Determinación del volumen y el coste; recuento de palabras y análisis de un proyecto. Programas de ayuda

Antoni Oliver

P08/B0240/00290

UOC
www.uoc.edu

Índice

Objetivos 5

1. Problemática general 7

2. Clasificación de los trabajos de traducción - localización .. 8
2.1. En función de las lenguas de origen y destino 8
2.1.1. Codificación de lenguas 9
2.1.2. Detección automática de lengua 10
2.2. Tipo de documento 11
2.3. Por especialidad 11
2.4. Formato del documento 15
2.5. Por plazo de entrega 19

3. Modelos de facturación 20
3.1. Traducción 20
3.2. Otras tareas 21

4. Programas de ayuda a la facturación 23
4.1. Recuento de palabras y análisis de proyectos 23
4.2. Programas específicos de ayuda a la facturación 23
4.3. Herramientas de análisis incorporadas a los sistemas de traducción asistida 25

5. Conclusiones 28

6. Para ampliar conocimientos 29

Objetivos

1. Analizar los factores que influyen en la tarificación de los servicios de traducción

2. Aprender a determinar estos factores de manera rápida y esmerada

3. Ver la importancia del proceso de realización de presupuestos y las consecuencias que se pueden derivar de un presupuesto mal elaborado

1. Problemática general

Los clientes, por regla general, quieren saber a priori cuánto tendrán que pagar por un trabajo de traducción. En la mayoría de los casos no se conforman con saber un precio genérico unitario, sino que quieren saber el precio total exacto que tendrán que pagar por un trabajo. Una excepción la pueden constituir los clientes habituales, por los cuales ya se ha acordado algún tipo de tarifa.

Los **presupuestos se tienen que hacer de manera rápida y esmerada:**

- **Rápida**: porque el cliente no está dispuesto a esperar mucho de tiempo para recibir el presupuesto y porque no podemos dedicar muchas horas, cosa que equivale a mucho dinero, para realizarlo. Se debe tener en cuenta que no todos los presupuestos que realizamos se aceptarán.

- **Esmerada**: si nos pasamos al alza no seremos competitivos y perderemos trabajos; si nos pasamos a la baja podemos perder el margen de beneficio y acabar perdiendo dinero.

2. Clasificación de los trabajos de traducción - localización

La tarifa aplicable a un trabajo de traducción - localización dependerá de diversas características. Estas características no son estándar y pueden variar considerablemente dependiendo de las empresas. Es muy importante definir qué características tendremos en cuenta y cómo influirán en el presupuesto. De la misma manera, es importante definir métodos para determinar de manera rápida y efectiva estas características

2.1. En función de las lenguas de origen y destino

Existe la tendencia a tener tarifas diferentes según los pares de lengua implicados en un trabajo de traducción. Esta tendencia responde, entre otros motivos, a la mayor facilidad o dificultad de encontrar un traductor para la combinación de lenguas determinada. Esta facilidad o dificultad dependerá también del lugar en qué se encuentre la empresa de traducciones (por ejemplo es más fácil encontrar a un traductor de catalán a castellano en Barcelona que en Helsinki, pero será más fácil encontrar uno de finés a sueco en Helsinki que en Barcelona). Ahora bien, el tema ubicación geográfica cada día pierde más importancia y no existen impedimentos técnicos para trabajar con traductores que estén en otros países. Sí que hay diferencias importantes en las tarifas habituales de traducción que se paga en un país o en otro (por desgracia en España es donde se pagan una de las tarifas más bajas de Europa). Por este motivo será difícil convencer a un traductor finés-sueco que esté trabajando en Helsinki que nos haga una traducción por las tarifas que se pagan en España. El hecho de que las tarifas en otros países sean más altas constituyen una buena oportunidad para los traductores de aquí: trabajar para empresas extranjeras. Sin embargo, no hay que "reventar" el mercado extranjero aplicando nuestras tarifas; lo procedente es trabajar para empresas extranjeras con las tarifas de aquel país.

En el apartado 3.1 de este tema veremos que no todas las lenguas "ocupan" lo mismo. Es decir, un mismo documento en diversas lenguas puede tener una cantidad de caracteres o palabras sustancialmente diferente. Este factor también habrá de ser tenido en cuenta cuando haya que definir un sistema de tarifas.

Hay algunas empresas, pocas, que no aplican diferencias de precio entre los diferentes pares de lenguas. La motivación que se argumenta recae en el hecho de que el trabajo de traducir de la lengua A a la B es el mismo que el de traducir

de la C en la D, y por lo tanto es injusto pagar tarifas distintas. ¿Es más difícil traducir del ruso al catalán que del castellano al catalán? Dejo la pregunta abierta para dar pie a una posible discusión.

