

Pràctiques d'Informàtica Aplicada a la Traducció*

Mikel L. Forcada, Carlos Pérez Sancho,
Raül Canals i Marote, Professor per contractar
Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant, E-03071 Alacant

Curs 2008–2009, versió del 17 de setembre de 2008

Índex

0	Prefaci	2			
0.1	Consideracions generals	2			
0.2	El portafolis electrònic	2			
0.3	Recomanacions	3			
0.4	Avaluació	4			
1	Pràctica 1: Iniciació	5			
1.1	Primera part: Obtenció d'algunes dades sobre el maquinari i la connexió a Internet	5			
1.1.1	El maquinari (o <i>hardware</i>)	5			
1.1.2	La connexió a Internet	5			
1.2	Segona part: creació del vostre portafolis electrònic	5			
1.2.1	Sistema operatiu	5			
1.2.2	Accés al portafolis des de Windows XP o Windows 2000	6			
1.2.3	Accés des de Windows 98	6			
1.2.4	Accés des de Windows Vista	6			
1.2.5	Accés des de GNU/Linux	7			
1.2.6	Canvi de contrasenya	7			
1.2.7	Creació del document mestre	7			
1.2.8	Visualització del portafolis amb el navegador	10			
1.3	Tercera part: creació d'una pàgina de recursos d'interés.	11			
1.3.1	Recerca	11			
1.3.2	Creació del document <code>iatpr1.html</code> :	11			
2	Pràctica 2: Ús avançat del processador de textos	12			
3	Pràctica 3: XML	14			
3.1	Introducció	14			
3.2	Primera part: validació de documents XML usant DTD	15			
3.2.1	Estudi de la DTD	15			
3.2.2	Validació d'un document de fax	15			
3.2.3	Validació de modificacions	16			
3.3	Segona part: traducció d'un text en XML: nou text	16			
3.3.1	La DTD per a textos literaris	16			
3.3.2	Traducció del document XML	17			
4	Pràctica 4: No és mot per mot	18			
4.1	Introducció	18			
4.2	El <i>model zero</i>	18			
4.3	Detecció de desviacions respecte del <i>model zero</i>	19			
4.4	Descobrint millores al <i>model zero</i>	19			
4.5	Estructura de l'informe	20			
4.6	Sistemes que es poden usar en línia	20			
5	Pràctica 5: Avaluació de la traducció automàtica	22			
5.1	Introducció	22			
5.2	Estructura de l'informe	23			
6	Pràctica 6: Memòries de traducció	24			
6.1	Introducció	24			
6.1.1	OmegaT	24			

*© 2008 Universitat d'Alacant. Aquest material pot ser distribuït, copiat i exhibit si els noms dels autors es mostren en els crèdits. Les obres derivades han de distribuir-se sota els mateixos termes de llicència que el treball original. Més detalls: <http://creativecommons.org/licenses/by-sa/3.0/deed.ca>. Podeu demanar els fonts LaTeX a Mikel L. Forcada (mlf@ua.es).

6.2	Esquema de la pràctica	25
6.3	Instal·lació	25
6.3.1	Instal·lació de Java i d’Ome- gaT	25
6.3.2	Instal·lació de l’alineador <i>bi- text2tmx</i>	26
6.4	Creació d’una memòria de traduc- ció a partir d’un text i la seua tra- ducció	26
6.4.1	Descàrrega de fitxers	26
6.4.2	Preparació i conversió dels fitxers	26
6.4.3	Creació de la memòria de traducció	26
6.5	Creació d’un projecte de traducció i traducció d’un text nou	27
6.5.1	Creació d’un nou projecte	27
6.5.2	Col·locació dels fitxers en el directori del projecte	27
6.5.3	Traducció del nou text	28
A	Informe d’autoavaluació (obligatori)	29
B	Part pública del portafolis (opcional)	30
C	Documents XML ben formats	30

0 Prefaci

0.1 Consideracions generals

Aquest document recull els enunciats dels treballs pràctics (de laboratori) del curs i les instruccions per a realitzar-los. L’objectiu és que pugueu disposar d’aquesta documentació per endavant per a preparar-vos per a les sessions de laboratori de l’assignatura. S’ha intentat ser exhaustiu i detallat on era necessari, però en altres llocs també s’ha deixat una miqueta de marge perquè hi exploreu; recordeu que si us perdeu sempre teniu a mà el professorat de l’assignatura (en el laboratori o en hores d’assistència a l’alumnat) per a orientar-vos.

Llegiu sempre l’enunciat de cada treball pràctic abans de la sessió corresponent de laboratori.

En efecte, és molt important que us llegiu molt atentament tota la documentació abans d’anar al laboratori. El 90% de les preguntes que feu en el laboratori o quan veniu al despatx estan resoltes en els textos que se us distribueixen: heu de llegir millor (sobretot si voleu ser professionals de la traducció!).

S’especifica, per a cada treball pràctic, les sessions que s’hi dedicaran, tal com s’indica en la programació general del curs.

0.2 El portafolis electrònic

Cada un o cada una de vosaltres mantindrà un *portafolis electrònic* individual en format de *web*, i ho podrà fer en principi des de qualsevol ordinador connectat a Internet. El portafolis tindrà una part *privada* que només podrà ser vista i editada per vosaltres i pel professorat, i una part *pública*, que podrà veure la resta de l’alumnat i qualsevol persona amb accés a Internet. Dedicarem les primeres sessions de laboratori a aprendre com crear i gestionar el portafolis.

El vostre portafolis electrònic serà una carpeta que residirà en un ordinador del Departament de Llenguatges i Sistemes Informàtics anomenat `cercador.dlsi.ua.es` i a la qual podreu accedir amb un nom (el vostre NIF) i una contrasenya (que podreu canviar vosaltres quan vulgueu) des de qualsevol ordinador connectat a Internet.

Objectiu: El portafolis servirà:

- perquè tragueu el màxim profit possible del treball que feu durant aquest quadrimestre en aquesta assignatura (per exemple, l'esforç de manteniment i organització del portafolis servirà perquè tingueu una visió general i connectada de les diverses parts de l'assignatura);
- per a establir el 50% de la vostra qualificació;
- per a organitzar el vostre treball de manera que sempre el tingueu a mà per si voleu referir-vos-hi en el futur.

Els portafolis desapareixeran al final de cada curs. És convenient que mantingueu una còpia de seguretat del vostre portafolis en algun altre lloc. També pot convenir que sol·liciteu a la Universitat una pàgina personal (<http://www.alu.ua.es/va/>) per a situar-hi una còpia de la part pública del portafolis.

Contingut: El portafolis podrà contenir qualsevol material que creieu que pugua servir-vos per a complir els objectius citats més amunt, però, com a mínim, ha de contenir:

- En el directori principal del portafolis, un *document mestre* en HTML (o en XHTML), amb el nom `index.html`, que ha de contenir enllaços a la resta de documents del portafolis; els enllaços han d'incloure una breu descripció de cada document enllaçat. Tots els documents del portafolis han d'estar en formats oberts, accessibles a tothom (preferiblement HTML (o XHTML), en el cas dels textos); no hi inclogueu documents en formats exclusius de programes comercials (com ara els de determinats processadors de textos).¹
- Un document d'auto-avaluació a mitjan quadrimestre (el 29 de novembre de 2008), que completareu i tornareu a presentar el 13 de gener de 2009 (veure l'apèndix A).

¹Tots els documents HTML (o XHTML) han de ser vàlids segons algun dels estàndards d'HTML existents (HTML 4.01 o XHTML 1.0); podeu usar validadors com ara <http://validator.w3.org> o www.validome.org, o programes com `tidy` (<http://www.w3.org/People/Raggett/tidy/>, per a instal·lar) o *tidy online* (<http://infohound.net/tidy/>, en línia), per a reparar petites desviacions de la validesa.

Exemples de materials que pot contenir opcionalment el portafolis (en la part privada o en la pública):

- Enllaços a documents o recursos externs al portafolis, relacionats amb l'assignatura.
- Les síntesis o esquemes que hàgeu pogut fer mentre completàveu les activitats a casa o estudiàveu un tema concret.
- El vostre *mapa de l'assignatura* on situeu els conceptes i habilitats més importants que heu après i les relacions entre elles.
- Informes corresponents a parts opcionals de pràctiques o a projectes opcionals que haja pogut proposar el professorat o hàgeu decidit dur endavant.
- Pistes i trucs que us han ajudat i que poden ajudar en l'ús de la informàtica a qui es dedique professionalment a la traducció.
- Una descripció *professional* (sincera, és clar) de les vostres habilitats informàtiques, com la que inclouríeu en un *curriculum vitae*.

Durant el curs, el professorat podrà proposar projectes opcionals d'aplicació de la Informàtica a la Traducció (projectes de traducció assistida per l'ordinador, projectes d'avaluació o manteniment de sistemes de traducció automàtica desenvolupats en la Universitat, etc.); vosaltres també podeu fer propostes al professorat en aquest sentit.

0.3 Recomanacions

Noms de fitxer: Respecteu els noms de fitxer que s'hi indiquen en cada pràctica.

Per a evitar els problemes tècnics que es produeixen per les divergències existents entre diferents plataformes i sistemes operatius quant als noms de fitxers i quant als jocs de caràcters i assegurar-vos així que els enllaços del vostre portafolis funcionaran sempre i des de qualsevol ordinador, no useu majúscules, lletres accentuades o modificades de fora del joc de caràcters ASCII ("ç", "ï", "à", etc.) ni espais en blanc en els noms de fitxer que elegiu vosaltres (molts noms de fitxer estan ja especificats en els enunciats de les pràctiques).

Us recomanem que useu només els següents caràcters: les 26 lletres minúscules de la "a" a la "z", les xifres del "0" al "9" i els signes "_", "-" i ".". No s'admetran reclamacions causades per noms de fitxers que no complisquen aquesta recomanació.

Estàndards: Per a garantir l'accés de tothom als documents, s'han de respectar els formats estàndards. Us recomanem que valideu sempre els documents HTML (o XHTML) després de qualsevol canvi. En la primera pràctica us explicarem en què consisteix la validació i com fer-la. No es corregiran documents HTML (o XHTML) no vàlids.

Els perills del doble clic: Teniu tendència a obrir els fitxers fent doble clic sobre l'icona que els representa. Quan feu això, deixeu en mans de la configuració actual del sistema operatiu en aquella màquina concreta la decisió sobre quin serà el programa que obrirà el vostre fitxer. Moltes vegades, no és el programa que voldríeu o esperaríeu i això us crea confusió. Per a obrir un fitxer amb un programa concret, obriu primer el programa, i després, des del programa, obriu el fitxer. Així teniu vosaltres el control de la situació.

0.4 Avaluació

El portafolis serà avaluat dues vegades:

- La primera, parcial, el 29 de novembre de 2008, a partir de les 23.59 hores. Aquesta avaluació serà general i servirà sobretot perquè el professorat valore el vostre treball, llija els vostres primers informes d'autoavaluació i faci recomanacions generals de cara a la segona entrega i a la marxa general del curs. També és molt útil perquè us familiaritzeu amb les condicions tècniques de presentació del portafolis. No tindrà efecte sobre la qualificació. El professorat només avaluarà aquells portafolis que continguen, a més de l'informe d'autoavaluació, almenys totes les pràctiques fetes fins aquell dia al laboratori menys una.
- La segona, definitiva, per a establir la qualificació del vostre portafolis tal com estiga a les 23.59 hores del 13 de gener de 2009.

Si l'assignatura se suspén en febrer i cal repetir el portafolis per a setembre, es tornarà a qualificar el que hàgeu entregat el 4 de juliol de 2009 a les 23.59.

Nota important sobre el procediment de presentació del portafolis: L'únic mètode vàlid de presentació dels treballs pràctics és a través del vostre portafolis web (per exemple, no s'accepta l'enviament de fitxers adjunts a tutories o a missatges electrònics, ni l'entrega de CD-ROMs o DVDs). Sigueu prudentes i no ho deixeu tot per a última hora. No admetrem cap reclamació al respecte.

A més dels que s'especifiquen en cada treball pràctic, aquests són els aspectes generals dels portafolis que es valoraran:

- Que no s'hi mesclen idiomes diferents (tot el portafolis està en el mateix idioma, espanyol o català)
- Que els textos siguin lingüísticament correctes en la llengua elegida²
- Que s'usen les convencions tipogràfiques i de presentació més comunes en publicacions professionals
- Que hi haja el mínim possible d'errades tipogràfiques
- Que l'estructura i la presentació de cada document es corresponguen amb una estructura lògica adequada
- Que el material tinga un to o un estil professional i un nivell de formalitat adequat.
- Que l'aparença (la presentació) dels documents siga clara i legible

Els següents aspectes són obligatoris, i el seu incumpliment comportarà una qualificació de zero en la pràctica corresponent.

- Tots els documents han d'estar en formats oberts (preferiblement HTML o XHTML, però també PDF), excepte quan s'indique expressament una altra cosa.

²Sorprén la quantitat de errades tipogràfiques, faltes d'ortografia i sintaxi i aplicació incorrecta de convencions tipogràfiques, sobretot quan es considera que una part de l'alumnat es dedicarà professionalment a la traducció.

- El portafolis ha de tenir un document mestre anomenat `index.html` situat en la part privada del portafolis web.

