la seva llengua per defecte. L'usuari veurà en tot moment tots els missatges en la llengua volguda, sigui aquesta la llengua original o bé una traducció. Ara bé, saber si llegim una traducció o un original pot ser important, i per aquest motiu aquesta informació sempre és present. D'aquesta manera, podem saber si llegim un original o bé una traducció i, a més, en cas de llegir una traducció sabrem quina és la llengua original i tindrem un accés immediat a la versió original.

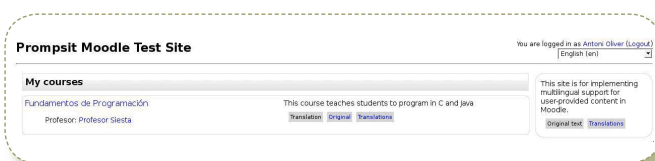


Figura 3. Plana d'inici del sistema si escollim com a llengua per defecte l'anglès

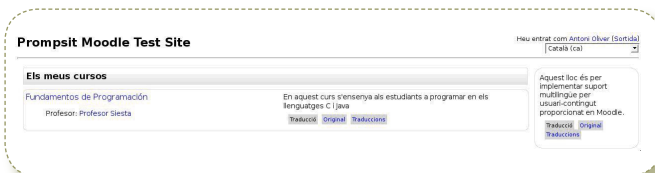


Figura 4. Plana d'inici del sistema si escollim com a llengua per defecte el català.

A la figura 3 podem veure la plana d'inici al sistema si escollim com a llengua per defecte l'anglès. Fixem-nos que el text de

la dreta, «This site is for implementing...», ens apareix en anglès i ens indica que és el text original. D'altra banda, a la figura 4 observem la mateixa pantalla si escollim com a llengua per defecte el català. Com podem veure, el text de la dreta apareix en català, però amb la informació que és una traducció. Mitjançant els enllaços sota el text podem visualitzar el text original (figura 5) o totes les traduccions (figura 6).

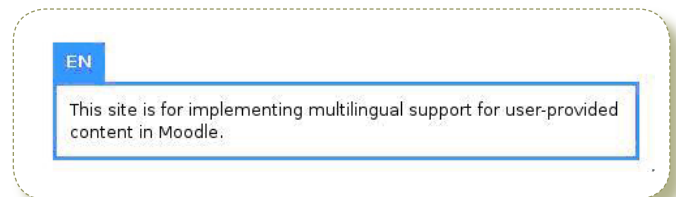


Figura 5. Pantalla que mostra el text original.

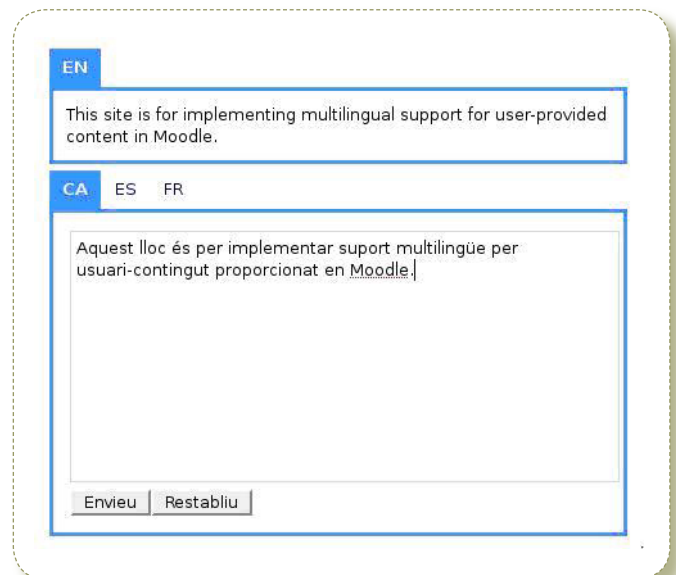


Figura 6. Pantalla que mostra el text original i les diferents traduccions (que es poden seleccionar amb les pestanyes).

