

Determinació del volum i el cost; recompte de paraules i anàlisi d'un projecte. Programes d'ajuda

Antoni Oliver

P08/C0240/00290

Índex

Objectius	5
1. Problemàtica general	7
2. Classificació dels treballs de traducció – localització	8
2.1. Per llengües d'origen i arribada	8
2.1.1. Codificació de llengües	9
2.1.2. Detecció automàtica de llengua	10
2.2. Tipus de document	11
2.3. Per especialitat	11
2.4. Format del document	15
2.5. Per termini de lliurament	18
3. Models de facturació	19
3.1. Traducció	19
3.2. Altres tasques	20
4. Programes d'ajuda a la facturació	22
4.1. Recompte de paraules i anàlisi de projectes	22
4.2. Programes específics d'ajuda a la facturació	22
4.3. Eines d'anàlisi incorporades als sistemes de traducció assistida	24
5. Conclusions	27
6. Per ampliar coneixements	28

Objectius

- 1.** Analitzar els factors que influeixen en la tarificació dels serveis de traducció
- 2.** Aprendre a determinar aquests factors de manera ràpida i acurada
- 3.** Veure la importància del procés de realització de pressupostos i les conseqüències que es poden derivar d'un pressupost malament elaborat

1. Problemàtica general

Els clients, per regla general, volen saber a priori quant hauran de pagar per un treball de traducció. En la majoria dels casos no es conformen amb saber un preu genèric unitari, sinó que volen saber el preu total exacte que hauran de pagar per un treball. Una excepció la poden constituir els clients habituals, pels quals ja s'ha acordat algun tipus de tarifa.

Els pressupostos s'han de fer de manera ràpida i acurada:

- **Ràpida:** perquè el client no està disposat a esperar molt de temps per rebre el pressupost i perquè no podem dedicar moltes hores, cosa que equival a molts diners, a realitzar-lo. S'ha de tenir en compte que no tots els pressupostos que realitzem s'acceptaran.
- **Acurada:** si ens passem a l'alça no serem competitius i perdrem feines; si ens passem a la baixa podem perdre el marge de benefici i acabar perdent diners.

2. Classificació dels treballs de traducció – localització

La tarifa aplicable a un treball de traducció – localització dependrà de diverses característiques. Aquestes característiques no son estàndard i poden variar considerablement entre diverses empreses. És molt important definir quines característiques tindrem en compte i com influiran en el pressupost. De la mateixa manera, és important definir mètodes per determinar de manera ràpida i efectiva aquestes característiques

2.1. Per llengües d'origen i arribada

Existeix la tendència a tenir tarifes diferents segons els parells de llengua implicats en un treball de traducció. Aquesta tendència respon, entre d'altres motius, a la més gran facilitat o dificultat de trobar un traductor per a la combinació de llengües determinada. Aquesta facilitat o dificultat dependrà també del lloc on es trobi l'empresa de traduccions (per exemple és més fàcil trobar un traductor de català a castellà a Barcelona que a Helsinki, però serà més fàcil trobar un de finès a suec a Helsinki que a Barcelona). Ara bé, el tema ubicació geogràfica cada dia perd més importància i no existeixen impediments tècnics per a treballar amb traductors que estiguin en altres països. El que sí que hi ha diferències importants en les tarifes habituals de traducció que es paga en un país o en un altre (per desgràcia a Espanya és on es paguen una de les tarifes més baixes d'Europa). Per aquest motiu serà difícil convèncer a un traductor finès-suec que treballi a Helsinki que ens faci una traducció a les tarifes que es paguen a Espanya. El fet que les tarifes a altres països siguin més altes donen una bona oportunitat de feina als traductors d'aquí de treballar per empreses estrangeres. Ara bé, no s'ha de "rebotar" el mercat estranger aplicant les nostres tarifes, sinó treballar per empreses estrangeres amb les tarifes d'aquell país.

A l'apartat 3.1. d'aquest tema veurem que no totes les llengües "ocupen" el mateix. És a dir, un mateix document en diverses llengües poden tenir nombres de caràcters o paraules substancialment diferents. Aquest factor també caldrà tenir-lo en compte a l'hora de definir un sistema de tarifes.

Hi ha unes poques empreses que no fan diferències de preu entre els diferents parells de llengua. La motivació per no fer diferenciació, argumenten, recau en el fet que és la mateixa feina traduir de la llengua A a la B, que de la C a la D, i per tant és injust pagar diferents tarifes. És més difícil traduir del rus al català que del castellà al català? Deixo la pregunta oberta per donar peu a una possible discussió.