2.1.1. Codificación de lenguas

Existe una gran cantidad de lenguas en el mundo; de hecho no se sabe con certeza cuántas hay, ya que ni todas están documentadas ni existen criterios únicos para distinguir entre lenguas, variantes y dialectos. Se puede hablar, sin embargo, del orden de 3500 lenguas en todo el mundo. Es interesante tener disponible información sobre las lenguas del mundo e información sobre sus variantes y dialectos. Para ello podéis consultar www.ethnologue.com

También es interesante observar que de las 3500 lenguas sólo 150 tienen más de 3.000.000 de hablantes.

Existen codificaciones estándar para el nombre de las lenguas. Es importante hacer este tipo de codificaciones para evitar problemas (si al alemán le ponemos el código "al", ¿cuál le pondremos al albanés?; ¿que pasará si una persona de habla inglesa tiene que utilizar estos códigos?). Es muy aconsejable no poner códigos numéricos, porque pueden inducir a error (01- castellano; 02. - catalán; 03 - inglés; 04. - francés). Una buena opción es utilizar las codificaciones **ISO 639** de 2 (**ISO 639-**1) y 3 letras (**ISO 639-**2). Aconsejamos especialmente utilizar las codificaciones ISO de tres letras, ya que ofrece la posibilidad de codificar más lenguas que las ISO de dos letras.

Podréis encontrar estas codificaciones a los siguientes enlaces web:

- En inglés: http://www.loc.gov/standards/iso639-2/php/English_list.php

- La ISO 639-1 en diversas lenguas, incluidas catalán y castellano: http://www.traduim.com/recursos/propis/index.php?fuseaction=babel

Si nos fijamos en estas codificaciones, veremos que no codifican las diferentes variantes de una misma lengua. Por ejemplo, no hacen distinción entre el castellano de España y el de Argentina, o entre el inglés de Inglaterra y el de Estados Unidos. A menudo interesa tener codificadas estas variantes. Una opción muy extendida es utilizar los códigos ISO de lenguas de dos letras más la codificación **ISO 3166-1** de estados. Esta norma ISO tiene tres variantes:

- 3166-1 alfa 2: códigos de 2 letras
- 3166-1 alfa 3: códigos de 3 letras
- 3166-1 numérica: código numérico.

En el mundo de la traducción se acostumbran a utilizar los códigos alfa 2. Así, el castellano de España se codifica como se-SE, el inglés de Estados Unidos como en-OS, y el catalán de Cataluña como casa-SE.

Podéis encontrar la lista correspondiente a la ISO 3166-1 a los siguientes enlaces web:

- En inglés: http://userpage.chemie.fu-berlin.de/diverse/doc/ISO_3166.html

- En diversas lenguas, incluidas el catalán y el castellano: http://www.traduim.com/recursos/propis/index.php?fuseaction=pangea

Si os fijáis en estos códigos que hemos propuesto, sólo codifican las lenguas mayoritarias. Ethnologue publica un código de tres letras que pretende codificar todas las lenguas conocidas. Dicho código se puede descargar de la siguiente página web: http://www.ethnologue.com/codes/default.asp#standards

2.1.2. Detección automática de lengua

Habitualmente la lengua original nos será conocida (o la conocemos o el cliente nos dice cuál es). Pero podría darse el caso de que un cliente nos pidiera traducir un texto original de una lengua que ni el cliente ni nosotros sabemos cuál es. Es posible detectar la lengua de un documento de manera automática. Aunque no será habitual utilizar este tipo de tecnología, explicaremos su funcionamiento y presentaremos algún programa gratuito de detección de lengua.

En general la detección de lengua es un problema lo bastante conocido y satisfactoriamente resuelto en tecnología lingüística. Lo primero que normalmente se nos ocurre para solucionar el problema de la detección automática de lengua es mirar si el texto contiene una serie de palabras que sean propias de una lengua y no del resto de lenguas. Algunos sistemas se basan en este principio, pero existe una solución más sencilla y mucho más eficaz, que se basa en el hecho de que las combinaciones de caracteres en una lengua son diferentes a las de otras. Así pues, la mayoría de sistemas de detección automática de lengua se basan en estadísticas de *n-gramas* de caracteres. Las *n-gramas* de caracteres son las combinaciones únicas de de *n* caracteres presentes en el texto (por tanto se puede hablar de *unigramas, bigramas, trigramas, etc.*). Pongamos por caso la siguiente frase del catalán:

"El meu germà es diu Jordi."