Una característica interessant de la manera amb què s'ha implementat la visualització de les traduccions és que es pot establir una sèrie de permisos que permetin a tots els usuaris o a un grup d'ells modificar les traduccions fetes automàticament. D'aquesta manera un usuari amb permisos suficients pot corregir els possibles errors de traducció comesos pel traductor automàtic.

## Conclusions

L'entorn d'aprenentatge que hem presentat en aquest article està a punt per a acollir cursos i alumnes multilingües, mostrar

<http://uocpapers.uoc.edu>

Com es treballa en una aula virtual multilingüe?

els missatges en català, castellà, anglès o francès o traduir els missatges entre els usuaris. Per a saber si el funcionament general és satisfactori o no, si l'usuari troba la interfície agradable i fàcil d'usar i si el traductor tradueix correctament i en temps real, l'única etapa que ens queda per recórrer és el període de proves. Així, doncs, caldria posar l'entorn d'aprenentatge a la disposició dels usuaris en una situació real i dissenyar una enquesta per a descobrir el grau de satisfacció dels alumnes i dels consultors. Aquesta etapa de proves està a punt de dur-se a terme.

## Bibliografia

- CAVNAR, W. B.; TRENKLE, J. M. (1994). «N-Gram-Based Text Categorization». A: *Proceedings of Third Annual Symposium on Document Analysis and Information Retrieval*. Las Vegas: NV, UNLV Publications/Reprographics. Pàg. 161-175.
- CLIMENT, S.; MORÉ, J.; OLIVER, A. [et al.] (2005). «Tecnologies de la traducció per a la gestió de la doble oferta docent en català i castellà a la UOC». *Zeitschrift für Katalanistik/Revista d'Estudis Catalans*. Núm. 18, pàg. 31-57. ISSN 0932-2221.
- MORÉ, J.; CLIMENT, S.; OLIVER, A. [et al.] (2005). «Análisis de los fenómenos lingüísticos de los mensajes de correo electrónico en catalán desde la perspectiva de la traducción automática». A: *XXI Congreso de la SEPLN* (2005: Granada).

## Citació recomanada:

OLIVER, Antoni; BORRELL, Cristina (2008). «Un entorn d'aprenentatge multilingüe amb Moodle i Apertium». *UOC Papers* [article en línia]. Núm. 6. UOC. [Data de consulta: dd/mm/aa].  
<[http://www.uoc.edu/uocpapers/6/dt/cat/oliver\\_borrell.pdf](http://www.uoc.edu/uocpapers/6/dt/cat/oliver_borrell.pdf)>  
ISSN 1885-1541



Aquesta obra està subjecta a la llicència Reconeixement-NoComercial-SenseObraDerivada 2.5 de Creative Commons. Podeu copiar-la, distribuir-la i comunicar-la públicament sempre que n'especifiqueu l'autor i la revista que la publica (*UOC Papers*); no en feu un ús comercial; i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/2.5/es/deed.ca>



**Antoni Oliver**

Professor dels Estudis de Llengües i Cultures de la UOC  
Coordinador acadèmic del postgrau Traducció i tecnologies de la UOC  
[aoliverg@uoc.edu](mailto:aoliverg@uoc.edu)

Doctor en Lingüística, llicenciat en Filologia Eslava i enginyer tècnic de telecomunicacions. La seva àrea de recerca se centra en l'aplicació de tècniques de processament del llenguatge natural en les tasques de traducció. Participa en la docència en diversos màsters i cursos de postgrau a la Universitat Autònoma de Barcelona i a la Universitat Pompeu Fabra.



**Cristina Borrell**

Investigadora de l'IN3 (UOC)  
Tècnica de projectes de la càtedra UOC-Linguamón de Multilingüisme  
[cborrellc@uoc.edu](mailto:cborrellc@uoc.edu)

Llicenciada en Traducció i Interpretació (francès i alemany) i Lingüística i Noves Tecnologies. Ha col·laborat durant un any amb Linguamón - Casa de les Llengües en qualitat de becària. Actualment està a punt de començar l'especialització de Polítiques europees de l'Institut Internacional de Postgrau de la UOC.