2.1.1. Codificació de llengües

Hi ha una gran quantitat de llengües al món; de fet no se sap amb certesa quantes, ja que ni totes estan documentades ni la distinció entre variants, dialectes i llengua està ben establerta. Es pot parlar, però, d'un ordre de 3500 llengües a tot el món. És interessant tenir disponible informació sobre les llengües del món i informació sobre les seves variants i dialectes. Podeu consultar: www.ethnologue.com.

També és interessant observar que de les 3500 llengües només 150 tenen més de 3.000.000 de parlants.

Existeixen codificacions estàndard per al nom de les llengües. És important fer aquest tipus de codificacions per evitar problemes (si a l'alemany li posem el codi "al", quin li posarem a l'albanès?; i que passarà si una persona de parla anglesa ha de fer servir aquests codis). És molt aconsellable NO posar codis numèrics, perquè poden induir a error (01- castellà; 02.- català; 03 – anglès; 04.- francès). Una bona opció és fer servir les codificacions **ISO639 de 2 (ISO 639-1)** i 3 lletres (**ISO 639-2**). Aconsellem especialment fer servir les codificacions ISO de tres lletres, ja que dóna la possibilitat de codificar més llengües que les ISO de dues lletres.

Podreu trobar aquestes codificacions als següents enllaços web:

- En anglès: http://www.loc.gov/standards/iso639-2/php/English_list.php
- La ISO 639-1 en diverses llengües, incloses el català i el castellà:
<http://www.traduim.com/recursos/propis/index.php?fuseaction=babel>

Si ens fixem en aquestes codificacions, veurem que no codifiquen les diferents variants d'una mateixa llengua. Per exemple, no fa distinció entre el castellà d'Espanya i el d'Argentina, o l'anglès d'Anglaterra i el d'Estats Units. Sovint interessa tenir codificades aquestes variants. Una opció molt estesa és fer servir els codis ISO de llengües de dues lletres més la codificació **ISO 3166-1** d'estats. Aquesta norma ISO té tres variants:

- 3166-1 alfa 2: codis de 2 lletres
- 3166-1 alfa 3: codis de 3 lletres
- 3166-1 numèrica: codi numèric.

Al món de la traducció s'acostumen a fer servir els codis alfa 2. Així, el castellà d'Espanya es codifica com a es-ES, l'anglès d'Estats Units com a en-US, i el català de Catalunya com a ca-ES.

Podeu trobar la llista corresponent a la ISO 3166-1 als següents enllaços web:

- En anglès: http://userpage.chemie.fu-berlin.de/diverse/doc/ISO_3166.html
- En diverses llengües, incloses el català i el castellà:
<http://www.traduim.com/recursos/propis/index.php?fuseaction=pangea>

Si us fixeu en aquests codis que hem proposat, només codifiquen les llengües majoritàries. Ethnologue publica un codi de tres lletres que pretén codificar totes les llengües conegudes. Aquest codi es pot descarregar de la plana web <http://www.ethnologue.com/codes/default.asp#standards>

2.1.2. Detecció automàtica de llengua

La majoria de vegades la llengua original sabrem quina és (o bé la reconeixem o bé el client ens la diu). Podria donar-se el cas que un client ens demanés traduir un text en una llengua que ni el client ni nosaltres sabem quina és. És possible detectar la llengua d'un document de manera automàtica. Tot i que no serà habitual fer servir aquest tipus de tecnologia, explicarem el seu funcionament i presentarem algun programa gratuït de detecció de llengua.

En general la detecció de llengua és un problema prou conegut i solucionat satisfactòriament en tecnologia lingüística. La primera idea que ve al cap per solucionar el problema de la detecció automàtica de llengua és mirar si el text conté una sèrie de paraules que siguin pròpia d'una llengua i no de la resta de llengües. Alguns sistemes es basen en aquest principi, però hi ha una solució més senzilla i molt més eficaç, que es basa en el fet que les combinacions de caràcters en una llengua són diferents que en altres. Així doncs, la majoria de sistemes de detecció automàtica de llengua es basen en estadístiques d'*n*-grames de caràcters. Els *n*-grames de caràcters són combinacions úniques d'*n* caràcters presents en el text (podem parlar d'*unigrames*, *bigrames*, *trigrames*,...). Posem per cas la frase: "El meu germà es diu Jordi."

Els *unigrames* presents al text són: E, l, m, e, u, g, r, à, s, d, i, J, o, ., _ (l'espai en blanc)

Els *bigrames* presents al text són: El, l_, _m, me, eu, u_, _g, ge, er, rm, mà, à_, _e, es, _s, s_, _d, di, iu, _J, Jo, or, rd, di, i.

Els *trigrames* presents al text són: El_, l_m, _me, meu, eu_, u_g, _ge, ger, erm, rmà, mà_, à_e, _es, es_, s_d, _di, diu, iu_, u_J, _Jo, Jor, ord, rdi, di.