1 Pràctica 1: Iniciació

[Sessions L_1 i L_2]

Aquesta pràctica té tres parts. En la primera estudiareu una miqueta el maquinari de l'ordinador i algunes dades bàsiques de la connexió a Internet; en la segona, accedireu al vostre portafolis electrònic en el servidor de pràctiques; en la tercera, generareu el document mestre del portafolis i un altre document corresponent a aquesta primera pràctica. Durant el procés, aprendreu a fer algunes operacions bàsiques amb el sistema operatiu i amb la xarxa i també aprendreu els conceptes d'HTML necessaris per a escriure pàgines senzilles.

1.1 Primera part: Obtenció d'algunes dades sobre el maquinari i la connexió a Internet

1.1.1 El maquinari (o *hardware*)

Enceneu l'ordinador i seguiu les instruccions del professor per a accedir-hi (en alguns ordinadors necessitareu potser una contrassenya basada en el vostre NIF o NIE).

En el sistema operatiu Windows XP, per a veure la informació sobre el sistema operatiu i el maquinari del vostre ordinador haureu d'obrir la finestra *Propietats del Sistema*. Hi podeu accedir fent clic en la icona *Sistema* del *Panel de Control*. Dins d'aquesta finestra trobareu diferents pestanyes. En la primera que veieu (*General*), podeu trobar la versió del sistema operatiu instal·lat, el tipus i velocitat del processador i la quantitat de memòria RAM. Per a veure més informació referent al maquinari, cliqueu en la pestanya *Hardware*, i a continuació en el botó *Administrador de dispositius*. **Aneu amb compte perquè qualsevol modificació que feu en aquesta finestra pot tenir repercussions greus per a l'ordinador.** Ací trobareu una llista amb tots els dispositius instal·lats en l'ordinador. Per a aquesta pràctica us heu de fixar únicament en les unitats de disc i DVD/CD-ROM instal·lades.

En acabar heu de saber aquestes coses sobre l'ordinador: quantes unitats de disquet i de disc dur

té, i de quin tipus i capacitat són; quin processador té i quina és la freqüència del rellotge; quanta memòria RAM té i, si ho descobriu, quins sistemes operatius té instal·lats. Afegireu aquesta descripció més endavant al vostre portafolis.

1.1.2 La connexió a Internet

En aquest apartat obtindrem algunes dades interessants sobre la connexió a Internet com són la velocitat de connexió, la de baixada i la de pujada, i l'adreça IP que estem usant per a la connexió (quan ens connectem a Internet l'adreça que usem no és la del nostre ordinador, que és local, sinó que és la del dispositiu que ens permet la connexió; per exemple, la de l'encaminador o *router*). Amb ajuda d'un buscador web haurem de trobar algun lloc web que mitjançant un test de velocitat ens permeti saber quina és la nostra velocitat de connexió amb aquell lloc web. De la mateixa manera haureu de trobar un lloc web que us permeti esbrinar quina és la vostra IP pública per a eixir a Internet. Heu d'anotar les velocitats de pujada i de baixada, l'adreça IP pública i els llocs web que us han permès obtenir aquesta informació, perquè posteriorment les afegireu al vostre portafolis.

1.2 Segona part: creació del vostre portafolis electrònic

Important: Tingueu sempre una còpia de seguretat del portafolis, per exemple en un disquet o en una memòria USB. En el laboratori podreu preguntar com fer-la.

Aquestes instruccions *no estan completament detallades*: caldrà que us "busqueu la vida" una miqueta. En qualsevol cas, sempre teniu un professor a mà per als dubtes que us puguin sorgir.

1.2.1 Sistema operatiu

Si l'ordinador del laboratori us demana que elegiu un sistema operatiu, elegiu Windows XP. Si l'ordinador us demana que us identifiqueu, seguiu les instruccions del professor.

Totes les pràctiques es poden fer també en GNU/Linux; només cal substituir les indicacions que es fan per a Windows per les anàlogues en GNU/Linux. Si us interessa, consulteu-ho al professorat de l'assignatura.

1.2.2 Accés al portafolis des de Windows XP o Windows 2000

Per a poder treballar amb el vostre portafolis des d'ordinadors amb sistema operatiu Windows 2000 o Windows XP, haureu de buscar la carpeta *Mis sitios de red*, que de vegades està dins de *Mi PC* i d'altres vegades en l'escriptori. Dins d'aquesta carpeta heu de fer clic sobre la icona *Agregar sitios de red*. En el quadre de diàleg que us apareixerà heu d'escriure l'adreça (l'URI³) de la vostra carpeta *web*, que té la forma `https://cercador.dlsi.ua.es/iat/X/` on *X* és el vostre NIF o NIE sense la lletra final. Una vegada escrit, heu de fer clic sobre *Siguiente*. Us apareixerà un altre quadre de diàleg on haureu d'escriure la contrasenya, que inicialment serà idèntica al vostre NIF o NIE sense la lletra del final; assegureu-vos que deixeu sense marcar la casella *Guardar esta contraseña...* i cliqueu en *Aceptar*. Us apareixerà un tercer quadre de diàleg on podreu canviar el nom de la carpeta (recurs) si voleu. Cliqueu sobre *Finalizar* i ja està. A partir d'ara, i en l'ordinador on hàgeu fet això, la carpeta web us apareixerà amb aquest nom dins de la carpeta *Mis sitios de red*, i podreu treballar-hi com si fóra una carpeta local.

Inicialment el vostre portafolis està buit. Si dins del vostre portafolis teniu creada o creeu una carpeta que es diga *pub*, el que aquesta carpeta continga podrà ser vist públicament en l'adreça `http://cercador.dlsi.ua.es/pubX`

1.2.3 Accés des de Windows 98

Si hi accediu des d'ordinadors que tinguen Windows 98, el procediment és lleugerament diferent: cal crear en *Mi PC* una *carpeta web*.

1.2.4 Accés des de Windows Vista

Hi ha alguns problemes per a accedir als portafolis a través de Windows Vista perquè sembla que el *component* necessari (WebDAV o WebDAVS) no està instal·lat.

Solució mitjançant instal·lació d'un nou component: El primer que heu de fer⁴ és instal·lar-

vos una actualització per a poder usar les carpetes web des de `http://www.webdavsystem.com/server/access/windows`. L'actualització està enllaçada al principi de la pàgina pàgina com *Software Update for Web Folders*⁵ (la versió en espanyol no val per a Vista, o almenys, no està en la llista de sistemes operatius).

Després:

1. Arranqueu l'Explorador de Windows
2. Feu clic amb el botó dret damunt d'on diu *Equipo* (en anglés *Computer*)
3. Useu l'opció *Añadir ubicación de red* (en anglés *Add a Network Location*). No us recomanem que useu l'opció *Map Network Drive...*, perquè en proves fetes pels nostres tècnics no ha funcionat.
4. Quan us demane la URL escriviu: `http://cercador.dlsi.ua.es/iat/X` on *X* és el vostre NIF o NIE.
5. Us demanarà usuari/contrasenya i ja està.

Finalment, heu de tenir amb compte, abans d'apagar l'ordinador d'esborrar l'entrada que us crea per a *cercador...* en l'Explorador perquè ningú pugui fer res amb els vostres documents.

Nota: aquesta solució només permet l'accés a través de WebDAV no segur (`http:`) (en Windows XP podem usar WebDAV segur (`https:`)).

Instal·lant un programari comercial: La única solució que hem provat i que funciona és usar un programari (desgraciadament privatiu i de pagament) anomenat Bitkinex (`http://www.bitkinex.com/`). Té una versió d'avaluació per a 30 dies.

Altres solucions: Pot ser interessant provar les solucions descrites en aquests documents:

- http://www.yale.edu/its/software/win/configuration_instructions/WebDAV-Vista.pdf

³*Uniform Resource Identifier* o identificador uniforme de recursos

⁴Text preparat amb el suport de M.A. Varó, tècnic del DLSI

⁵URI directe: <http://www.microsoft.com/downloads/details.aspx?familyid=17C36612-632E-4C04-9382-987622ED1D64&displaylang=en>.

- http://www.webdavsystem.com/server/access/map_drive

canviant l'adreça per les que usem en les nostres pràctiques, però no ho les hem pogut provar.

1.2.5 Accés des de GNU/Linux

Escriptori Gnome: Aneu al menú *Llocs* de l'escriptori, i seleccioneu l'opció *Connecta al servidor*. En *Tipus de Servei*, elegiu "Web-DAV segur (HTTPS)". En *Servidor* teclegeu "cercador.dlsi.ua.es". En *directori* teclegeu "iat/X/", on *X* és el vostre NIE o NIF. El *nom d'usuari* ha de ser el vostre NIE o NIF i en *Nom a utilitzar per a la connexió* podeu posar, per exemple "Portafolis IAT". Cliqueu sobre *Connecta*. Es crearà una carpeta web amb aquest nom en el vostre escriptori. Quan hi vulgueu accedir, feu-hi doble clic. El sistema us demanarà la contrasenya si la connexió té èxit, i l'obrirà. Podreu usar aquesta carpeta tantes vegades com vulgueu connectar-vos-hi.

Escriptori KDE: Potser el més senzill és llançar el navegador Konqueror i obrir l'adreça `webdavs://cercador.dlsi.ua.es/iat/\ (X\)`, on *X* és el vostre NIE o NIF.

1.2.6 Canvi de contrasenya

És obligatori que canvieu la contrasenya per una altra que no sàpia ningú. Per a fer-ho, heu de connectar-vos a la pàgina `https://cercador.dlsi.ua.es/iat/`, i elegir l'opció *Canviar contrasenya*.

1.2.7 Creació del document mestre

El *document mestre* serà la pàgina de presentació del vostre portafolis i des d'on s'enllaçaran la resta de pràctiques i documents que heu de presentar. Aquest document s'anomenarà `index.html`, estarà escrit en HTML i residirà en la carpeta arrel (en el directori arrel) del vostre portafolis. Per a crear-lo en aquesta pràctica usarem el redactor del programa Mozilla.

Poseu en marxa el programa *Mozilla*.⁶ En el menú *Finestra (window)*, elegiu l'opció *Redactor (composer)*. Us obrirà una pàgina en blanc que podeu editar. No hi escrigueu res encara.

Com que un dels objectius de l'assignatura és que aprengueu HTML, elegiu de les pestanyes de la part inferior esquerra la que diu *Font HTML* o *HTML source*. Veureu que el document realment no està en blanc. La primera línia, que comença amb "`<!DOCTYPE`" declara que el document és un document HTML estàndard segons la versió 4.01 *transicional* de HTML. Veureu que el redactor no us deixa modificar aquesta línia.

En la segona línia, l'etiqueta "`<html>`" indica el començament del document HTML, i l'etiqueta "`</html>`" del final indica el final del document. En HTML, i en molts altres formats, les *etiquetes* que indiquen el començament d'un *element* de tipus *nom* del document tenen la forma "`<nom...>`" i les etiquetes que indiquen el final de l'element tenen la forma "`</nom>`".

Dins de l'element `html` trobem dos elements: `head` (l'*encapçalament*) i `body` (el *cos* del document). Fixeu-vos on s'obrin i on es tanquen i examineu el que contenen.

Dins de l'encapçalament, un element *meta* (especial, perquè s'obri però no es tanca, ja que no té contingut) indica, a través de dues assignacions del tipus *atribut="valor"*, que el *joc de caràcters* que usa el document és l'ISO-8859-1, el més comú a Europa occidental (per exemple, per a escriure documents en *txec* o en *coreà*, caldria canviar part del valor de l'atribut `content`). Dins de `head` també trobem l'element `title`; inicialment buit, on heu d'escriure, entre "`<title>`" i "`</title>`" el títol del vostre document (per exemple, "Portafolis IAT de *Nom Cognom Cognom*"). Feu-ho. Aquest títol es presentarà, quan obriu el document amb un navegador, en la barra del navegador, però *no com a part del text del document*.

Dins de `body` només hi ha un element "`
`" (un d'aquells que no es tanquen), que indica un salt (*break*) de línia. El primer que farem és llevar aquest salt de línia, esborrant-lo, de manera que l'element `body` quede buit.

Ara, el primer que inserirem entre "`<body>`"

⁶Si no el teniu instal·lat a l'ordinador, el podeu descarregar de <http://www.softcatala.org> (Mozilla 1.7.12 per a Windows). Mozilla ha canviat recentment de nom i a partir d'ara s'anomenarà *SeaMonkey*.

i “</body>” és un *títol de secció de nivell 1*, per a posar un títol que sí que forme part del text del document; per exemple, escriurem “<h1>Portafolis IAT de...</h1>” com hem fet abans en *title*.⁷ Darrere d’aquest element, inseriu un petit paràgraf de prova; entre “<p>” i “</p>”, escriviu, per exemple “Aquest és el meu portafolis per a l’assignatura Informàtica Aplicada a la Traducció.”. Podeu anar mirant com queda l’aparença del vostre document usant la pestanya *Normal*, on també podeu escriure i fer correccions: fixeu-vos en la grandària i el tipus de lletra que el programa elegit automàticament per als elements *h1* i *p*.

Quan torneu a la vista *Font HTML*, veureu que els caràcters accentuats han estat substituïts per *entitats* del tipus “&nom;”: per exemple, la ó és “ó”. Aquestes entitats asseguren que el caràcter que representen es veurà correctament independentment del joc de caràcters seleccionat.