Las *unigramas* presentes en el texto son: E, l, m, e, u, g, r, à, s, d, i, J, o, ., _ (el carácter de espacio en blanco)
Las *bigramas* presentes en el texto son: 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.

Las *trigramas* presentes en el texto son: 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.

Se podría continuar trabajando con *n* superiores. Estas combinaciones son muy características de una determinada lengua. Si hacemos estadísticas de las *n-gramas* de un conjunto de lenguas y después las comparamos con las *n-gramas* del texto del cual queremos conocer la lengua, con estas estadísticas podremos determinarlo con un grado de acierto muy elevado.

Si queréis probar un programa de detección de lengua, os recomiendo el TextCat (http://odur.let.rug.nl/~vannoord/TextCat/). Podéis trabajar "on-line" o descargaros el programa y trabajar en vuestro ordenador (funciona bajo Unix-Linux). Aunque no lo queráis ejecutar en vuestro ordenador, si descargáis el programa tendréis acceso a los modelos de las *n-gramas* de las diferentes lenguas que puede detectar.

2.2. Tipo de documento

El parámetro de tipo de documento no se utiliza tanto como las combinaciones idiomáticas para distinguir tarifas de traducción. No existe un listado estándar de tipo de documento. A continuación hacemos una propuesta, que sería necesario completar:

- carta
  - comercial
  - personal

- catálogo
- abstract
- artículo científico
- artículo divulgativo
- artículo de prensa
- libros (novela, teatro, poesía, ensayo, de texto...)
- manual de instrucciones
- páginas web (¿es más un tipo de documento o un formato?)
- guiones cinematográficos

2.3. Por especialidad

En muchas tarifas de traducción es habitual distinguir entre lenguaje general y técnico. No existe una clasificación estándar de especialidades con respecto a la traducción, pero se pueden utilizar diversas propuestas de clasificación temática.

- La clasificación temática servirá tanto para clasificar los trabajos de traducción, como los traductores disponibles, así como para clasificar las memorias de traducción y los glosarios terminológicos

- Presentaremos las siguientes propuestas de clasificación temática:
  - Clasificación Decimal Universal de Dewey (http://www.tnrdlib.bc.ca/dewey.html)

  - Áreas de conocimiento de la Unesco (http://www.ub.es/medicina/doctorat/unesco-2.htm)

  - Clasificación temática de Google (http://www.google.com/dirhp?hl=ca)

  - Clasificación temática propia. A partir de la información de las propuestas presentadas, la mejor opción es crear una clasificación temática propia

Creación de una clasificación temática propia

Para crear una clasificación temática propia tendremos en cuenta los siguientes puntos:

- Basarse en la realidad de la empresa

- Organización jerárquica, en temas y subtemas

- No hay que utilizar demasiados niveles, probablemente con 2 niveles será suficiente

- Basarse en sistemas clasificatorios existentes, adaptándolos a las necesidades propias

- ¿Decimal o alfabético? Si hacemos una clasificación decimal, cada uno de los niveles estará limitado a 10 posibles valores. En algunos casos esto puede ser insuficiente (pensemos, por ejemplo, en las especialidades de medicina). Podemos utilizar una clasificación alfabética para poder disponer de más valores a cada nivel.

- Si queremos hacer una diferenciación en la tarificación no utilizaremos un sistema clasificatorio complejo.

A continuación presentamos una posible clasificación temática de ejemplo, con un sistema de clasificación que utiliza letras en vez de números.

A) Legal - empresa

AA) Economía
AB) Banca
AC) Bolsa
AD) Seguros
AE) Comercio
AF) Organización y dirección de empresas
AG) Derecho

AGA) Derecho privado
AGB) Derecho público
AGC) Derecho penal
AGD) Derecho internacional
AGE) Derecho canónico
AGF) Legislación

AH) Política

B) Humanidades

BA) Filosofía
BB) Antropología
BC) Sociología
BD) Pedagogía
BE) Historia
BF) Lingüística
BG) Geografía
BH) Religión
BI) Arte y cultura

BIA) Artes plásticas
BIB) Música
BIC) Danza y baile
BID) Cine
BIE) Teatro
BIF) Literatura