Es podria continuar treballant amb *n* superiors. Aquestes combinacions són molt característiques d'una determinada llengua. Si fem estadístiques d'*n*-grames d'un conjunt de llengües i després comparem els *n*-grames del text del qual volem descobrir la llengua amb aquestes estadístiques podrem determinar-la amb força grau d'encert.

Si voleu provar un programa de detecció de llengua, us recomano el TextCat (<http://odur.let.rug.nl/~vannoord/TextCat/>). Podeu treballar "on-line" o descarregar-vos el programa i treballar al vostre ordinador (funciona sota Unix-Linux). Tot i que no el vulgueu executar al vostre ordinador, si descarregueu el programa tindreu accés als models d' n -grames de les diferents llengües que pot detectar.

2.2. Tipus de document

El paràmetre de tipus de document no es fa servir tant com les combinacions idiomàtiques a l'hora de distingir tarifes de traducció. No existeix un llistat estàndard de tipus de document. A continuació fem una proposta, que caldria completar:

- carta
 - comercial
 - personal
- catàleg
- abstract
- article científic
- article divulgatiu
- article de premsa
- llibres (novel·la, teatre, poesia, assaig, de text...)
- manual d'instruccions
- planes web (és més un tipus de document o un format?)
- guions cinematogràfics

2.3. Per especialitat

En moltes tarifes de traducció és habitual fer diferenciació entre llenguatge general i tècnic. No hi ha una classificació estàndard d'especialitats pel que fa a la traducció, però es poden fer servir diverses propostes de classificació temàtica.

- La classificació temàtica servirà tant per classificar els treballs de traducció, com els traductors disponibles, així com per classificar les memòries de traducció i els glossaris terminològics
- Presentarem les següents propostes de classificació temàtica:
 - Classificació Decimal Universal de Dewey (<http://www.tnrndlib.bc.ca/dewey.html>)
 - Àrees de coneixement de la Unesco (<http://www.ub.es/medicina/doctorat/unesco-2.htm>)
 - Classificació temàtica de Google (<http://www.google.com/dirhp?hl=ca>)
 - Classificació temàtica pròpia. A partir de la informació de les propostes presentades, la millor opció és crear una classificació temàtica pròpia

Creació d'una classificació temàtica pròpia

Per crear una classificació temàtica pròpia tindrem en compte els següents punts:

- Basar-se en la realitat de l'empresa
- Organització jeràrquica, en temes i subtemes
- No cal fer servir massa nivells, probablement amb 2 nivells hi haurà prou
- Basar-se en sistemes classificatoris existents, adaptant-los a les necessitats pròpies
- Decimal o alfabètic? Si fem una classificació decimal, cada un dels nivells estarà limitat a 10 possibles valors. Això en alguns casos pot ser insuficient (pensem, per exemple, en les especialitats de medicina). Podem fer servir una classificació alfabètica per poder disposar de més valors a cada nivell.
- Si volem fer una diferenciació en la tarificació no farem servir un sistema classificatori complex.

A continuació presentem una possible classificació temàtica d'exemple, amb un sistema de classificació que fa servir lletres en comptes de números.

- A. Legal - empresa
 - AA. Economia
 - AB. Banca
 - AC. Borsa
 - AD. Assegurances

- AE. Comerç
- AF. Organització i direcció d'empreses
- AG. Dret
 - AGA. Dret privat
 - AGB. Dret públic
 - AGC. Dret penal
 - AGD. Dret internacional
 - AGE. Dret canònic
 - AGF. Legislació
- AH. Política
- B. Humanitats
 - BA. Filosofia
 - BB. Antropologia
 - BC. Sociologia
 - BD. Pedagogia
 - BE. Història
 - BF. Lingüística
 - BG. Geografia
 - BH. Religió
 - BI. Art i cultura
 - BIA. Arts plàstiques
 - BIB. Música
 - BIC. Dansa i ball
 - BID. Cinema
 - BIE. Teatre
 - BIF. Literatura
- C. Ciències
 - CA. Biologia
 - CB. Geologia
 - CD. Física
 - CE. Química
 - CF. Matemàtiques
 - CG. Lògica
- D. Medicina – farmàcia
 - DA. Patologia
 - DB. Terapèutica
 - DC. Especialitats
 - DCA. Andrologia
 - DCB. Cardiologia i sistema limfàtic
 - DCD. Dermatologia
 - DCE. Hepatologia
 - DCE. Medicina forense
 - DCG. Gastroenterologia
 - DCH. Geriatria i gerontologia
 - DCI. Ginecologia
 - DCJ. Hematologia
 - DCK. Medicina de l'aparell muscular