HTML conté alguns elements que serveixen per a marcar parts del text com text amb èmfasi (*em*), text amb èmfasi fort (*strong*), codi (*code*), etc.. Els navegadors assignen tipus de lletra diferents a cada un d’ells. Per exemple, en la vista *Font HTML* enfatitzeu el mot *portafolis* del paràgraf posant-lo entre “” i “”. També podeu assignar al text aquestes característiques en la vista *Normal* a través del menú *Format, Estil de text*: només cal seleccionar una part del text i elegir l’estil que volem que tinga.⁸

És millor usar aquests estils genèrics (*èmfasi, èmfasi més forta*, etc.) que les característiques visuals (*negreta* ..., *cursiva* <i>...</i>, etc.), ja que aquestes últimes només són útils per a les persones vidents i a més no indiquen què és cada part del document, sinó només com es presenta.⁹

Un altre element que us pot ser útil és *ul*, que indica una llista sense numerar (*unnumbered list*). Darrere del paràgraf, en la vista *Text HTML*, inseriu un “”...“” i dins, dos o tres

⁷Anàlogament, *h2*, *h3*, etc. són els títols de secció de nivell 2, 3, etc.

⁸És interessant anar canviant de la vista *Normal* a la vista *Font HTML* per a veure l’efecte d’editar l’HTML directament o per a observar l’HTML que es genera quan es modifica el text en el mode *Normal*.

⁹Si us interessa, el professorat us pot explicar com es pot canviar la presentació de cada tipus d’element usant fulls d’estil.

“”...“”; dins de cada *li* podeu posar text. Fixeu-vos en la vista *Normal* i veureu com es presenta la llista. Si canvieu els *ul* per *ol* (*ordered list*, llista ordenada) veureu una llista numerada automàticament (aquestes proves les haureu d’eliminar del portafolis definitiu quan l’entregueu).

Abans de continuar, guardeu el document en l’escriptori amb el nom *index.html*. Ara en el menú de *Tasques*, elegiu *Validar HTML*. Us obrirà una finestra del navegador que apuntarà a la pàgina <http://validator.w3.org/file-upload.html>, la pàgina del *World Wide Web Consortium* que us diu si la pàgina és vàlida segons els estàndards internacionals establerts per a HTML. Cliqueu sobre el botó que hi ha a la dreta de *File: (Browse, Navegar, etc.)* i busqueu el vostre fitxer *index.html* en l’escriptori. Quan l’hàgeu seleccionat, cliqueu sobre el botó *Validate this file*. El validador us dirà si la pàgina és vàlida, és a dir, si, com es declara en el document, és HTML 4.01 *transicional*. Si no ho és, us dirà en quina part del fitxer HTML que heu escrit hi ha un error. Podreu corregir la pàgina, guardar-la i tornar a validar-la.

Encara que no estigues usant Mozilla sempre podeu usar qualsevol navegador per a connectar-vos als validadors en línia com ara <http://validator.w3.org> o <http://www.validome.org> i validar els vostres documents HTML, tant els que teniu al vostre disc com els que ja estan publicats en el portafolis (en aquest últim cas el programa validador us demanarà el vostre nom d’usuari i contrasenya).

Un altre element molt important, a, permet fer (hiper)enllaços a altres documents. Per exemple, si voleu escriure el text “Un buscador molt interessant és Google.”, on la paraula *Google* aparega destacada i funcione com un enllaç a l’adreça (o URI) <http://google.com/intl/ca/>, escriureu: Un buscador molt interessant és Google.; és a dir, en l’element *a*, l’atribut *href* indica l’adreça i el *contingut* de l’element, el text de l’enllaç. Proveu si voleu.¹⁰

Si voleu afegir una imatge al vostre document, podeu fer-ho amb l’element buit *img*, amb un atribut *src* que indicarà l’adreça (l’URI) de la imatge.

¹⁰Aquest enllaç de prova l’eliminaréu més endavant quan entregueu el portafolis

Taula 1: Alguns elements bàsics d'HTML

Element	Descripció	Més informació
<html>...</html>	Conté tot el document	
<head>...</head>	Encapçalament	
<body>...</body>	Cos	
<meta...>	Informació sobre el document	L'element és buit
<title>...</title>	Conté el títol del document	
 	Salt de línia forçat	L'element és buit.
<h1>...</h1>	Encapçalament de nivell 1	
<h2>...</h2>	Encapçalament de nivell 2	
...
<h6>...</h6>	Encapçalament de nivell 6	
<p>...</p>	Paràgraf	El tancament "</p>" és opcional
...	Llista sense numerar	Conté elements <code>li</code>
...	Llista numerada	Conté elements <code>li</code>
...	Element de llista	Pot contenir una altra llista en el seu interior. El tancament "" és opcional.
...	Èmfasi	
...	Èmfasi fort	
<code>...</code>	Exemple de codi	
<a...>...	"Àncora"	Si porta un atribut <code>href="URI"</code> , el text entre "<a...>" i "" funciona com un enllaç al document que hi ha en l'URI.
<img...>	Imatge	L'atribut <code>src="URI"</code> indica l'adreça on és la imatge. L'atribut <code>alt="text"</code> descriu la imatge amb paraules. L'element és buit.

ge i un altre atribut `alt` que la descriu per a qui no puga veure l'imatge (invidents, navegadors no gràfics, etc.). Cal que el vostre document mestre continga una foto vostra: connecteu-vos al campus virtual amb Mozilla, i quan tingueu la foto de la vostra fitxa a la vista, cliqueu-hi amb el botó secundari, guardeu-la en l'escriptori, per exemple amb el nom `foto.jpg` i arrossegueu-la al portafolis. Després, poseu en el portafolis ``; on poseu aquest enllaç apareixerà la vostra foto.¹¹

1.2.8 Visualització del portafolis amb el navegador

Podeu accedir als documents que hi ha al nostre portafolis usant un programa navegador. Anem a comprovar-ho.

Copieu el document mestre (`index.html`) i el document que conté la vostra foto al vostre portafolis. Una vegada, i amb l'única finalitat que comproveu com es visualitzen els documents al navegador, canvieu el nom del document mestre a `mestre.html` (en lloc d'`index.html`). Ara, amb el navegador, accediu al vostre portafolis, `https://cercador.dlsi.ua.es/iat/X` i comproveu com es veu la llista de documents: hi veureu (`mestre.html` i el document corresponent a la vostra foto. Comproveu que podeu accedir a aquests documents fent-hi clic.

Ara, restablireu el nom original del document mestre, `index.html` (des del navegador només podem veure documents però no podem modificar-los, així que per tal de dur a terme aquest canvi de nom haureu d'usar la carpeta d'accés al portafolis que heu creat anteriorment). Quan hàgeu restablert el nom del document mestre, actualitzeu la pàgina del navegador corresponent al vostre portafolis i veureu com ara ja no us mostra els documents que hi ha, sinó que directament us mostra el document mestre. Això no vol dir que els altres documents hagen desaparegut: el document amb la vostra foto continua allà i podem accedir-hi simplement escrivint el seu nom a continuació de l'adreça web que ens mostra el navegador. Proveu-ho.

¹¹Podeu posar-hi una altra foto vostra si us agrada més.

Reparació de documents HTML invàlids amb

Tidy: Podeu intentar reparar un document HTML que s'ha tornat invàlid usant el programa *tidy* que està en línia en `http://www.infohound.net/tidy/`. Cal que envieu (*Upload*) el vostre document HTML invàlid, elegiu les opcions correctes (per exemple, elegint `latin1` en *Char encoding* o desactivant l'opció *Output XHTML* i cliqueu sobre la fletxa que diu *Tidy!* i en la nova pantalla cliqueu (si no hi ha cap error sever) sobre *Download Tidied File* i guardeu la pàgina amb el mateix nom i en el mateix directori.

Requisits tècnics que ha de complir aquest treball pràctic per a ser avaluat (si no es compleix algun dels requisits, la qualificació serà zero):

- El document mestre es diu `index.html` i resideix al directori arrel del portafolis de manera que és directament accessible a través de l'adreça `https://cercador.dlsi.ua.es/iat/X` on *X* és el NIF o el NIE sense lletra final.
- Hi ha, a la carpeta arrel del vostre portafolis, una carpeta anomenada `maquina`, que conté un document HTML anomenat `maquina.html` on apareix la informació que heu trobat a la primera part de la pràctica. Aquest document està correctament enllaçat des del document mestre.
- El document mestre, i el document `texttmaquina.html` són vàlids.

Aspectes que es valoraran en aquest treball pràctic:

- Tots els enllaços són correctes (funcionen i indiquen clarament en què consisteix el document que enllacen)
- Hi ha enllaços als documents corresponents a cada un dels treballs pràctics
- Conté foto o justifica per què no n'hi ha
- Conté la secció que descriu l'ordinador usat en la primera sessió.
- Usa adequadament les marques HTML

- No conté continguts de prova procedents del desenvolupament de la pràctica (llista de prova, enllaç de prova a Google, etc.).
- El contingut és adequat

1.3 Tercera part: creació d'una pàgina de recursos d'interés.

En aquesta part de la pràctica:

1. Us familiaritzareu amb l'ús d'un programa navegador per a visitar servidors d'Internet i per a buscar documents i recursos d'interés per als traductors.
2. Creareu un document HTML (una *pàgina web*) que anomenareu `iatpr1.html` on anunciareu els recursos que heu trobat. La pàgina contindrà enllaços actius, de manera que, clicant-hi, es pugui arribar al recurs corresponent.
3. Quan el tingueu llest, publicareu en el vostre portafolis *web*, i, si voleu, en la part pública d'aquest, el document que heu creat.

1.3.1 Recerca

Heu de localitzar en Internet almenys tres documents *relacionats amb un aspecte concret de la traducció o la interpretació* que siguin d'interés per a vosaltres. Useu el menú de *favorits* del navegador per a guardar-los de manera que després hi pugueu accedir fàcilment quan estigueu editant la pàgina.

Heus ací els URI d'alguns buscadors:¹²

- Google (<http://www.google.com>)
- Alltheweb (<http://www.alltheweb.com>)
- Altavista (<http://www.altavista.com>)

En principi, els buscadors es poden configurar perquè es presenten en altres idiomes.

La recerca més típica consisteix a buscar documents que continguin determinats mots. Però els buscadors d'Internet no només permeten buscar documents que continguin uns mots concrets,

¹²Recordeu que la majoria dels navegadors sobreentenen que el protocol d'accés és el dels hipertextos (`http://`) i que, per tant, no cal teclejar aquest prefix quan es tecleja un URI.

sinó que també permeten buscar documents que continguin grups de mots consecutius si els posem *entre cometes*; alguns fins i tot permeten restringir l'idioma dels documents buscats. El resultat de la recerca sol indicar també quants documents s'han trobat.

Ha de quedar clar que els buscadors realment *no busquen* documents en Internet sinó que consulten *índexs* que han anat construint a partir de documents visitats anteriorment. Per tant, pot haver-hi documents en Internet que els buscadors no troben perquè mai no els han visitat.

Hi ha alguns tipus de recerques que poden ser d'utilitat per a traductors: per exemple, si no sabem quina és la preposició anglesa més natural per a traduir "ballar a la llum del sol", podem buscar "dance *by* the sunlight", "dance *in* the sunlight", "dance *at* the sunlight", etc. i elegir la que apareixi més vegades en Internet (alerta: això dóna prioritat a l'ús front a la *norma*, és clar).

1.3.2 Creació del document `iatpr1.html`:

De la mateixa manera que heu creat el document mestre `index.html` d'abans, creeu ara un petit document, bàsicament una llista de recursos, en el qual:

- Els títols (tant el de l'encapçalament `<title>...</title>` com el del cos del document `<h1>...</h1>`) indiquen el contingut del document.
- Els enllaços formen part del vostre text on es descriuen els recursos, estiguen descrits breument, i permeten accedir als recursos amb un clic.
- Cada recurs es descriu en un paràgraf contingut en un element `...` d'una llista numerada `...` o no numerada `...`.

Guardau aquest document en l'escriptori amb el nom `iatpr1.html` i arrossegueu-lo al vostre portafolis. Després, obriu des del *redactor* de Mozilla el document mestre `index.html` del vostre portafolis i, en la secció (probablement una secció `h2`) de *Treballs pràctics*, poseu un text que expliqui què és `iatpr1.html` i que continga un enllaç `<a...>...` a aquesta pràctica. Com que

el fitxer `iatpr1.html` estarà en la mateixa carpeta que el document mestre, l'URI que es dona en l'atribut `href="..."` pot ser *relatiu* a la carpeta on resideix el document actual, és a dir, pot ser simplement `href="iatpr1.html"`.

Recomanació important sobre enllaços i estructura del portafolis: Quan en els enllaços feu referència als documents del vostre portafolis, useu sempre enllaços *relatius*. D'aquesta manera, els enllaços funcionaran exactament igual des del disquet, des del disc dur o des del vostre portafolis *web*. Perquè els vostres enllaços relatius siguin senzills (un simple nom de fitxer), us suggerim que tots els documents que s'enllacen des del document mestre estiguen en la carpeta arrel del vostre portafolis, és a dir, que no feu subcarpetes si no és estrictament necessari.

Requisits tècnics que ha de complir aquest treball pràctic per a ser avaluat (si no es compleix algun dels requisits, la qualificació serà *zero*):

- La pràctica és accessible des del document mestre
- El document és vàlid
- El document s'anomena `iatpr1.html` i és directament accessible a través de l'URI `https://cercador.dlsi.ua.es/iat/X/iatpr1.html` on *X* és el NIF o el NIE sense la lletra final.