C) Ciencias

CA) Biología
CB) Geología
CD) Física
CE) Química
CF) Matemáticas
CG) Lógica

D) Medicina - farmacia

DA) Patología
DB) Terapéutica

DC) Especialidades
DCA) Andrología
DCB) Cardiología y sistema linfático
DCD) Dermatología
DCE) Hepatología
DCF) Medicina forense
DCG) Gastroenterología
DCH) Geriatría y gerontología
DCI) Ginecología
DCJ) Hematología
DCK) Medicina del aparato muscular
DCL) Medicina interna
DCM) Neurología
DCN) Obstetricia
DCO) Oftalmología.
DCP) Oncología
DCR) Otorrinolaringología
DCS) Pediatría
DCT) Neumología
DCU) Psiquiatría
DCV) Reumatología
DCW) Sexología
DCX) Estomatología y odontología
DCY) Uronefrología
DCZ) Toxicología
DC0) Traumatología
DC1) Alergología

DD) Medicina preventiva y salud pública
DE) Medicina del trabajo
DF) Epidemiología
DG) Medicina Legal
DH) Anatomía
DI) Histología
DJ) Fisiología
DK) Farmacia

E) Ingeniería - tecnología

EA) Ingeniería civil
EB) Ingeniería mecánica
EC) Ingeniería electrónica
ED) Ingeniería aeronáutica
EE) Ingeniería naval

EF) Ingeniería industrial
EG) Energía
EH) Metalurgia y siderurgia
EI) Minería
EJ) Industria de la madera, el papel y el corcho
EK) Industria editorial
EL) Arquitectura
EM) Agricultura
EN) Transporte

F) Telecomunicaciones - informática

FA) Informática hardware
FB) Informática software
FC) Programación
FD) Bases de datos
FE) Protección de datos
FF) Inteligencia artificial
FG) Internet
FH) Telecomunicaciones
FI) Telefonía

G) Otros

GA) Ocio
GB) Turismo
GC) Gastronomía
GD) Deportes

2.4. Formato del documento

El formato de los documentos o de los ficheros a traducir - localizar es un factor importante, que influirá en:

- Presupuesto
- Flujo de trabajo
- Software a utilizar
- Profesional que realizará el trabajo

Los avances técnicos hacen que cada día sea menos frecuente que lleguen trabajos de traducción en papel. Se están imponiendo los formatos electrónicos. Aun así aún continúan llegando algunos trabajos en formato papel. En el tema 7 *Formatos de codificación de texto y datos. Características y tratamiento,*de la asignatura "Traducción y tecnologías" vimos con detalle los formatos más habi-

tuales y cómo tratarlos. En este apartado reapasaremos alguno de éstos formatos y veremos como afectan a la manera de hacer el presupuesto. Recordemos que es importante hacer un presupuesto de una manera rápida y esmerada.

Formato papel

El formato papel no permite su tratamiento informático. Esto influye en la manera de hacer el presupuesto y también en el flujo de trabajo y el software a utilizar.

- Existe la posibilidad de realizar un OCR del documento en papel: habrá que escanear el documento y aplicarle un programa de Reconocimiento Óptico de Caracteres (siglas en inglés: OCR). Es un proceso lento y laborioso, ya que a menudo requiere de mucha revisión. Para hacer el presupuesto, por tanto, no es una buena solución. Para traducir ciertos documentos si que puede ser de utilidad la utilización de escáneres y OCR:
  - Si se disponen de memorias de traducción y glosarios y se utiliza un sistema de traducción asistida
  - Para reducir los gastos de envío de los trabajos (no será estrictamente necesario hacer OCR, será suficiente con escanear y enviarlos por correo electrónico en algún formato gráfico)
- Para hacer el presupuesto: se puede hacer por páginas o bien estadísticamente (contando manualmente las palabras de unas cuantas páginas significativas y multiplicando por el número de páginas)
- Aplicar incrementos si el cliente quiere un mantenimiento del formato del documento ya que habrá que hacer una tarea importante de maquetación

Formato procesador de textos

El formato procesador de textos permite un tratamiento informático sencillo y un presupuesto rápido. El problema es si disponemos del programa y versión adecuados para abrir y tratar el formato. Si no es así habrá que pedir al cliente que nos lo pase en otro formato, cosa que acabará teniendo alguna consecuencia sobre el mantenimiento del formato.

- Los programas de traducción asistida permiten la entrada de un amplio abanico de formatos de procesador de textos
- Uno buen formato intermedio, en caso de incompatibilidad, es RTF (Rich Text Format).
- Los mismos procesadores de textos permiten el recuento de palabras, pero puede haber discrepancias entre la manera de contar de uno y otro programa.

PDF (Portable document format)

- El formato PDF está pensado para que sea compatible entre diferentes plataformas y pueda ser visualizado exactamente igual en todas ellas.