- DCL. Medicina interna
- DCM. Neurologia
- DCN. Obstretícia
- DCO. Oftalmologia.
- DCP. Oncologia
- DCR. Otorrinolaringologia
- DCS. Pediatria
- DCT. Pneumologia
- DCU. Psiquiatria
- DCV. Reumatologia
- DCW. Sexologia
- DCX. Estomatologia i odontologia
- DCY. Urologia
- DCZ. Toxicologia
- DC0. Traumatologia
- DC1. Alergologia
- DD. Medicina preventiva i salut pública
- DE. Medicina del treball
- DF. Epidemiologia
- DG. Medicina legal
- DH. Anatomia
- DI. Histologia
- DJ. Fisiologia
- DK. Farmàcia
- E. Enginyeria – tecnologia
 - EA. Enginyeria civil
 - EB. Enginyeria mecànica
 - EC. Enginyeria electrònica
 - ED. Enginyeria aeronàutica
 - EE. Enginyeria naval
 - EF. Enginyeria industrial
 - EG. Energia
 - EH. Metal·lúrgia i siderúrgia
 - EI. Minería
 - EJ. Indústria de la fusta, del paper i del suro
 - EK. Indústria editorial
 - EL. Arquitectura
 - EM. Agricultura
 - EN. Transport
- F. Telecomunicacions – informàtica
 - FA. Informàtica maquinari
 - FB. Informàtica programari
 - FC. Programació
 - FD. Bases de dades
 - FE. Protecció de dades
 - FF. Intel·ligència artificial
 - FG. Internet

- FH. Telecomunicacions
- FI. Telefonia
- G. Altres
 - GA. Oci
 - GB. Turisme
 - GC. Gastronomia
 - GD. Esports

2.4. Format del document

El format dels documents o dels fitxers a traduir – localitzar és un factor important, que influirà en:

- Pressupost
- Flux de treball
- Programari a fer servir
- Professional que farà la feina

Els avenços tècnics fan que cada dia sigui menys freqüent que arribin feines de traducció en paper. S'estan imposant els formats electrònics. Tot i així encara continuen arriben treballs en format paper. Al tema 7 *Formats de codificació de text i dades. Característiques i tractament* de l'assignatura "Traducció i tecnologies" vam veure en detall els formats més habituals i la manera de tractar-los. En aquest apartat farem el repàs d'algun d'aquests formats i veurem com afecta a la manera de fer el pressupost. Recordem que és important fer un pressupost d'una manera ràpida i acurada

Format paper

El format paper no permet un tractament informàtic. Això té influència en la manera de fer el pressupost i també en el flux de treball i programari a fer servir.

- Existeix la possibilitat de fer OCR del document en paper.: cal escanejar el document i aplicar un programa de Reconeixement Òptic de Caràcters (OCR). És un procés lent i laboriós, ja que sovint requereix molta revisió. Per fer el pressupost no és una bona solució. Per traduir certs documents si que pot ser d'utilitat la utilització d'escàners i OCR:
 - Si es disposen de memòries de traducció i glossaris i es fa servir un sistema de traducció assistida
 - Per reduir les despeses de tramesa dels treballs (no serà estrictament necessari fer OCR, n'hi haurà prou amb escanejar i enviar-los per correu electrònic en algun format gràfic)

- Per fer el pressupost: es pot fer per planes o bé estadísticament (contant manualment les paraules d'unes quantes planes significatives i multipliant pel número de planes)
- Aplicar increments si el client vol un manteniment del format del document ja que caldrà fer una tasca important de maquetació

Format processador de textos

El format processador de textos permet un tractament informàtic senzill i un pressupost ràpid. El problema és si disposem del programa i versió adequats per obrir i tractar el format. Si no és així caldrà demanar al client que ens el passi en un altre format, cosa que acabarà tenint alguna conseqüència sobre el manteniment del format.

- Els programes de traducció assistida permeten un ampli ventall de formats de processador de textos d'entrada
- Un bon format intermedi en cas d'incompatibilitat és l'RTF (Rich Text Format).
- Els mateixos processadors de textos permeten el recompte de paraules, però poden haver discrepàncies entre la manera de contar d'un i un altre programa.

PDF (Portable Document Format)

- El format PDF està pensat per què sigui compatible entre diferents plataformes i que es vegin exactament igual en totes elles.
- És un format final que es crea a partir d'altres (per exemple Word, latex, etc...)
- Té l'avantatge de ser fàcilment distribuïble però és difícilment tractable
- Té característiques de seguretat que poden fer-lo totalment intractable
- Si no està protegit es pot exportar el text, però és difícil mantenir el format. També hi ha aplicacions per contar paraules. Compte amb aquests recomptes!
- Com que no és un format final el millor és demanar els fitxers en format original, però el client no sempre els té o els vols buscar

- En alguns casos el tractament d'un PDF s'assemblarà al tractament d'un document en paper i caldrà aplicar els increments necessaris de les tarifes, especialment pel que fa a la posterior maquetació.