Aspectes que es valoraran en aquest treball pràctic:

- El contingut és adequat: s'explica què tenen en comú els recursos anunciats.
- Els enllaços són correctes (funcionen) i estan clarament descrits. Els enllaços formen naturalment part del text de la descripció (és a dir, el text del enllaç forma part de la resta del text i es pot llegir tot junt).
- S'hi usen adequadament les marques HTML explicades

2 Pràctica 2: Ús avançat del processador de textos

[Sessió *L₃*]

En aquesta pràctica aprendreu a formatar un text amb OpenOffice (2.0 o superior) usant estils en comptes de treballar directament amb les característiques visuals. Per a la realització de la pràctica necessitareu dos fitxers. El primer s'anomenarà `lutherking.pdf` i el segon `lutherking_net.odt`. Els podeu descarregar de la secció de pràctiques de la pàgina de l'assignatura. Com podreu observar, `lutherking_net.odt` conté el mateix text que `lutherking.pdf` —a excepció d'uns pocs mots que haureu d'escriure vosaltres— però sense formatatge visual. L'objectiu de la pràctica és que modifiqueu el fitxer `lutherking_net.odt` aplicant estils, fins que l'aparença siga idèntica a la de `lutherking.pdf`.

Per evitar la temptació de modificar directament les característiques visuals del text, us recomanem d'anar al menú *Visualització*, elegir l'opció *Barra d'eines*, i desactivar les barres anomenades *Estàndard* i *Formatatge*.

Instruccions: Per a produir el document final, haureu d'usar, modificar o crear estils de pàgina, de paràgraf i de caràcter, a més d'inserir salts de pàgina i un índex.

Els salts de pàgina serveixen per a canviar de pàgina en el lloc que vulguem, així podem escriure fins a mitjan pàgina i saltar de pàgina deixant la resta de la pàgina buida. Ho farem amb l'opció de menú *Insereix* → *Salt manual...* Hi ha un salt de pàgina en la primera pàgina i un altre en la quarta.

Per utilitzar els estils triarem l'opció de menú *Format* → *Estils i formatació*, així accedirem a la finestra que ens permetrà triar d'entre les diferents llistes d'estils —de paràgraf, de caràcter, etc.— l'estil que desitgem tot situant-nos o seleccionant el text on volem aplicar l'estil i fent clic sobre el nom que identifica l'estil. També podrem modificar les característiques dels estils fent-hi clic amb el botó dret i triant l'opció *Modifica...*, i crear nous estils a partir d'un altre fent-hi clic amb el botó dret i triant *Nou...* En el nostre cas crearem estils nous quan siga necessari a partir de l'estil *Per defecte*.

La primera pàgina pertany a l'estil de pàgina anomenat *Primera pàgina*, el paràgraf *Tinc un som-*

ni pertany a l'estil *Títol*, el paràgraf *Martin Luther King Jr.* a l'estil *autor*, que heu de crear vosaltres, i el paràgraf *Traducció del discurs[...]* a l'estil *Subtítol*, a més, conté un índex generat automàticament. A continuació de l'índex hi ha un salt de pàgina.

L'índex s'insereix amb l'opció de menú *Insereix* → *Índexs i taules* → *Índexs i taules...* Una volta generat, podem actualitzar-lo fent-hi clic amb el botó dret del ratolí i elegint l'opció *Actualitza l'índex o la taula*.

La resta de pàgines pertanyen a l'estil de pàgina *pàgina interior*. Aquest estil l'heu de crear vosaltres a partir de l'estil de pàgina *per defecte* i ha de contenir l'encapçalament i el peu de pàgina que veieu al PDF.

El paràgraf *Tinc un somni* de la segona pàgina pertany a l'estil de paràgraf anomenat *Títol*.

Els paràgrafs que comencen amb un numeral seguit d'un punt pertanyen a l'estil *Encapçalament 1*, i els que comencen amb dos numerals separats per un punt pertanyen a l'estil *Encapçalament 2*. El text d'aquests paràgrafs l'heu d'escriure vosaltres perquè no apareix al document de text que us hem facilitat. La numeració no l'heu d'escriure, farem que hi aparega automàticament. Per fer això triarem l'opció de menú *Eines* → *Numeració d'esquemes...*, assignarem al nivell 1 l'estil de paràgraf *Encapçalament 1*, li assignarem numeració, i després de la numeració, com a separador, un punt.¹³ De la mateixa manera assignarem al nivell 2 l'estil *Encapçalament 2*, amb numeració, amb dos subnivells i amb un punt seguit d'un espai en blanc a continuació del número.

L'encapçalament pertany a l'estil *Capçalera*, i el peu de pàgina a l'estil *Peu de pàgina*. La resta de paràgrafs del text, sense excepció, pertanyen a l'estil *Cos del text*. A més de pertànyer a l'estil de paràgraf *Cos del text*, els paràgrafs que comencen amb *Tinc un somni* o amb *Tinc el somni* pertanyen a l'estil de llista *llista somnis*, aquest estil l'heu de crear vosaltres a partir de l'estil *Llista 1*.

Heu d'escriure el vostre nom a l'encapçalament.

La numeració de les pàgines es posa de manera automàtica amb l'opció de menú *Insereix / Camps*, tant el número de pàgina com el recompte total de pàgines.

¹³En algunes instal·lacions d'OpenOffice.org, potser hàgeu de posar un punt seguit d'un espai en blanc.

A més d'això, les frases que apareixen en negreta i cursiva pertanyen a l'estil de caràcter *Èmfasi*. Es tracta de *No, no, no estem satisfets...*, *Tinc el somni que els meus quatre fills...* i *els nens negres i les nenes negres...*

Els estils de pàgina tenen les característiques següents:

- *Primera pàgina*: no conté encapçalament ni peu de pàgina.
- *Pàgina interior*: conté encapçalament i peu de pàgina.

Els estils de paràgraf tenen les característiques següents:

- *Títol*: lletra Arial, negreta, de 24 punts.
- *Encapçalament 1*: Arial, negreta, 115
- *Encapçalament 2*: Arial, negreta, cursiva, de 14 punts.
- *Cos del text*: Times New Roman, 12 punts.
- *Capçalera*: Times New Roman, 12 punts.
- *Peu de pàgina*: Times New Roman, 12 punts.

Els estils de caràcter tenen les característiques següents:

- *Èmfasi*: Times New Roman, negreta, cursiva, 12 punts.

La resta de característiques es veuen clarament al model de document en format PDF aportat pel professor, i totes elles han d'estar regulades necessàriament a través dels estils aplicats (a excepció dels salts de línia).

Heu d'escriure el vostre nom en la fitxa de propietats del document, *Fitxer* → *Propietats...* → *Descripció*.

El document resultat del vostre treball s'anomenarà *lutherking.odt*. Haureu de generar un document PDF equivalent, que s'anomenarà *lutherking.pdf*. A més, haureu de produir un altre document d'OpenOffice.org anomenat *lutherking2.odt* que tindrà el mateix contingut i estils que *lutherking.odt* però una aparença significativament diferent, obtinguda fent només canvis als estils, i generareu el document PDF corresponent, *lutherking2.pdf*.

Instal·lació d'OpenOffice.org: Si l'OpenOffice.org no està instal·lat en el vostre ordinador, podeu instal·lar-lo a partir de la *web* <http://www.softcatala.org>, secció d'*Ofimàtica*.

El fitxer és un arxiu executable (en el moment d'escriure aquestes línies s'anomena `OOo_2.4.1.Win32Intel.install.ca.exe`) que s'instal·la automàticament. Si teniu algun problema amb la instal·lació, podeu consultar el professorat de laboratori.

Tipus de lletra: Si la vostra instal·lació d'OpenOffice.org no té algun dels tipus de lletra que es demanen en les instruccions, useu un tipus que sigui aproximadament igual en aparença.

Requisits tècnics que ha de complir aquest treball pràctic per a ser avaluat (si no es compleix algun dels requisits, la qualificació serà zero):

- La pràctica és accessible des del document mestre
- Es presenta tant en format d'OpenOffice.org (extensió `.odt`) com en format PDF (`.pdf`).
- Els documents s'anomenen `lutherking.odt`, `lutherking2.odt`, `lutherking.pdf` i `lutherking2.pdf` i són directament accessibles a través dels URIs `https://cercador.dlsi.ua.es/iat/X/lutherking.odt`, `https://cercador.dlsi.ua.es/iat/X/lutherking2.odt`, `https://cercador.dlsi.ua.es/iat/X/lutherking.pdf` i `https://cercador.dlsi.ua.es/iat/X/lutherking2.pdf` respectivament, on *X* és el NIF o el NIE sense la lletra final.

Aspectes que es valoraran en aquest treball pràctic:

- S'usen només estils per a regular la presentació (no modifica directament característiques visuals del text).
- No hi ha cap paràgraf en l'estil `Per defecte`
- Els estils estan correctament definits

- Els estils estan correctament emprats
- Es numeren automàticament els apartats, subapartats, etc.
- El nom de l'autor apareix en l'encapçalament i en la fitxa Propietats.

3 Pràctica 3: XML

[Sessions L_4 i L_5]

3.1 Introducció

En aquesta pràctica, que dura dues sessions, us familiaritzareu amb l'ús i la generació de documents marcats seguint l'estàndard XML (*eXtensible Markup Language* o llenguatge extensible de marques). XML permet definir *tipus* o *famílies* de documents usant DTD (*document type definitions* o definicions del tipus de document) que especifiquen l'estructura de marques que poden tenir els documents d'un tipus determinat.¹⁴ En particular, amb un programa validador i una DTD es pot comprovar si un document XML *ben format* (mireu l'apèndix C) és *vàlid*, és a dir, si segueix la definició especificada per la DTD.

La pràctica té dues fases, una per sessió:

1. En la primera fase,
 - (a) estudiareu una DTD molt senzilla que serveix per a definir l'estructura de documents hipotètics de fax com el de l'exemple que apareix en el capítol 4 de les notes de classe;
 - (b) validareu un document de fax d'exemple a través d'una de les webs següents: <http://www.stg.brown.edu/service/xmlvalid/>, <http://www.validome.org> o <http://validator.w3.org>, i
 - (c) canviareu aquest document perquè no complisca la DTD i veurem els missatges d'error que produeix el validador.
2. En la segona fase,

¹⁴Les DTD no són l'única manera d'especificar famílies de documents XML; una altra manera més potent són els anomenats *esquemes XML* (anglès *XML schema*).

- (a) estudiareu breument una DTD que serveix per a etiquetar documents tals com obres literàries,
- (b) traduireu un document curt en XML de manera que la traducció també siga vàlida segons aquesta DTD,
- (c) validareu el vostre etiquetatge enviant el document al programa validador.

5. El NOM, el número NUM i la DATA contenen text sense marques (indicat amb #PCDATA).

```
<!ELEMENT NOM (#PCDATA)>
<!ELEMENT NUM (#PCDATA)>
<!ELEMENT DATA (#PCDATA)>
```

6. El TEXT es compon d'un o més (+) paràgrafs (P).

```
<!ELEMENT TEXT (P)+>
```

7. Finalment, els paràgrafs contenen text.

```
<!ELEMENT P (#PCDATA)>
```

3.2 Primera part: validació de documents XML usant DTD

3.2.1 Estudi de la DTD

La DTD (*fax.dtd*) que defineix documents hipotètics de fax de l'estil del que apareix en les notes de classe la podeu trobar en <http://www.dlsi.ua.es/~mlf/iat/prac3/>. Descarregueu el fitxer (usant *Anomena i desa l'enllaç*), guardeu-lo en el vostre disquet i obriu-lo amb un editor, per exemple amb el *Bloc de Notes* o *Notepad* (el professor us indicarà com accedir-hi). Si el *Bloc de Notes* no troba el fitxer, assegureu-vos d'haver-li indicat que llegirà fitxers de tots els tipus (tipus de fitxer "tots els fitxers", és a dir, *.*). Examinem-la línia a línia:

1. La primera línia és un comentari. Els comentaris comencen amb `<!--` i acaben amb `-->` i es poden situar en qualsevol part d'una DTD.

```
<!-- Exemple de DTD de FAX -->
```

2. Les línies següents defineixen l'estructura del document definint els seus *elements*. La línia

```
<!ELEMENT FAX (DESTINATARI,
  REMITENT?, DATA, TEXT)>
```

defineix l'element *arrel* o *principal*, FAX, i especifica que es compon (en l'ordre indicat) d'un DESTINATARI, un REMITENT opcional (indicat amb ?), d'una DATA i d'un TEXT.

3. El DESTINATARI del fax té dues parts: el NOM (opcional) i el número (NUM).

```
<!ELEMENT DESTINATARI (NOM?, NUM)>
```

4. El remitent es defineix igual:

```
<!ELEMENT REMITENT (NOM?, NUM)>
```

3.2.2 Validació d'un document de fax

Ara validarem el fax (*fax.xml*) que trobareu en <http://www.dlsi.ua.es/~mlf/iat/prac3/>. Descarregueu-lo i guardeu-lo en el mateix directori on heu guardat la DTD. Examinem ara (per exemple, amb el *Bloc de Notes*) el contingut del fitxer:

```
<?xml version="1.0"
encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/css"
  href="fax.css" ?>
<!DOCTYPE FAX SYSTEM
"http://www.dlsi.ua.es/%7Emlf/iat/
prac3/fax.dtd">
<FAX>
<DESTINATARI>
  <NOM>Mikel L. Forcada</NOM>
  <NUM>+34-96-590-9326</NUM>
</DESTINATARI>
<REMITENT>
  <NOM>Letícia Ibaizábal</NOM>
  <NUM>+34-96-999-9999</NUM>
</REMITENT>
<DATA>23 de setembre de 2000</DATA>
<TEXT>
<P>Mikel, m'agradaria que m'enviaves
el text de l'última pràctica d'IAT.
</P>
<P>Ah, i també; l'examen de febrer.
Gràcies.</P>
</TEXT>
</FAX>
```

Fixeu-vos en les primeres línies del fax (el *pròleg* del document):

- La primera línia declara que el document és un document XML de la versió 1.0 i que el joc de caràcters que usa és l'ISO-8859-1 (també anomenat Latin-1).
- La segona línia,

```
<?xml-stylesheet type="text/css"
             href="fax.css" ?>
```

associa al document un *full d'estil* situat en el mateix directori (usant un URI relatiu) i anomenat `fax.css`; el full d'estil, escrit en aquest cas en el llenguatge CSS (anglès *cascaded style sheets*, fulls d'estil en cascada), especifica com s'ha de presentar el document XML en els programes navegadors (*nota*: només alguns dels navegadors més recents, com Mozilla 1.1, Mozilla Firefox, Netscape 6.0, SeaMonkey, etc., entenen acceptablement el llenguatge CSS; els d'altres instal·lacions poden no ser capaços).