- Es un formato final que se crea a partir de otros (por ejemplo Word, Latex, etc.)

- Tiene la ventaja de ser de fácil distribución pero sin embargo es difícilmente tratable

- Tiene características de seguridad que pueden hacerlo totalmente intratable

- Si no está protegido se puede exportar el texto, pero es difícil mantener el formato. También hay aplicaciones para contar palabras. Pero ¡mucho cuidado con estos recuentos!

- Como no es un formato final siempre es mejor pedir los ficheros en formato original, pero el cliente no siempre los tiene o los quieres buscar

- En algunos casos el tratamiento de un PDF se parecerá al tratamiento de un documento en papel y habrá que aplicar los incrementos necesarios de las tarifas, especialmente con respecto a la posterior maquetación.

WEB

- Los formatos Web se van diversificando enormemente y ya no son simples documentos html, pueden contener scripts (java, VB...), accesos a bases de datos, etc.

- Los filtros de los programas de traducción asistida no siempre están lo bastante actualizados y pueden dejar de importar secciones enteras

- Recordad que podéis utilizar las utilidades de pseudotraducción para verificar si los filtros de entrada han funcionado correctamente

- El gestor de proyectos, si se enfrenta a traducción de páginas web deberá tener unos conocimientos medios de toda la tecnología relacionada. Para una muy buena introducción, y gratuita, dirigíos a la web W3: www.w3schools.com

- Los proyectos web son un reto importante desde el punto de vista tecnológico y organizativo: multiplicidad de archivos, tecnologías y formatos diferentes. Además, muchas webs son absolutamente caóticas.

Localización de software

Los proyectos de localización de software pueden ser extremadamente variados y estar relacionados con tecnologías diferentes. Por regla general nos podemos encontrar con tres tipos de ficheros a traducir:

- Ficheros fuente: son los ficheros escritos en el lenguaje de programación concreto.

- Ficheros de recursos (*resource files*): son ficheros que contienen el texto localizable. Contienen la definición de los cuadros de diálogo, cadenas e iconos utilizados por el programa.

- Ejecutables: nos podemos encontrar que el cliente no aporte los ficheros anteriores y que tengamos que trabajar directamente con los ficheros ejecutables.

Imágenes

- Las imágenes por regla general no son tratables mediante programas de traducción asistida

- Pueden llevar un tiempo considerable de maquetación

- Un procedimiento posible es el de no considerar el texto a traducir, sino un precio por imagen, considerando que por regla general el precio de maquetación será superior al de traducción

- Contar las palabras de las imágenes es una tarea pesada. Existen algunos programas para facilitarla.

Audio - vídeo

- Nos podemos encontrar con un encargo de traducir mensajes que estén en algún formato de audio o de vídeo.

- Se acostumbra a presupuestar por minuto de grabación. No es conveniente transcribir la cinta para hacer el presupuesto y en muchos casos no hay que transcribir para traducir, sino que se puede traducir directamente del mensaje oral.

2.5. Por plazo de entrega

A menudo se aplica un recargo si el tiempo necesario para realizar un trabajo de traducción es inferior a un cierto límite. La justificación de este aumento de tarifa por urgencia es que será necesario distribuir el trabajo entre un número superior de traductores y que las tareas de gestión, de revisión y de control de calidad se incrementarán.

3. Modelos de facturación

En este apartado presentaremos los modelos más habituales de facturación para diferentes tareas relacionadas con la traducción. Presentaremos básicamente las unidades a facturar para cada tarea. Con respecto a las tarifas, no las presentaremos ya que no existe uniformidad suficiente.

3.1. Traducción

En tareas de traducción la unidad de facturación más habitual es la palabra, pero existen otras posibilidades. También se pueden dar dos casos: que las palabras se cuenten del original a traducir o del documento traducido.

Las posibles unidades de facturación para el caso de la traducción son:

- **Palabra**: es la unidad más habitual

- **Carácter**: es menos frecuente pero se acostumbra a utilizar para lenguas de tipo aglutinante, ya que el número de palabras acostumbra a ser más bajo

- **Línea**: se acostumbra a utilizar muy poco, y cuando se utiliza acostumbra a tratarse de una línea estándar que contiene un número determinado de caracteres o palabras (y por lo tanto pasa a ser un simple recuento de caracteres o palabras).

- **Página:** se utiliza en algunos casos con originales en papel para evitar tener que hacer el recuento de palabras. En otros casos se trata de páginas estándar, con un número determinado de caracteres o palabras y, por lo tanto, como en el caso de las líneas estándar, pasa a ser un recuento de caracteres o palabras).