WEB

- Els formats Web es van diversificant enormement i ja no són simples documents html, poden contenir scripts (java, VB..), accessos a bases de dades, etc.
- Els filtres dels programes de traducció assistida no sempre estan prou actualitzats i poden deixar d'importar seccions senceres
- Recordeu que podeu fer servir les utilitats de pseudotraducció per verificar si els filtres d'entrada han funcionat correctament
- El gestor de projectes, si s'enfronta a traducció de planes web haurà de tenir uns coneixements mitjans de tota la tecnologia relacionada. Per una molt bona introducció, i gratuïta, adreceu-vos a la plana web W3: www.w3schools.com
- Els projectes web són un repte important des del punt de vista tecnològic i organitzatiu: molts arxius, moltes tecnologies i formats diferents. Moltes webs són absolutament caòtiques.

Localització de programari

Els projectes de localització de programari poden ser extremadament variats i amb tecnologies diferents. Per regla general ens podem trobar amb tres tipus de fitxers a traduir:

- Fitxers font: són els fitxers escrits en el llenguatge de programació concret.
- Fitxers de recursos (*resource files*): són fitxers que contenen el text localitzable. Contenen la definició dels quadres de diàleg, cadenes i icones utilitzades pel programa
- Executables: ens podem trobar que el client no aporti els fitxers anteriors i que haguem de treballar directament amb els fitxers executables.

Imatges

- Les imatges per regla general no són tractables mitjançant programes de traducció assistida
- Poden portar un temps considerable de maquetació

- Un procediment és no considerar el text a traduir, sinó un preu per imatge, considerant que per regla general el preu de maquetació serà superior al de traducció
- Contar les paraules de les imatges és una tasca feixuga. Existeixen alguns programes per alleugerir-la.

Àudio – vídeo

- Ens podem trobar amb un encàrrec de traduir uns missatges que estan en algun format d'àudio o de vídeo.
- S'acostuma a pressupostar per minut de gravació. No és convenient transcriure la cinta per fer el pressupost i en molts casos no cal transcriure per traduir, sinó que es pot traduir directament del missatge oral.

2.5. Per termini de lliurament

Sovint s'aplica un recàrrec si el temps necessari per realitzar un treball de traducció és inferior a un cert límit. La justificació d'aquest augment de tarifa per urgència és que el treball caldrà distribuir-lo entre un nombre superior de traductors i que les tasques de gestió, de revisió i de control de qualitat s'incrementaran.

3. Models de facturació

En aquest apartat presentarem els models més habituals de facturació per a diferents tasques relacionades amb la traducció. Presentarem bàsicament les unitats a facturar per a cada tasca. Pel que fa referència a les tarifes, no les presentarem ja que no existeix uniformitat suficient.

3.1. Traducció

En tasques de traducció la unitat de facturació més habitual és la paraula, però existeixen altres possibilitats. També es poden donar dos casos, que les paraules es comptin de l'original a traduir o del document traduït.

Les possibles unitats de facturació pel cas de la traducció són:

- **Paraula:** és la unitat més habitual
- **Caràcter:** és menys freqüent però s'acostuma a fer servir per a llengües de caire aglutinant, ja que el nombre de paraules acostuma a ser més baix
- **Línia:** s'acostuma a fer servir molt poc, i quan es fa servir acostuma a tractar-se d'una línia estàndard que conté un nombre determinat de caràcters o paraules (i per tant passa a ser un simple recompte de caràcters o paraules).
- **Plana:** es fa servir en alguns casos d'originals en paper per evitar haver de fer el recompte de paraules. En altres casos es tracta de planes estàndard, amb un nombre determinat de caràcters o paraules i, per tant, com en el cas de les línies estàndard, passa a ser un recompte de caràcters o paraules).
- **Per obra o document:** es fa servir en alguns casos de traducció literària o d'obres de divulgació. Es pacta un preu per la totalitat de l'obra.

A la següent taula podem observar diversos recomptes sobre un mateix document en castellà (spa), anglès (eng), francès (fra) i alemany (deu). Els recomptes que presentem són planes, paraules, caràcters (sense comptar espais), caràcters (comptant espais), paràgrafs i línies.

	Planes	Paraules	Caract (s/e)	Caract. (a/e)	Paràgrafs	Línies
spa	29	11.610	61.501	73.041	263	1.044
eng	27	10.163	55.187	65.105	265	937
fra	31	11.754	65.796	77.315	265	1.090

	Planes	Paraules	Caract (s/e)	Caract. (a/e)	Paràgrafs	Línies
deu	30	8.882	63.207	71.827	265	1.080

Fixeu-vos com el mateix document té moltes menys paraules en alemany que en castellà, però com en canvi té més caràcters (sense espais) que el castellà. Tot això caldrà tenir-ho en compte a l'hora de crear el sistema de tarifes.