- La tercera línia especifica el tipus del document tot indicant d'una banda l'etiqueta arrel o principal del document (`FAX`) i l'URI (`SYSTEM`) on s'ha de buscar la DTD.¹⁵

Copieu-vos també en el disquet el full d'estil `fax.css` que trobareu en <http://www.dlsi.ua.es/~mlf/iat/prac3/> i podreu veure el document XML amb el navegador (si aquest és recent), que el presentarà seguint les directrius del full d'estil (si obriu el full d'estil `fax.css` veureu que el full d'estil CSS té una sintaxi molt senzilla i que està comentat per facilitar-vos-en la comprensió).

Aneu a un dels tres validadors en línia esmentats més amunt i en el formulari corresponent (*Local file* o *Upload*) busqueu en el disquet el fitxer XML que acabeu de guardar i envieu-lo a validar. Com que el document és vàlid perquè segueix l'estructura indicada en la DTD, el validador ens ho confirmarà.

Nota: de vegades hi ha problemes amb algun validador perquè el servidor local no declara correctament el joc de caràcters. Si un dels tres validadors diu que el document `fax.xml` sense modifi-

¹⁵En aquest context el caràcter `'` no és admissible; usem `'%7E'` que és equivalent.

car és vàlid, useu aquest validador per a la resta del treball pràctic.

3.2.3 Validació de modificacions

Modifiqueu amb l'editor el document que teniu guardat en el disc perquè:

- li falte l'etiqueta `</NOM>` al final del nom del remitent o l'etiqueta `</P>` al final d'un paràgraf,
- no tinga número del destinatari (és a dir, li falte l'element `<NUM> ... </NUM>`)
- etc.,

de manera que no siga vàlid, guardeu-lo, i comproveu en cada cas els missatges d'error que dona el validador. (podeu deixar l'editor obert i, en el cas de <http://validator.w3.org> i <http://www.validome.org>, simplement demanar al navegador que recarregue o actualitze la pàgina amb el resultat de la validació anterior).

Feu també canvis que preserven la validesa del document (com ara eliminar completament l'element `<REMITENT>...</REMITENT>` opcional) i mireu el que diu el validador.

3.3 Segona part: traducció d'un text en XML: nou text

3.3.1 La DTD per a textos literaris

La DTD que especifica l'etiquetatge que usarem per als textos literaris la podeu trobar en <http://www.dlsi.ua.es/~mlf/iat/prac3/> i s'anomena `microtei.dtd`. Aquesta DTD és una versió molt simplificada d'una de les DTD que defineixen un esquema d'etiquetatge estàndard d'obres literàries usat en moltes biblioteques digitals, anomenat TEI (anglès *text encoding initiative*¹⁶); els documents XML vàlids segons aquesta DTD també són vàlids segons una de les DTD del TEI. La DTD està comentada perquè pugueu llegir-hi el significat de cadascun dels elements que s'hi defineixen. Guardeu-la i obriu-la amb l'editor (per exemple, amb el *Bloc de Notes*).

Quant a la declaració dels elements, hi ha algunes novetats respecte a l'anterior DTD. La línia

¹⁶<http://www.tei-c.org>

```
<!ELEMENT p (#PCDATA|hi)*>
```

defineix l'element `p` (que representa un paràgraf) i indica que es compon de text sense marques (assenyalat amb `#PCDATA`) o d'elements de tipus `hi` i que tots poden aparèixer en qualsevol ordre i quantitat ("*" vol dir "zero o més vegades" i "|" separa opcions possibles).

D'altra banda, les línies

```
<!ATTLIST language
id ID #IMPLIED >
```

defineixen per a l'element `language` un atribut del tipus IDENTIFICADOR (ID) opcional (`#IMPLIED`) anomenat `lang` per als elements de tipus `text`. Un exemple d'ús d'aquest atribut en un document XML és:

```
<text lang="ca">
...
</text>
```

Fixeu-vos que l'atribut només s'especifica en la marca d'obertura. La definició

```
<!ATTLIST text
lang IDREF #REQUIRED >
```

és molt similar a l'anterior. En aquest cas es defineix `lang` com un atribut de "referència a identificador"¹⁷ (`IDREF`) obligatori (`#REQUIRED`) de l'element `text`. Finalment, la definició

```
<!ATTLIST div
type CDATA #IMPLIED
n CDATA #IMPLIED >
```

Defineix per a l'element `div` dos atributs opcionals `type` i `n`, que són de tipus `text` (`CDATA`).

3.3.2 Traducció del document XML

Guardeu el document XML `microtext.xml` que trobareu en <http://www.dlsi.ua.es/~mlf/iat/prac3/> en el directori `prac3` del disquet amb el nom `text.xml`.

Obriu aquest document amb l'editor. Observareu al principi del text les línies següents

```
<?xml version="1.0"
encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/css"
href="microtei.css"?>
```

¹⁷Fa referència a un atribut definit en altre lloc com a `ID`

```
<!DOCTYPE TEI.2 SYSTEM
"http://www.dlsi.ua.es/%7Emlf/iat/
prac3/microtei.dtd">
```

les quals indiquen

- que el text que escriureu serà un document XML escrit amb el joc de caràcters ISO-8859-1,
- que a aquest document se li assigna, per a permetre'n la presentació en navegadors capaços d'usar-lo, el full d'estil CSS `microtei.css` que es troba en el mateix directori, i
- que l'element arrel del document s'anomenarà `TEI.2` i que la DTD que defineix el tipus de document es pot trobar en el mateix directori `microtei.dtd`.

El text està escrit en català (`<text lang="ca">`), encara que conté una cita en llatí (`<hi lang="la">`) i un parlament en italià (`<q who="Galileo" lang="it">`).

Ara, traduïeu les parts en català del text a l'anglès (`en`), al francès (`fr`) o a l'alemany (`de`), deixant la resta en la llengua original, i prestant atenció a mantenir un etiquetatge ben format i vàlid d'acord amb la DTD. Assegureu-vos que el nou idioma està definit en la secció `langUsage` del preàmbul del document. Canvieu també la descripció dels elements `sourceDesc` i `publicationStmt` perquè estiguen en la llengua meta i expliquen que el document és ara una traducció. Guardeu-lo en el disc. Envieu-lo al validador com heu fet abans amb els documents de fax. És probable que la primera vegada el text no siga vàlid. Useu les indicacions que dóna el validador per a buscar els errors, i corregiu-los fins que el text siga vàlid.

Copieu-vos també en el directori `prac3` el full d'estil `microtei.css` que trobareu en <http://www.dlsi.ua.es/~mlf/iat/prac3/> i podreu veure el document XML amb un navegador recent (si us fixeu, hi ha parts del document que no es presenten amb aquest full d'estil, però heu de traduir-les igualment). Opcionalment, també podeu provar a canviar el full d'estil per a veure com canvia la presentació. En <http://flanagan.ugr.es/xml/css/> podeu trobar una introducció a la creació de fulls d'estil CSS per a documents XML.

Quan hàgeu acabat, guardeu el text correctament etiquetat (`text.xml`) en el disquet i pugeu-lo al portafolis, conjuntament amb el full d'estil `microtei.css` perquè es presente correctament en el navegador.

Requisits tècnics que ha de complir aquest treball pràctic per a ser avaluat (si no es compleix algun dels requisits, la qualificació serà *zero*):

- El text traduït és accessible des del document mestre
- El document traduït s'anomena `text.xml` i és accessible directament a través de l'URL `https://cercador.dlsi.ua.es/iat/X/text.xml`
- El document traduït és vàlid d'acord amb la DTD
- El full d'estil `microtei.css` està correctament enllaçat (és a dir, és en el mateix directori que `text.xml`)

Alguns aspectes que es valoraran del treball pràctic:

- S'hi explica de manera entenedora (en el document mestre al costat de l'enllaç o en un breu document HTML o XHTML vàlid o PDF a banda) el que s'ha fet en la primera part de la pràctica
- S'ha canviat adequadament el contingut de l'atribut `lang` i s'han declarat les llengües en `langUsage`
- S'han traduït o modificat correctament les parts del document que no es presenten
- L'estructura XML és com la del document original
- El document està ben traduït a la llengua meta elegida

4 Pràctica 4: No és mot per mot

[Sessió L_6 i la meitat de la L_7]

4.1 Introducció

Aquesta és una pràctica *oberta*: haureu d'experimentar amb un sistema de traducció automàtica dels que hi ha accessibles per internet i amb un parell de llengües concret¹⁸ (mireu la taula 2) seguint una estratègia ben dissenyada. Per això, haureu de recollir dades, analitzar-les i, finalment, escriure un breu informe que publicareu en el vostre portafolis *web*.

L'**objectiu** és descobrir, estudiar i caracteritzar objectivament algunes de les operacions que fa el sistema a més de substituir cada mot de la llengua origen per un mot equivalent de la llengua meta.

L'**objectiu no és** discutir els errors comesos pel sistema de traducció automàtica (hi ha un altre treball pràctic que aborda aquest aspecte) ni tampoc donar opinions subjectives sobre la utilitat o el futur de la traducció automàtica en la professió.

Segurament el temps en el laboratori no serà suficient per a acabar la pràctica, però sí que servirà per a recollir dades i perquè discutiu amb altres companyes i companys i el professorat us oriente quant a com plantejar-la i realitzar-la. S'entén que necessitareu una miqueta més de temps per a completar i publicar l'informe sol·licitat en el vostre portafolis *web*.

Podeu fer la pràctica individualment o en grups de dues persones; en aquest cas, publiqueu un informe idèntic en els dos portafolis. Sigueu conscients que l'avaluació d'un informe realitzat per dues persones serà òbviament més exigent que la d'una. El professorat podrà formular qualsevol pregunta sobre l'informe a qualsevol de les persones autores per a comprovar la veracitat de l'autoria declarada.

4.2 El model zero

El *model zero* és un model intencionadament rudimentari de traducció automàtica, que servirà com a referència per a estudiar els sistemes reals, que són sempre més avançats. Aquest model tradueix cada mot del text origen aïlladament (sense tenir en compte el context) i el substitueix per un equi-

¹⁸El professorat us demana que en la mesura del possible, la llengua d'arribada siga l'espanyol o el català. En aquesta pràctica no es recomana l'ús del traductor de Google perquè és un sistema de traducció automàtica estadístic, no basat en regles.

valent fix (d'un mot, de més mots, o fins i tot de zero mots), de manera que el text meta és el resultat d'escriure aquests equivalents fixos en el mateix ordre que tenien les paraules del text original. Per exemple, si la frase té N mots,

$$m_1 m_2 m_3 \cdots m_N \quad (1)$$

la traducció *model zero* és

$$T(m_1) T(m_2) T(m_3) \cdots T(m_N) \quad (2)$$

on $T(m)$ representa l'equivalent fix del mot m en la llengua meta, que pot tenir zero, un o més mots. Per exemple, la traducció *model zero* a l'anglès de la frase catalana *Aquest exercici pràctic és molt senzill*, amb $m_1 = \text{Aquest}$, $m_2 = \text{exercici}$, etc., podria ser **This exercise practical it is very simple* (incorrecta), on, per exemple, $T(m_4) = T(\text{és}) = \text{it is}$.

4.3 Detecció de desviacions respecte del *model zero*

Es pot forçar un sistema de traducció automàtica perquè funcione com a *model zero*. Per exemple, escrivint cada mot de la frase en un paràgraf independent (amb una línia en blanc completa entre mot i mot), en alguns casos amb majúscula inicial i acabat en punt, com si formara una frase per ell mateix. La comparació de la traducció de la frase completa amb la traducció *model zero* obtinguda així dona molta informació sobre altres operacions.

Convé realitzar la pràctica creant, mantenint i ampliant un document que continga un corpus de frases numerades de l'estil de

- (1) My friend is Italian
(2) My friends are Italian

una per paràgraf. Per a crear aquest corpus de frases,

- podeu inspirar-vos en els resultats obtinguts per a textos reals, reduint les frases al mínim necessari per a il·lustrar una operació concreta, o
- creant frases senzilles que posen de manifest una operació concreta.

No té sentit que discutiu frases on no heu pogut identificar clarament una operació concreta.

L'informe hauria de contenir totes les oracions que siguin necessàries i suficients per a demostrar l'existència de les operacions descobertes (acompanyades de la traducció mot per mot i la traducció de la frase completa) i, si convé, de regles concretes, numerant cada oració o cada exemple per si cal fer-hi referència una altra vegada en el document.