Un altre aspecte a tenir en compte és sobre quina versió, original o traduïda, es fa el recompte de paraules. Quan el document original està en un format electrònic que sigui fàcil de comptar, s'acostuma a fer el recompte sobre l'original. En canvi, si l'original està per exemple, en paper, és habitual fer el recompte sobre la versió traduïda. Ara bé, cal tenir en compte que només és possible fer el recompte sobre la versió traduïda quan no és necessari fer un pressupost previ. Analitzem uns quants casos a partir de les xifres d'aquesta taula:

Si fem el recompte de paraules amb un preu unitari de 0.05 € per paraula i comptem les paraules sobre l'original en anglès, el preu de la traducció a les altres llengües seria de $10163 \times 0.05 = 508.15$ €. Si l'original és l'anglès, però el tenim en paper i decidim fer el recompte sobre la versió traduïda i volem traduir-lo al castellà i a l'alemany ens trobem que:

- El traductor al castellà cobra $11.610 \times 0,05 = 580.5$ €
- El traductor a l'alemany cobra $8.882 \times 0,05 = 441.1$ €

Una diferència de 136,4 € per la mateixa tasca de traducció.

Un altre factor a tenir en compte és que poden haver discrepàncies entre els recomptes fets amb diferents programes. Això no és degut a què contin malament, sinó a què no tenen una mateixa definició de paraula (guions separen paraules, o apòstrofs, etc) i a més a més, potser no contenen els mateixos elements (nota a peu, anotacions, capçaleres, peus de plana, quadres de text, etc).

3.2. Altres tasques

Altres tasques relacionades amb la traducció tenen models de facturació diferents. A la següent taula presentem una proposta.

Adaptat de: *The guide to translation and localization*. Lingo Systems. Podeu demanar-ne una còpia gratuïta a: www.lingosys.com.

A nivell de projecte	Preu
Gestió de projecte	Normalment el 10-15 % del pressupost total
Documentació	Preu
Traducció	Per paraula, caràcter, línia o plana

Maquetació	Per paraula, plana o hora
Revisió	Per paraula, plana o hora
Creació de terminologia o glossaris	Per hora
Assegurament de la qualitat	Per hora
Programari, webs i documentació "on line"	Preu
Traducció, maquetació i revisió	El mateix que per documentació
Creació de terminologia o glossaris	El mateix que per documentació
Captures de pantalla	Per captura de pantalla o per hora
Enginyeria	Per hora
Proves funcionals	Per hora
Gràfics i captures de pantalla	Per hora
Ús de memòries de traducció	Preu
Coincidència exacta (exact match)	Sense càrrec o un percentatge del preu de traducció
Coincidència parcial (fuzzy match)	Un percentatge del preu de traducció
Segments o oracions noves	Per paraula
Administració i manteniment de la memòria de traducció	Per hora

4. Programes d'ajuda a la facturació

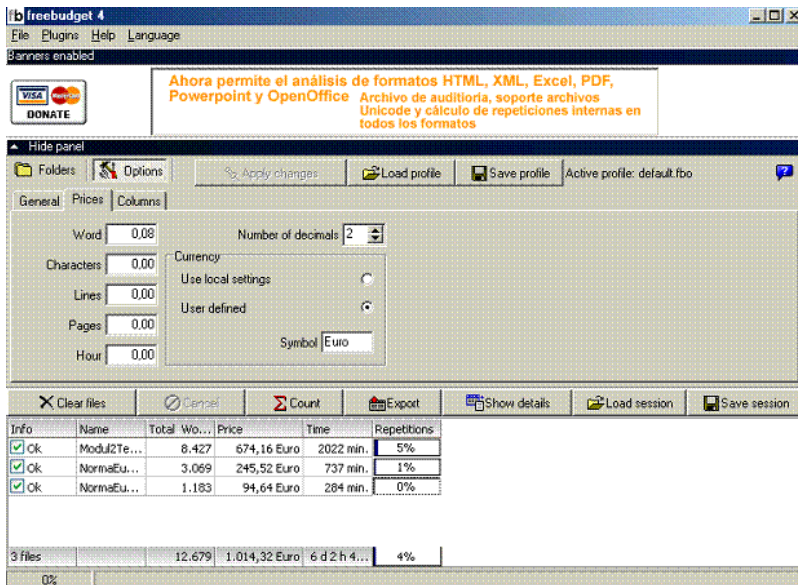
4.1. Recompte de paraules i anàlisi de projectes