4.4 Descobrint millores al *model zero*

Hi ha, entre altres, dues formes de descobrir què fa el sistema de traducció automàtica a més del que fa el *model zero*:

- Comparar el resultat de traduir la frase original completa amb l'obtingut amb el *model zero* forçat descrit en l'apartat anterior. Un possible format per a cada exemple numerat seria:

- (1) My friends are Italian
El meu amics són italià.
Els meus amics són italians.

La primera línia és l'oració original, la segona la traducció de cada mot solt (*model zero*) i la tercera és la traducció feta pel sistema de traducció automàtica.

- Trobar parells de frases on canviar només una paraula fa que canvie la traducció d'una altra i provar a entendre per què succeeix ("parells mínims"). En aquest cas podríeu usar un altre format per a representar els exemples:

- (1a) The red **car**
El **cotxe** vermell
(1b) The red **cars**
Els cotxes vermells

Heus ací una llista (no exhaustiva) d'operacions que podeu buscar i documentar amb exemples:

- Concordança de gènere, de nombre, de persona, etc. entre mots contigus o allunyats.
- Canvis de l'ordre de les paraules (reordenaments)

- Resolució d'ambigüitats lèxiques (homografia, polisèmia)
- Processament conjunt d'unitats lèxiques de més d'un mot (expressions multimot): unitats terminològiques, expressions idiomàtiques, col·locacions, etc.

El vostre informe ha de discutir operacions d'almenys tres d'aquests quatre grups.

Si elegiu bé les frases, és possible que fins i tot pugueu descobrir alguns dels mecanismes o de les regles que el sistema aplica. En aquest sentit, cal tenir en compte que les traduccions *incorrectes* donen tanta o més informació sobre aquestes estratègies que les *correctes*. Una vegada identificades algunes operacions importants, podeu estudiar-ne alguna amb més detall.

4.5 Estructura de l'informe

L'informe que publicareu en el vostre portafolis *web* ha tenir un estil objectiu i professional i ha d'estar escrit de manera que una persona exterior a l'assignatura el pugui entendre i reproduir completament el vostre treball, independentment d'aquest enunciat, i seguint en la mesura del possible el format d'un article en una revista especialitzada. Podeu redactar-lo usant OpenOffice.org i exportar-lo com a PDF, però potser és suficient usar un document HTML escrit amb el redactor de Mozilla. Es valorarà molt la *claredat* i la *brevetat* de l'informe; no s'acceptaran informes sense un nivell mínim de correcció lingüística i d'estil, i que no segueixen les convencions tipogràfiques usuals. Si useu OpenOffice.org, heu d'aplicar el que heu après en la pràctica 2 sobre estils.

L'informe ha de recollir almenys les següents informacions:

- Un títol clar i explicatiu, el nom de les persones que l'han realitzat, l'adreça de correu electrònic de cada una d'elles, la data d'escriptura de l'informe, etc.
- L'objectiu d'aquest treball, descrit clarament i amb les vostres paraules.
- Una identificació clara del sistema de traducció automàtica usat (incloent el parell de

llengües) i com s'hi ha d'accedir (URI, etc.).¹⁹

- Una descripció del procés seguit per a generar el conjunt de frases de prova (a partir de textos reals o no).
- Una descripció de la metodologia o de l'estratègia usada per a detectar què fa el sistema a més de substituir cada mot aïlladament per un equivalent constant.
- Una descripció de cada una de les operacions addicionals que heu descobert i estudiat, amb exemples suficients d'observacions per a justificar que realment es fan (les frases usades per a aquesta justificació s'han de posar numerades en el text, preferentment la primera vegada que s'hi fa referència).
- Si podeu, una explicació de la regla o del mecanisme concret que usa el sistema de traducció automàtica en aquests casos.
- Si voleu, comentaris sobre sobre l'eficàcia i les limitacions de les regles que sembla usar el sistema.
- Referències bibliogràfiques, si n'heu usat per a guiar el vostre estudi.

L'informe ha de tenir el nom `iatpr4.html` si és un document HTML, o `iatpr4.pdf` si és un document PDF i ha d'estar enllaçat des del document mestre `index.html` del portafolis.

4.6 Sistemes que es poden usar en línia

La taula 2 (actualitzada en setembre de 2006) dóna una llista dels traductors automàtics que es poden usar, tot indicant el nom del sistema, els parells de llengües disponibles i l'URI.

Requisits tècnics que ha de complir aquest treball pràctic per a ser avaluat (si no es compleix algun dels requisits, la qualificació serà *zero*):

- La pràctica és accessible des del document mestre.

¹⁹Si A i B són dues llengües, el parell $A - B$ és diferent del parell $B - A$. No canvieu de parell durant el treball.

NOM DEL SISTEMA	PARELL(S)	URI
interNOSTRUM	ca-es, es-ca	http://www.internostrum.com
Apertium	ca-en, ca-eo, ca-es, ca-fr, ca-oc, ca-pt, en-ca, en-es, es-ca, es-en, es-eo, es-fr, es-gl, es-oc, es-pt, fr-ca, fr-es, gl-es, gl-pt, oc-ca, pt-ca, pt-es, pt-gl, ro-es	http://apertium.org
Translendum	ca-en, ca-es, ca-fr, de-en, de-es, de-fr, de-ru, en-ca, en-de, en-es, en-fr, en-it, en-ru, es-ca, es-de, es-en, es-fr, fr-ca, fr-de, fr-en, fr-es	http://www.translendum.com/ (fer clic sobre <i>Demo de traducció</i>)
Babelfish	el-en, el-fr, en-de, en-el, en-es, en-fr, en-it, en-ja, en-ko, en-nl, en-pt, en-ru, en-zh, fr-de, fr-el, fr-en, fr-es, fr-it, fr-nl, fr-pt, it-en, it-fr, ja-en, ko-en, nl-en, nl-fr, pt-en, pt-fr, ru-en, zh-en	http://babelfish.altavista.com
Freetranslation	de-en, en-de, en-es, en-fr, en-it, en-ja, en-nl, en-no, en-pt, en-zh, es-en, it-en, ja-en, nl-en, pt-en, ru-en	http://www.freetranslation.com
Reverso	de-en, de-fr, en-de, en-es, en-fr, en-it, en-ru, es-en, es-fr, fr-de, fr-en, fr-es, fr-it, it-en, it-fr, ru-en	http://www.reverso.net
El traductor de Google*	ar-en, de-en, de-fr, en-ar, en-de, en-es, en-fr, en-it, en-ja, en-ko, en-pt, en-zh, es-en, fr-de, fr-en, it-en, ja-en, ko-en, pt-en, zh-en	http://translate.google.com

Taula 2: Alguns traductors automàtics disponibles en línia. Les abreviatures usades per a les llengües són els codis ISO-639 (de dues lletres) corresponents (http://www.loc.gov/standards/iso639-2/php/code_list.php). Aquesta llista pot no estar actualitzada. (*) Google usa traducció automàtica estadística.

- L'informe està en HTML o PDF i s'anomena `iatpr4.html` o `iatpr4.pdf` i és directament accessible a través de l'URI `https://cercador.dlsi.ua.es/iat/X/iatpr4.html` o `https://cercador.dlsi.ua.es/iat/X/iatpr4.pdf` respectivament, on *X* és el NIF o el NIE sense la lletra final.
- Si es fa usant HTML, el document `iatpr4.html` és vàlid.²⁰

Aspectes que es valoraran en aquest treball pràctic:

- L'informe segueix l'esquema de l'enunciat.
- Es compara un hipotètic model zero amb el sistema real.
- La redacció del document és adequada (el treball pot ser entès i reproduït per una persona aliena a l'assígnatura).
- L'anàlisi és adequada.
- Els exemples són clars i senzills, estan numerats i estan situats en el text en el lloc on s'hi fa referència per primera vegada.
- L'informe no inclou informació prescindible (p.ex., valoracions sobre la qualitat de les traduccions).
- S'hi estudien casos interessants

5 Pràctica 5: Avaluació de la traducció automàtica

[segona meitat de la sessió *L*₇ i sessió *L*₈]

5.1 Introducció

En aquesta pràctica elegireu un sistema de traducció automàtica accessible per Internet dels que hem estudiat en la pràctica 4 i un parell de llengües i haureu de fer una *avaluació quantitativa de la qualitat* de les traduccions en brut que produeix per a aquest parell de llengües;²¹ l'avaluació ha de con-

²⁰Atenció: els documents HTML que genera OpenOffice.org poden no ser vàlids.

²¹El professorat us demana que en la mesura del possible, la llengua d'arribada siga l'espanyol o el català.

siderar l'*adopció* del sistema per part d'una empresa o institució per a la *disseminació*, amb postedició (imagineu-vos que esteu assessorant-la en una situació real i concreta). Si hi ha més d'un sistema de traducció automàtica per al parell de llengües en qüestió, també podeu fer una avaluació comparativa entre els sistemes.

Aquesta pràctica, com la pràctica 4, és oberta, en el sentit que sereu vosaltres qui haureu de dissenyar l'estratègia d'avaluació, elegir els textos de manera que siguin suficientment grans i representatius de la tasca elegida, definir la *mesura* o el *barrem quantitativu* que usareu per a puntuar la qualitat, detectar possibles limitacions de la vostra estratègia d'avaluació, etc. Per descomptat, us podeu basar en la informació que s'haja discutit en les classes de teoria i consultar el professorat.

Per això, haureu de recollir dades, analitzar-les i, finalment, escriure un breu informe (en català o en espanyol) que publicareu en el vostre portafolis d'Internet.

Heus ací alguns exemples d'indicadors quantitativus de la qualitat:

- La quantitat *mínima* de mots que cal esborrar, substituir o inserir per convertir la traducció en brut en una traducció adequada per al propòsit previst, expressada com a percentatge del nombre total de mots de la traducció en brut.²²
- El temps necessari per convertir la traducció en brut en una traducció adequada per al propòsit previst, expressat, per exemple, en minuts per cada 1.000 mots del text en brut.

Ja veieu que haureu de fer alguna regla de tres, perquè els textos no seran sempre ni de 100 ni de 1.000 paraules.

Una manera de saber si la quantitat de text que heu avaluat és representativa és estudiar, per exemple, si els indicadors varien molt de la primera meitat del text a la segona meitat. Això també té implicacions a l'hora de presentar els resultats. Per exemple, no té sentit dir que la postedició requereix 52 minuts i 29 segons per cada 1.000 mots (amb tanta exactitud) quan aquest indicador per a la primera meitat del text és 49 minuts i 21 segons i per a la segona meitat és 55 minuts i 37 segons.

²²De vegades, aquest indicador s'anomena *word error rate* o taxa d'error per mot.

Potser fóra més lògic dir “al voltant de 50 minuts cada 1.000 mots”.

Podeu fer la pràctica individualment o en grups de dues persones; en aquest últim cas, publiquen un informe idèntic en els dos portafolis. Siguen conscients que l'avaluació d'un informe realitzat per dues persones serà òbviament més exigent que la d'una. El professorat podrà formular qualsevol pregunta sobre l'informe a qualsevol de les persones autores per a comprovar la veracitat de l'autoria declarada.

5.2 Estructura de l'informe

L'informe que publicareu en el vostre portafolis *web* ha de tenir un estil professional i objectiu i ha d'estar escrit de manera que una persona exterior a l'assignatura el pugui entendre i reproduir completament el vostre treball, independentment d'aquest enunciat, i seguint en la mesura del possible el format d'un article en una revista especialitzada. Podeu redactar-lo usant OpenOffice.org i exportar-lo com a PDF, però potser és suficient usar un document HTML escrit amb el redactor de Mozilla. Es valorarà molt la *claredat* i la *brevetat* de l'informe; no s'acceptaran informes sense un nivell mínim de correcció lingüística i d'estil o que no segueixen les convencions tipogràfiques usals. Si useu OpenOffice.org, heu d'aplicar el que heu après en la pràctica 2 sobre estils.

L'informe ha de recollir almenys les següents informacions:

- Un títol clar i explicatiu, el nom de les persones que l'han realitzat, l'adreça de correu electrònic de cada una d'elles, la data d'escriptura de l'informe, etc.
- L'objectiu d'aquest treball, descrit clarament i amb les vostres paraules.
- Una identificació clara del sistema de traducció automàtica usat (incloent el parell de llengües) i com s'hi ha d'accedir (URI, etc.).²³
- Una identificació clara de la situació concreta de traducció automàtica per a la disseminació que s'aborda.

²³Si *A* i *B* són dues llengües, el parell *A – B* és diferent del parell *B – A*. No canvieu de parell durant el treball.

- Una descripció de la metodologia o de l'estratègia usada per a avaluar el sistema: com s'han elegit els textos, per què, quina mesura o índex de qualitat s'usa per a valorar la qualitat dels textos i com s'ha calculat.
- Limitacions conegudes del sistema d'avaluació que heu dissenyat.
- Resultats de l'avaluació: quantitativs i qualitativs.²⁴
- Conclusions, si n'hi ha.
- Referències bibliogràfiques, si n'heu usat per a guiar el vostre estudi.
- Un apèndix amb els textos originals, les traduccions en brut, i, si n'heu fet, les traduccions corregides o marcades (aquests textos poden ser fitxers a banda, enllaçats i publicats en el portafolis com a documents separats).

L'informe tindrà el nom `iatpr5.html` si és un document HTML, o `iatpr5.pdf` si és un document PDF i haurà d'estar enllaçat des del document mestre `index.html` del portafolis.

Requisits tècnics que ha de complir aquest treball pràctic per a ser avaluat (si no es compleix algun dels requisits, la qualificació serà *zero*):

- La pràctica és accessible des del document mestre.
- L'informe està en HTML o PDF i s'anomena `iatpr5.html` o `iatpr5.pdf` i és directament accessible a través de l'URI `https://cercador.dlsi.ua.es/iat/X/iatpr5.html` o `https://cercador.dlsi.ua.es/iat/X/iatpr5.pdf` on *X* és el NIF o el NIE sense la lletra final.
- S'hi inclouen els textos original, traduït en brut i corregit en format de text pla (`.txt`) i aquests textos estan convenientment enllaçats des del document mestre o des de l'informe.

²⁴Si necessiteu comptar quants mots té un text, podeu usar *OpenOffice.org*: en el menú *Fitxer*, elegiu l'opció *Propietats*: la informació la trobareu en la pestanya *Estadística*.

- Si es fa usant HTML, el document `iatpr5.html` és vàlid²⁵.

Aspectes que es valoraran en aquest treball pràctic:

- L'informe té l'estructura adequada (segons s'indica en l'enunciat).
- El sistema i el parell de llengües estan clarament identificats.
- La metodologia està clarament explicada (de manera que siga reproducible) i justificada.
- La situació concreta d'ús de la traducció automàtica està ben descrita i és creïble.
- S'hi ha fet una avaluació quantitativa de la qualitat.

6 Pràctica 6: Memòries de traducció

[Sessions L_9 , L_{10} i L_{11}]

6.1 Introducció

En els sistemes de traducció automàtica que hem estudiat fins ara, el programa traductor produïa, seguint un conjunt de regles i algorismes basats en coneixements lingüístics, una traducció en llengua meta per a un text escrit en llengua origen. En aquesta pràctica, en canvi, considerarem una altra forma d'utilitzar els ordinadors en tasques de traducció i utilitzarem un programa de traducció humana assistida per ordinador basat en les anomenades memòries de traducció.

Hi ha molts programes comercials de traducció assistida per ordinador. En aquesta pràctica, en canvi, usarem un programa gratuït i de codi obert, anomenat *OmegaT*, el qual pot constituir una alternativa interessant als programes comercials en el vostre treball.

²⁵Atenció: els documents HTML que genera OpenOffice.org poden no ser vàlids.

6.1.1 OmegaT

OmegaT (<http://www.omegat.org>) és un programa de traducció assistida per l'ordinador (basat en memòries de traducció) que té les característiques següents:

- El programa és de lliure distribució i de codi obert. Es pot descarregar del *web* del programa (<http://www.omegat.org/es/downloads.html>). Cal descarregar "OmegaT para Windows sin JRE" (versió actual: `OmegaT_1.7.3_04_Without_JRE.zip`). Les explicacions que es donen en aquest enunciat es refereixen als menús en català. El programa intenta detectar la llengua preferida del sistema operatiu i l'usa.²⁶
- El programa està escrit en Java; això vol dir que es pot usar sobre qualsevol sistema operatiu, sempre que s'hi haja instal·lat l'entorn d'execució de Java (*Java runtime environment*), disponible per a la majoria dels sistemes operatius (s'hi instal·la de vegades conjuntament amb alguns navegadors d'Internet).
- El programa és molt senzill (alguns dirien que rudimentari) i molt menut (es pot instal·lar i portar en una memòria USB).
- El programa és capaç de traduir textos plans en UTF-8 (`.txt`) i ISO-8859-1 (`.txt1`), hipertextos HTML i XHTML (`.html`) i fitxers d'OpenOffice.org 1.1 o superior (`.odt`, `.sxd`). Si es volen traduir documents de Word o d'altres processadors de textos, s'ha de fer a través d'OpenOffice.org.
- El programa genera memòries de traducció en format TMX²⁷ adequades per a altres programes de memòria de traducció.
- El programa usa traduccions de segments ja revisats del text actual per a generar pretraduccions actualitzades dels segments segons es van visitant.

²⁶Per a executar-lo en català podeu teclejar `java -Duser.language=ca -jar=OmegaT.jar` en l'entorn d'ordres del vostre sistema.

²⁷TMX, *translation memory exchange*, és un format estàndard d'intercanvi de memòries de traducció (<http://www.lisa.org/tmx>).

Algunes limitacions d'OmegaT:

- El programa no incorpora un *alineador* de bitextos (assumeix que qui tradueix *comença des de zero* i vol fer més eficient la seua faena usant OmegaT per a evitar traduir text repetit); com a contrapartida, pot llegir fitxers TMX de la versió 1.1 produïts per altres programes. Per a pal·liar aquesta limitació, en aquesta pràctica usarem el programa *bitext2tmx*, també de codi obert, ens pot ajudar a alinear dos textos que són un traducció de l'altre i generar un fitxer TMX 1.1 (vegeu la secció 6.3.2).
- La interfície d'ús és bastant *espartana*, comparada amb la d'altres programes de traducció assistida.
- OmegaT només té un suport molt limitat de consulta de glossaris o terminologia per a millorar les pretraduccions (tan limitat que no el considerarem en aquesta pràctica) i de moment no és capaç de fer substitucions simples de quantitats numèriques quan aquestes són la única diferència entre la frase a traduir i alguna de les frases de la memòria.

Per a saber més coses sobre OmegaT, podeu visitar el *web* del programa <http://www.omegat.org>, que inclou revisions crítiques de diversos autors. També pot ser interessant llegir un article de Marc Prior, disponible en <http://www accurapid.com/journal/23linux.htm>, que parla d'una de les primeres versions del programa.

Reiterem que OmegaT és un programa molt senzill i encara en desenvolupament, i que no es pot comparar amb els programes comercials, encara que us pot ser molt útil. Com es diu en el *web* d'OmegaT,

OmegaT no és especialment difícil d'instal·lar o d'usar però sí que requereix la voluntat de llegir i seguir les instruccions. Si teniu una aversió natural a llegir instruccions, i la vostra aproximació al programari nou és fer clic sobre tots els botons que veieu fins que passa alguna cosa similar a l'efecte desitjat, potser OmegaT no és el que busqueu.

Però com que en l'assignatura l'hem d'usar per a fer aquesta pràctica, *lleugi i seguiu atentament les instruccions*.

6.2 Esquema de la pràctica

En aquesta pràctica, a la qual dedicarem dues sessions, farem les operacions següents, típiques del treball amb memòries de traducció:

1. Instal·larem tot el programari necessari: el programa de traducció assistida OmegaT i l'alineador *bitext2tmx* (apartat 6.3).
2. Alinearem amb *bitext2tmx* un text OpenOffice.org amb la seua traducció (ja existent) i generarem una memòria de traducció. Caldrà fer alguns preparatius i transformacions (apartat 6.4)
3. Iniciarem un nou projecte de traducció amb OmegaT: proveïrem OmegaT amb dues memòries de traducció, la que acabem de crear i una altra que us passarà el professorat. Després traduirem el nou text amb la memòria. Com a resultat, tindrem una memòria de traducció i un text traduït en OpenOffice.org. (apartat 6.5).

6.3 Instal·lació

S'assumeix que teniu instal·lat OpenOffice.org.

6.3.1 Instal·lació de Java i d'OmegaT

Abans de res, heu de comprovar que teniu instal·lada una versió recent (1.4 o superior) de l'entorn d'execució de Java en el vostre ordinador. Potser una manera molt directa de comprovar-ho és descarregar-vos de www.omegat.org l'última versió d'OmegaT (sin JRE) i descomprimir-la en un directori `prac6` que creareu en algun lloc del disc dur o de la vostra memòria USB.²⁸ Si fent doble clic sobre l'arxiu `OmegaT.jar` o obrint-lo des de dins d'un navegador s'executa el programa, i us apareixen diverses finestres, no heu de fer res.²⁹ Si no, heu d'anar a <http://java.com/es/>, descarregar-vos-el (si heu d'elegir, que siga "Fuera de línea"), i seguir les instruccions, o bé baixar-vos la versió d'OmegaT "amb JRE".

²⁸Si el fitxer no es descomprimeix automàticament, instal·leu-vos el programa *7zip*, disponible en SoftCatalà (<http://www.softcatala.org/wiki/Rebost:7-Zip>) o directament des del web de *7zip* (<http://www.7-zip.org/>).

²⁹En algun sistema és possible que hàgeu de fer doble clic sobre l'arxiu `OmegaT.bat`.

6.3.2 Instal·lació de l'alineador *bitext2tmx*

El programa *bitext2tmx* (<http://bitext2tmx.sourceforge.net>) no forma part del paquet OmegaT però el necessitarem. Obriu un navegador i baixeu-vos la versió 0.9 des de <http://www.sourceforge.net/projects/bitext2tmx/>, pestanya *Download*, opció *Browse all packages*, clic sobre *bitext2tmx*, clic sobre *0.9* i clic sobre *bitext2tmx.zip*,³⁰ descomprimiu-lo de manera que tots els fitxers vagin dins del mateix directori on abans s'ha instal·lat OmegaT (prac6).

6.4 Creació d'una memòria de traducció a partir d'un text i la seua traducció

En aquesta pràctica traduirem textos amb informació tècnica sobre telèfons mòbils, de l'anglès o del francès a l'espanyol. Per a traduir les especificacions d'un nou telèfon mòbil, el P2000, aprofitarem la versió anglesa o francesa i la versió espanyola de les especificacions del Z741 per a crear una memòria de traducció. Podeu veure un esquema d'aquesta part del procés en la figura 1

6.4.1 Descàrrega de fitxers

Descarregueu-vos al vostre directori prac6 els fitxers OpenOffice.org següents:

- L'original <http://www.dlsi.ua.es/~mlf/iat/prac6/z741en.odt> (anglès) o <http://www.dlsi.ua.es/~mlf/iat/prac6/z741fr.odt> (francès).
- La traducció a l'espanyol <http://www.dlsi.ua.es/~mlf/iat/prac6/z741es.odt>.

6.4.2 Preparació i conversió dels fitxers

Obriu el fitxer original (*z741en.odt* o *z741fr.odt*) amb OpenOffice.org. Ara, el guardarem com a fitxer de text pla codificat en UTF-8: anirem al menú *Fitxer* → *Anomena i desa*, i el guardarem amb el mateix nom però en *Tipus de fitxer* seleccionarem *Text encoded (.txt)*

³⁰O directament de <http://downloads.sourceforge.net/bitext2tmx/bitext2tmx.zip>.

i clicarem sobre la casella *Edita la configuració de filtres / Edit filter settings*. Quan fem clic sobre *Desa*, ens apareixerà un quadre de *Propietats* on seleccionarem, com a *Lletra*, *Unicode (UTF-8)* i com a *Salt de paràgraf*, *CR i LF*. Quan cliquem sobre *D'acord*, ens guardarà el fitxer (*z741en.txt* o *z741fr.txt*). Assegureu-vos que el guardeu en el vostre directori prac6.

Feu el mateix amb la traducció *z741es.odt*.

6.4.3 Creació de la memòria de traducció

Per a crear la memòria de traducció a partir del fitxer original (*z741en.txt* o *z741fr.txt*) i el traduït *z741es.txt*, usarem el programa alineador *bitext2tmx*. Feu-ho així:

1. Aneu al vostre directori prac6.
2. Feu doble clic sobre *bitext2tmx.jar* i se us obrirà el programa *bitext2tmx*.
3. Si els menús estan en anglès aneu al menú *Tools* → *Options* → *Language* i elegiu el català.
4. Aneu al menú *Eines* → *Configuració* i activeu l'opció *Tallar per salt de línia*.
5. Aneu al menú *Arxiu* → *Obrir*. Seleccionau el text origen o font (*z741en.txt* o *z741fr.txt*) i meta (*z741es.txt*) i indiqueu l'idioma (*anglès* o *francès* i *espanyol*, respectivament), i la codificació (els dos són *UTF-8*).
6. Feu clic sobre *Acceptar*.
7. Veureu el text en dues columnes, l'esquerra per al text origen o font i la dreta per al text meta, amb un segment per fila.
8. Useu les tecles que hi ha davall de cada columna per a *Unir* dos segments en un, *Eliminar* un segment o *Tallar* un segment en dos fins que els textos estiguen *alineats*, és a dir, fins que cada segment estiga al costat de la seua traducció. Si us equivoqueu, podeu *Desfer*. Si voleu tornar a començar, feu clic en *Avortar*.

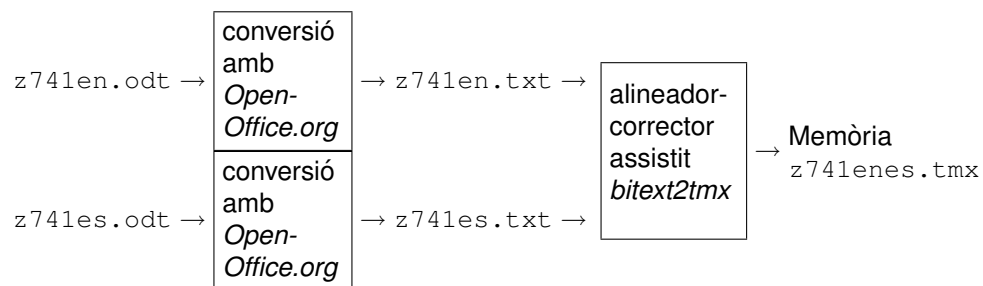


Figura 1: Esquema del procés de conversió i *alineament* dels textos `z741en.odt` i `z741es.odt` per a crear la memòria de traducció `z741enes.tmx` (els noms de fitxer per a la traducció francès–espanyol tenen `fr` en comptes de `en`).