Hi ha una diferència important entre el concepte de recompte de paraules (o altres unitats com per exemple caràcters) i l'anàlisi de projectes. Un recompte de paraules només ens dóna una idea de la quantitat total de paraules que conté un projecte de traducció, però no ens indica quantes d'aquestes paraules estan en segments que es poden recuperar de les memòries de traducció i quantes presents en segments que es repetiran dins del mateix projecte. Per saber això haurem de fer una anàlisi detallada del projecte. Una anàlisi ens donarà informació sobre:

- Les paraules totals del projecte
- Les paraules que estan en segments que es poden recuperar de la memòria de traducció per diferents graus de coincidència (coincidència exacta (*exact match*) o coincidència aproximada (*fuzzy match*)). Dins de les coincidències aproximades ens poden donar diferents xifres per a diferents ventalls (per exemple, entre 95 i 99%, entre 90 i 95%, etc).
- Les paraules que estan en segments que es repeteixen dins del mateix projecte. És el que es coneix com a repetició interna. Tot i que no estan presents a la memòria de traducció, només caldrà traduir-les una vegada, ja que després es podrà recuperar del propi projecte.

4.2. Programes específics d'ajuda a la facturació

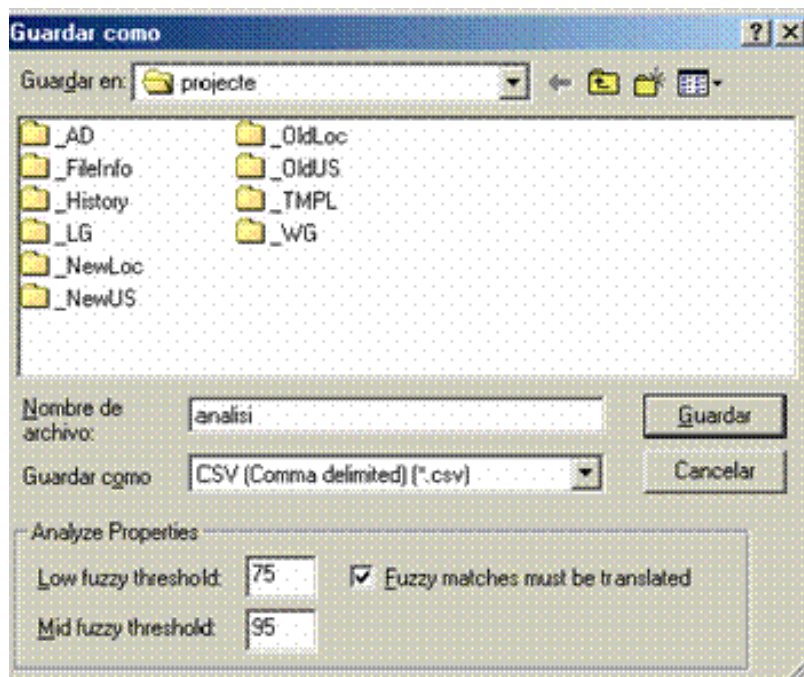
Existeixen una sèrie de programes per realitzar recomptes i diferents anàlisis de projectes de traducció. Entre aquests es pot destacar el **Freebudget**. Es tracta d'un programa gratuït que podeu obtenir a l'adreça www.webbudget.com/freebudget/index.htm. Aquesta eina permet fer recomptes i pressupostos de múltiples arxius txt, rtf i word (entre d'altres). Cal destacar les següents funcionalitats:



- Recompte de paraules o caràcters:
 - Permet comptar tots els arxius d'una carpeta, amb l'opció d'incloure també els arxius de totes les subcarpetes
 - Permet definir una plana estàndard com a una certa quantitat de paraules o caràcters
 - Permet seleccionar els diferents elements a comptar dels documents (capçaleres i peus de planes, notes a peu, quadres de text, etc.)
 - Permet adoptar el mateix criteri de recompte que el Microsoft Word
 - Permet fer una estimació de la repetitivitat interna dels documents
- Pressupost:
 - Permet definir un preu unitari per al càlcul automàtic del pressupost
- Càlcul de terminis:
 - Permet definir una productivitat diària dels traductors per fer el càlcul de terminis
 - Permet fer el càlcul de terminis per a equips de més d'un traductor
- Exportació:
 - Permet exportar el pressupost a diferents formats

4.3. Eines d'anàlisi incorporades als sistemes de traducció assistida

La majoria d'eines de traducció assistida ofereixen utilitats de recompte de paraules i anàlisi de projectes. El ForeignDesk també permet fer anàlisis de projectes. Si fem Project > Analyze, ens apareix la següent pantalla:



En aquesta pantalla podem donar el nom de l'arxiu csv (delimitat per comes) on s'emmagatzemarà l'anàlisi. També ens permet definir dos nivells de *Fuzzy match* un de baix (que per defecte està al 75%) i un de mig (que per defecte és del 95%). L'anàlisi que ens ofereix ForeignDesk té el següent aspecte:

Project	Total Words	Words Left	Real Words Left	Propagated Words
	1364	1364	1298	66

File Name	Leverage Words	95-100%		75-94%		New Words	Total Words	Leverage Segments	95-100%		75-94%		New Segments	Total Segments
		Words	Segments	Words	Segments				Words	Segments				
dogo3.html	0	0	7	147	154	2	0	1	14	17				
dogo2.html	0	0	7	171	178	2	0	1	14	17				
dogo1.html	0	4	0	435	439	4	4	0	27	35				
dogo4.html	0	584	6	3	593	0	17	1	1	19				

Com podeu veure, ForeignDesk ens ofereix estadístiques a nivell de paraula i a nivell de segment. A continuació oferim una explicació de les xifres de l'anàlisi, adaptada de la que ofereix el manual de ForeignDesk:

Informació de projecte:

Total words:	El nombre total de paraules del projecte
Words left:	El nombre de paraules que queden per traduir, descomptant aquelles reciclades ¹ o ja traduïdes.
Real words left:	El nombre de paraules que queden per traduir tenint en compte l'autopropagació interna
Propagated words:	El nombre de paraules que es propaguen dins del projecte

⁽¹⁾En ForeignDesk el concepte de "reciclat" consisteix en aprofitar les traduccions provinents d'un altre projecte de ForeignDesk, que pot estar totalment o parcialment traduït.

Informació per arxiu:

Leverage words/segments:	El nombre de paraules o segments que han estat reciclats
95%-100% words/segments:	El nombre de paraules o segments que s'aprofiten de les memòries de traducció amb una similitud entre el 95% i el 100%
75%-94% words/segments:	El nombre de paraules o segments que s'aprofiten de les memòries de traducció amb una similitud entre el 75% i el 94%
New words/segments:	El nombre de paraules o segments nous del projecte.
Total words/segments:	El nombre total de paraules o segments del projecte.

Informació de resum:

Total:	El nombre total de paraules i segments, comptant els reciclats, 95-100%, 74-95% i els nous.
Propagation:	El nombre total de paraules i segments propagats dins dels 95-100%, 75-94% i els nous.
Total less propagation:	El nombre total de paraules i segments de 95-100%, 75-94% i els nous sense tenir en compte els propagats.

Els diferents sistemes de traducció assistida ofereixen anàlisis de projectes substancialment diferents. A continuació presentem una anàlisi d'un projecte feta amb el Déjà Vu X:

Analysys

Language: Spanish

	Lines	Words	Characters	Codes
Exact	23	71	598	6
99%	9	0	55	0
98%	0	0	0	0
97%	0	0	0	0
96%	0	0	0	0
95%	0	0	0	0
90%-94%	0	0	0	0
80%-89%	0	0	0	0
70%-79%	0	0	0	0
60%-69%	1	3	16	0
50%-59%	0	0	0	0
40%-49%	1	20	119	0
30%-39%	1	9	70	0
20%-29%	2	35	226	0
10%-19%	7	137	809	4
0%-9%	0	0	0	0
No Match	151	1481	9240	94
Total:	195	1756	11133	104

Global

Duplicate Sentences: 18 Internal Repetition: 5%

Analyze Copy Save Close

A la següent figura podem observar la pantalla d'anàlisi de projectes del Trados:

Analyse Files - My First Memory (English-German)

Add Remove Clear Analyze Cancel Options Help

Files to analyse: Use TM from previous analysis

Log File: D:\

0%

Match types:	Segments	Words	Percent
Context TM:	0	0	0%
Repetitions:	0	0	0%
100%:	0	0	0%
95% - 99%:	0	0	0%
85% - 94%:	0	0	0%
75% - 84%:	0	0	0%
50% - 74%:	0	0	0%
No match:	0	0	0%
Total:	0	0	0%

Analysis results:

View Log File... Export Frequent Segments... 5 or more occurrences

Create Project TM... Export Unknown Segments... 05 % or lower match value

5. Conclusions

En aquest tema hem presentat els aspectes més rellevants d'un projecte de traducció que cal tenir en compte a l'hora de realitzar un pressupost. També hem presentat algunes tècniques i eines d'ajuda al recompte de paraules i anàlisi dels projectes. Hem de recordar aquí les dues característiques que ha de tenir tot pressupost: ha de ser ràpid de fer i acurat.

6. Per ampliar coneixements

Actualment s'està elaborant un esborrany de norma ISO 639 – 3 de codificació de llengües que inclogui totes les llengües del món. Us recomanem la lectura de l'article de Wikipedia http://es.wikipedia.org/wiki/ISO_639-3 i la consulta de l'esborrany a

<http://www.sil.org/iso639-3/codes.asp>