9. Quan hàgeu acabat, feu clic sobre *Guardar* i guardeu la memòria de traducció resultant en el vostre directori `prac6` amb el nom `z741enes.tmx` o `z741fres.tmx` segons quina haja estat la vostra llengua origen.

Ara ja teniu una memòria de traducció on es guarden les unitats de traducció corresponents als documents del model `z741`. Podeu obrir-la amb el *bloc de notes* o amb algun navegador recent (Mozilla, Mozilla Firefox) i examinar-la. Veureu:

- Que el document té com a element arrel `tmx`.
- Que dins de `tmx` hi ha un encapçalament header amb informació sobre el fitxer i un cos `body` que conté les unitats de traducció `tu`, cada una amb un identificador únic `tuid=`.
- Que les unitats de traducció `tu` es componen de variants `tuv`, cada una per a una llengua `lang=`, que contenen un `seg` amb el segment de text corresponent.

6.5 Creació d'un projecte de traducció i traducció d'un text nou

Per a traduir un text nou, primer hem de crear un projecte de traducció amb OmegaT. Aneu al vostre directori `prac6` i feu doble clic sobre `OmegaT.jar`.³¹ Us apareixeran diverses finestres: entre elles, Visor de concordances i de glossari.

³¹En algun sistema és possible que hàgeu de fer doble clic sobre l'arxiu `Omegat.bat`.

6.5.1 Creació d'un nou projecte

Aneu al menú *Fitxer* → *Crea un projecte nou*. Aneu al vostre directori `prac6`. En la línia *Selecció* o equivalent, indiqueu `mobils` i accepteu. OmegaT us respon amb un diàleg en el qual podeu especificar els directoris del projecte i les llengües origen i meta. No canvieu els directoris (però fixeu-vos-hi: seran subdirectoris d'`prac6\mobils`). Seleccioneu la llengua origen (`en` o `fr`) i la llengua meta (`es`) i assegureu-vos que l'opció "Habilita la segmentació per frases" està activada³² abans d'acceptar. El programa us obri una finestra de "Fitxers del projecte". Tanqueu-la, i tanqueu també OmegaT.

6.5.2 Col·locació dels fitxers en el directori del projecte

Ara situarem els fitxers que necessitem per a fer la traducció en el directori corresponents del projecte, és a dir, les dues memòries de traducció corresponents a textos similars.

1. Però abans, descarregueu-vos el fitxer d'especificacions original del model `P2000` (<http://www.dlsi.ua.es/~mlf/iat/prac6/p2000en.odt> o <http://www.dlsi.ua.es/~mlf/iat/prac6/p2000fr.odt>) i guardeu-lo en el vostre directori `prac6`.
2. Copieu la memòria de traducció que heu generat per al model `z741` (`z741enes.tmx` o

³²Les regles de segmentació es poden modificar també des d'ací per a cada idioma

z741fres.tmx) del directori prac6 al directori de memòries de traducció del projecte: prac6\mobils\tm.

3. Descarregueu-vos una altra memòria de traducció (<http://www.dlsi.ua.es/~mlf/iat/prac6/m988enes.tmx> o <http://www.dlsi.ua.es/~mlf/iat/prac6/m988fres.tmx>) i guardeu-la també en el directori de memòries de traducció del projecte: prac6\mobils\tm.

Nota: si no tingúreu una memòria de traducció anteriorment, OmegaT també us permetria traduir les frases del text una a una a mà. OmegaT crea i manté una memòria de traducció que guarda en el directori principal del projecte per a futures traduccions del mateix projecte.

6.5.3 Traducció del nou text

En la figura 2 podeu veure un esquema del procés.

Ara usareu OmegaT per a traduir el nou text (les especificacions del mòbil P2000) amb l'ajuda de les memòries de traducció instal·lades en el projecte. Aneu al menú → Obri i obriu, dins del vostre directori prac6\mobils el fitxer omegat.project. En una finestra anomenada *Fitxers del projecte* podeu carregar el fitxer corresponent al model P2000, que heu de traduir. El sistema el carregarà i us mostrarà el nombre de segments (per exemple, 112). Tanqueu la finestra.

Ara veieu una pantalla on podeu traduir els segments ajudant-vos de la memòria de traducció. Per a traduir a mà un segment, només heu de col·locar el cursor entre <segment...> i <final segment>, fer-hi doble clic, i modificar el text.

Si OmegaT troba segments similars a l'actual en la memòria, els presenta, numerats, en el *Visor de concordances...* de la dreta, indicant amb colors i percentatges les diferències entre el segment actual i els de la memòria, i ens en mostra les traduccions guardades. Si pensem que podem aprofitar la traducció d'un segment de la memòria per a traduir l'actual, podem elegir-la usant les combinacions Ctrl-1 per a la primera, Ctrl-2 per a la segona, etc.³³ i la combinació Ctrl-R per a substituir el text original per la traducció de la memòria, que podem editar fins que siga satisfactòria.

³³Eviteu fer-ho si no existeix la traducció corresponent

La combinació Ctrl-N ens porta al següent segment i la combinació Ctrl-P a l'anterior; per a situar-nos sobre un segment concret, hem de fer-hi doble clic. Quan hem acabat de traduir tot el document, seleccioneu Fitxer → Crea els documents traduïts i després Fitxer → Guarda i Fitxer → Tanca. Ja podeu eixir d'OmegaT (Fitxer → Surt). El resultat està en el vostre directori prac6\mobils\target i es diu igual que el fitxer inicial (p2000en.odt o p2000fr.odt).

Obriu aquest fitxer amb OpenOffice.org i veureu la traducció que acabeu de fer. Guardeu (Anomena i desa) el fitxer en el vostre directori prac6 amb el nom p2000es.odt. Exporteu-la també com a fitxer PDF p2000es.pdf.

Obriu (amb el *bloc de notes* o amb *Mozilla*) les memòries de traducció mobils-level1.tmx (TMX versió 1.1), mobils-level2.tmx (TMX versió 1.4) i mobils-omegat.tmx (idèntic a mobils-level1.tmx) que OmegaT ha creat en prac6\mobils, hi podreu veure les noves unitats de traducció que heu creat (si el *bloc de notes* o el navegador no entenen els caràcters UTF-8, els veureu malament).

Requisits tècnics que ha de complir aquest treball pràctic per a ser avaluat (si no es compleix algun dels requisits, la qualificació serà zero):

- Hi ha un informe anomenat iatpr6.html amb URI <https://cercador.dlsi.ua.es/iat/X/iatpr6.html> (on X és el NIF o el NIE sense la lletra final), accessible a través d'un enllaç des del document mestre. El document iatpr6.html explica breument el procés seguit en el treball pràctic realitzat incloent en el text de l'explicació enllaços a tots aquests documents:
 - Els documents originals corresponents al model z741, exportats com a PDF (z741en.pdf o z741fr.pdf, i z741es.pdf).
 - El document TMX resultant de l'alineament dels documents corresponents al model z741 (que s'ha de dir rz741enes.tmx o z741fres.tmx).
 - El document TMX del model m988 (m988enes.tmx o m988fres.tmx).

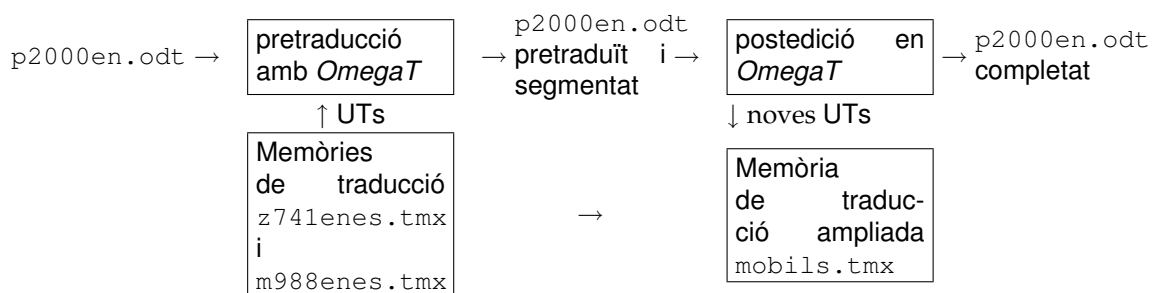


Figura 2: Esquema del procés de *pretraducció* i *postedició* del nou text `p2000en.odt` usant les memòries de traducció `z741enes.tmx` i `p988enes.tmx` (els noms de fitxer per a la traducció francès–espanyol tenen `fr` en comptes de `en`).

- El document original corresponent al model p2000, exportat com a PDF (`p2000en.pdf` o `p2000fr.pdf`).
- La traducció corresponent (`p2000es.pdf`)
- El document TMX resultant de tot el treball (`mobils.tmx`).

- El document `iatpr6.html` és vàlid.

Aspectes que es valoraran en aquest treball pràctic:

- El document està ben traduït
- El document traduït presenta correctament la imatge
- La redacció de l'informe `iatpr6.html` és adequada
- L'informe descriu bé l'objectiu i els passos de la pràctica
- L'informe analitza els diferents tipus de concordança observats durant la traducció de les especificacions del mòbil p2000.
- L'informe inclou una breu crítica d'*OmegaT* i *bitext2tmx* i idees sobre com millorar-los o com usar-los. És important llegir la documentació (ajuda) del programa abans d'escriure l'informe.

A Informe d'autoavaluació (obligatori)

Heu de presentar un únic document d'autoavaluació a mitjan quadrimestre (el 29 de novembre de 2008), que completareu i tornareu a presentar al final (el 13 de gener de 2009, o, si no supereu aquesta part, el 4 de juliol de 2009), on explicareu, molt concisament:

- com han evolucionat les vostres habilitats i coneixements sobre informàtica aplicada a la traducció
- què heu après en cada treball de laboratori — explicant amb les vostres paraules els objectius de cada pràctica— i en cada bloc de seccions d'aula
- on creieu que heu de fer més esforç
- les preocupacions que pugueu tenir
- la relació que pot tenir tot això amb la vostra futura professió
- Els informes i els resultats de cada un dels treballs pràctics que s'assignaran en el laboratori.

Aquests documents han d'estar com a mínim en la *part privada*.

Requisits tècnics que han de complir els informes d'autoavaluació per a ser tinguts en compte

(si no es compleix algun dels requisits, la qualificació d'aquesta part serà *zero*):

- Els autoinformes són accessibles des del document mestre.
- Es presenten en HTML o PDF.
- Si es presenten en HTML, són vàlids.³⁴

Aspectes que es valoraran en els informes d'autoavaluació:

- S'hi descriu l'evolució de les vostres habilitats i coneixements
- S'hi explica el que heu après en cada bloc d'aula i de laboratori
- S'hi parla de preocupacions i motivacions
- S'hi parla de la relació de l'assignatura amb el futur treball professional
- El document és concís però complet.

B Part pública del portafolis (opcional)

La part pública del portafolis és opcional i servirà:

- per a compartir amb la resta de la classe qualsevol material (vostre, o trobat en la xarxa) que pugui ser útil en l'assignatura;
- per a mostrar a futurs empleadors o clients el que heu après sobre Informàtica Aplicada a la Traducció.

Requisits tècnics que ha de complir la part pública del portafolis (si no es compleix algun dels requisits, la qualificació d'aquesta part serà *zero*):

- El document mestre del portafolis públic és accessible des del document mestre de la part privada.

³⁴Atenció: els documents HTML que genera OpenOffice.org poden no ser vàlids.

- El document mestre del portafolis públic s'anomena `index.html` i és accessible directament a través de l'URI `http://cercador.dlsi.ua.es/pubX/`, on *X* és el NIF o el NIE sense la lletra final.
- Tots els documents del portafolis públic estan en HTML o en PDF.
- Si es presenten en HTML, són vàlids.³⁵

Aspectes que es valoraran de la part pública:

- El contingut és adequat; és a dir, conté aspectes d'interès públic —per a la resta de l'alumnat, per a persones exteriors a l'assignatura— i no conté aspectes interns de l'assignatura.
- Serveix com a portafolis professional; és a dir, per a mostrar a futurs clients o empleadors el que heu après sobre informàtica aplicada a la traducció, perquè conté una descripció professional (sincera, és clar) de les vostres habilitats informàtiques, com la que inclouríeu en un *curriculum vitae*.

C Documents XML ben formats

Heus ací algunes de les característiques que fan que un document XML estiga *ben format*, és a dir, siga un document XML i no una altra cosa:

- Cada etiqueta d'inici d'element de la forma `<nom>`, `<nom atribut=valor>`, `<nom atribut=valor atribut=valor>`, etc. (amb zero o més assignacions de valors a atributs) ha d'estar emparellat amb una etiqueta de final d'element de la forma `</nom>`, sense atributs però amb el mateix nom. Si l'element és buit, `<nom...></nom>`, també es pot escriure `<nom.../>`
- Un element pot contenir qualsevol nombre d'elements.
- Els elements no es poden solapar o crear: no és possible escriure, per exemple, `<a>textmés textmés text encara `.

³⁵Atenció: els documents HTML que genera OpenOffice.org poden no ser vàlids.

- El document conté només un element *arrel* que conté tots els elements del text.
- El document pot contenir comentaris entre `<!-- i -->` o instruccions de processament del tipus `<?nom...?>` en qualsevol lloc excepte dins de les etiquetes.
- Els valors dels atributs han d'anar entre cometes dobles "*valor*" o simples '*valor*'.
- Un element no pot tenir dos atributs amb el mateix nom.
- Els caràcters `<` i `&` no poden aparèixer en el text dels elements ni dels atributs (si es necessiten aquests caràcters s'han d'escriure `<` i `&`).