- **Por obra o documento**: se utiliza en algunos casos de traducción literaria o de obras de divulgación. Se pacta un precio por la totalidad de la obra.

En la siguiente tabla podemos observar diversos recuentos sobre un mismo documento en castellano (spa), inglés (eng), francés (fra) y alemán (deu). Los recuentos que presentamos son páginas, palabras, caracteres (sin contar espacios), caracteres (contando espacios), párrafos y líneas.

	Páginas	Palabras	Caract (s/e)	Caract (a/e)	Párrafos	Líneas
spa	29	11.610	61.501	73.041	263	1.044
eng	27	10.163	55.187	65.105	265	937

	Páginas	Palabras	Caract (s/e)	Caract (a/e)	Párrafos	Líneas
fra	31	11.754	65.796	77.315	265	1.090
deu	30	8.882	63.207	71.827	265	1.080

Fijaos cómo el mismo documento tiene muchas menos palabras en alemán que en castellano, pero como en cambio tiene más caracteres (sin espacios) que el castellano. Todo eso habrá que tenerlo en cuenta a la hora de crear el sistema de tarifas.

Otro aspecto a tener en cuenta es sobre qué versión, original o traducida, se hace el recuento de palabras. Cuando el documento original está en un formato electrónico que sea fácil de contar, se acostumbra a hacer el recuento sobre el original. En cambio, si el original está por ejemplo, en papel, es habitual hacer el recuento sobre la versión traducida. Ahora bien, hay que tener en cuenta que sólo es posible hacer el recuento sobre la versión traducida cuando no es necesario hacer un presupuesto previo. Analizamos unos cuantos casos a partir de las cifras de esta tabla:

Si hacemos el recuento de palabras con un precio unitario de 0.05 € por palabra y contamos las palabras sobre el original en inglés, el precio de la traducción a las otras lenguas sería de 10163x0.05= 508.15 €. Si el original es el inglés, pero lo tenemos en papel y decidimos hacer el recuento sobre la versión traducida y queremos traducirlo al castellano y al alemán nos encontramos que:

- El traductor al castellano cobra 11.610 x 0,05=580,5 €
- El traductor al alemán cobra 8882 x 0,05=441,1 €

Una diferencia de 136,4 € por la misma tarea de traducción.

Otro factor a tener en cuenta es que puede haber discrepancias entre los recuentos hechos con diferentes programas. Eso no es debido en qué cuenten mal, sino en qué no tienen una misma definición de palabra (guiones o apóstrofos separan palabras, etc.) y además, quizás no cuentan los mismos elementos (notas al pie, anotaciones, cabeceras, pies de página, cuadros de texto, etc.).

3.2. Otras tareas

Otras tareas relacionadas con la traducción tienen modelos de facturación diferentes. En la siguiente tabla[1] presentamos una propuesta.

(1)Adaptado de: *The guide to translation and localization. Lingo Systems.* Podéis pedir una copia gratuita a www.lingosys.com.

A nivel de proyecto	Precio
Gestión de proyecto	Normalmente el 10-15%-15 del presupuesto total

Documentación	**Precio**
Traducción	Por palabra, carácter, línea o página
Maquetación	Por palabra, página u hora
Revisión	Por palabra, página u hora
Creación de terminología o glosarios	Por hora
Aseguramiento de la calidad	Por hora
Software, webs I documentación "on line"	**Precio**
Traducción, maquetación I revisión	Lo mismo que por documentación
Creación de terminología o glosarios	Lo mismo que por documentación
Capturas de pantalla	Por captura de pantalla o por hora
Ingeniería	Por hora
Pruebas funcionales	Por hora
Gráficos y capturas de pantalla	Por hora
Uso de memorias de traducción	**Precio**
Coincidencia exacta (exact match)	Sin cargo o un porcentaje del precio de traducción
Coincidencia parcial (fuzzy match)	Un porcentaje del precio de traducción
Segmentos u oraciones nuevas	Por palabra
Administración y mantenimiento de la memoria de traducción	Por hora

4. Programas de ayuda a la facturación

4.1. Recuento de palabras y análisis de proyectos

Hay una diferencia importante entre el concepto de recuento de palabras (u otras unidades como por ejemplo caracteres) y el análisis de proyectos. Un recuento de palabras sólo nos da una idea de la cantidad total de palabras que contiene un proyecto de traducción, pero no nos indica cuántas de estas palabras están en segmentos que se pueden recuperar de las memorias de traducción y cuántas presentes en segmentos que se repetirán dentro del mismo proyecto. Para saberlo habrá que hacer un análisis detallado del proyecto. Un análisis nos dará información sobre:

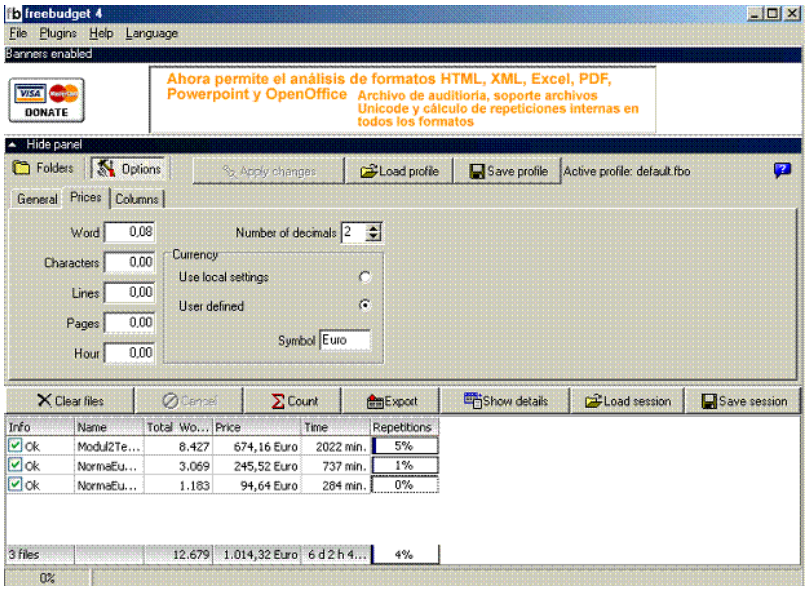
- Las palabras totales del proyecto

- Las palabras que están en segmentos que se pueden recuperar de la memoria de traducción por diferentes grados de coincidencia (coincidencia exacta (*exact match*) o coincidencia aproximada (*fuzzy match*)). Dentro de las coincidencias aproximadas nos pueden dar diferentes cifras para diferentes abanicos (por ejemplo, entre 95 y 99%, entre 90 y 95%, etc.).

- Las palabras que están en segmentos que se repiten dentro del mismo proyecto. Es lo que se conoce como repetición interna. Una de estas palabras, aunque no esté presente en la memoria de traducción, sólo habrá que traducirla una vez, ya que después se podrá recuperar del propio proyecto.

4.2. Programas específicos de ayuda a la facturación

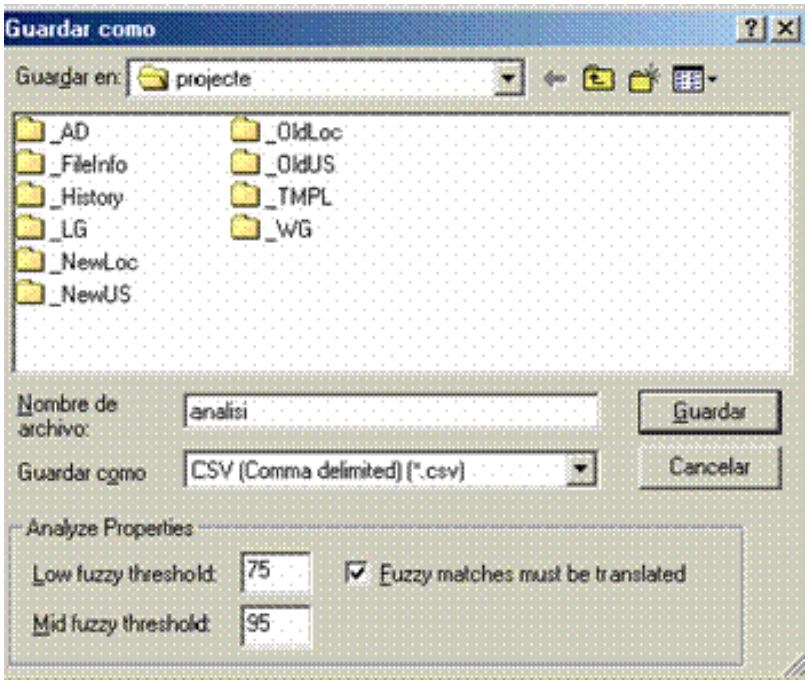
Existen una serie de programas dedicados a la realización de recuentos y diferentes análisis de proyectos de traducción. Entre éstos cabe destacar **Freebudget**. Se trata de un programa gratuito que podéis obtener en la dirección www.webbudget.com/freebudget/index.htm. Esta herramienta permite hacer recuentos y presupuestos de múltiples archivos txt, rtf y word (entre otros). Hay que destacar las siguientes funcionalidades:



- Revuento de palabras o caracteres:
  - Permite contar todos los archivos de una carpeta, con la opción de incluir también los archivos de todas las subcarpetas
  - Permite definir la página estándar como una cierta cantidad de palabras o caracteres
  - Permite seleccionar los diferentes elementos a contar de los documentos (cabeceras y pies de página, notas al pie, cuadros de texto, etc.)
  - Permite adoptar el mismo criterio de recuento que el Microsoft Word
  - Permite hacer una estimación de la repetitividad interna de los documentos
- Presupuesto:
  - Permite definir un precio unitario para el cálculo automático del presupuesto
- Cálculo de plazos:
  - Permite definir una productividad diaria de los traductores para hacer el cálculo de plazos
  - Permite realizar el cálculo de plazos para equipos de más de un traductor
- Exportación:
  - Permite exportar el presupuesto a diferentes formatos

4.3. Herramientas de análisis incorporadas a los sistemas de traducción asistida

La mayoría de herramientas de traducción asistida ofrecen utilidades de recuento de palabras y análisis de proyectos. ForeignDesk también permite realizar análisis de proyectos. Si ejecutamos Project > Analize, nos aparece la siguiente pantalla:



En esta pantalla podemos poner el nombre del archivo csv (delimitado por comas) en el que se almacenará el análisis. También nos permite definir dos niveles de *Fuzzy match* uno bajo (que por defecto es del 75%) y uno de medio (por defecto: 95%). El análisis que nos ofrece ForeignDesk tiene el siguiente aspecto:

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

File Name	Leverage Words	95-100% Words	75-94% Words	New Words	Total Words	Leverage Segments	95-100% Segments	75-94% Segments	New Segments	Total Segments
dogc3.html	0	0	7	147	154	2	0	1	14	17
dogc2.html	0	0	7	171	178	2	0	1	14	17
dogc1.html	0	4	0	435	439	4	4	0	27	35
dogc4.html	0	584	6	3	593	0	17	1	1	19

Como se puede ver, ForeignDesk nos ofrece estadísticas a nivel de palabra y a nivel de segmento. A continuación ofrecemos una explicación de las cifras del análisis, adaptada de la que ofrece el manual de ForeignDesk:

Información de proyecto:

Total words:	El número total de palabras del proyecto
Words left:	El número de palabras que quedan por traducir, descontando aquéllas recicladas[2] o ya traducidas.
Real words left:	El número de palabras que quedan por traducir teniendo en cuenta la autopropagación interna
Propagated words:	El número de palabras que se propagan dentro del proyecto

(2)En ForeignDesk el concepto de "reciclado" refiere al aprovechamiento de las traducciones que provienen de otro proyecto de ForeignDesk, que puede estar total o parcialmente traducido

Información por archivo:

Leverage words/segments:	El número de palabras o segmentos que han sido reciclados
95%-100% words/segments:	El número de palabras o segmentos que se aprovechan de las memorias de traducción con una similitud de entre el 95% y el 100%
75%-94% words/segments:	El número de palabras o segmentos que se aprovechan de las memorias de traducción con una similitud de entre el 75% y el 94%
New words/segments:	El número de palabras o segmentos nuevos del proyecto.
Total words/segments:	El número total de palabras o segmentos del proyecto.

Información de resumen:

Total:	El número total de palabras y segmentos, contando los reciclados, 95-100%, 74-95% y los nuevos.
Propagation:	El número total de palabras y segmentos propagados dentro de los 95-100%, 75-94% y los nuevos.
Total less propagation:	El número total de palabras y segmentos de 95-100%, 75-94% y los nuevos sin tener en cuenta los propagados.

Los diferentes sistemas de traducción asistida ofrecen análisis de proyectos sustancialmente diferentes. A continuación presentamos un análisis de un proyecto realizado con Déjà Vu X:

Analysis

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

En la siguiente figura podemos observar la pantalla de análisis de proyectos de Trados:

Analyse Files - My First Memory (English-German)

Add... Remove Clear Analyse

Files to analyse:

Use TM from previous analysis

Cancel

Options

Help

Log File D:\

0%

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 occurences

Create Project TM... Export Unknown Segments... 85 % or lower match value

5. Conclusiones

En este tema hemos presentado los aspectos más relevantes de un proyecto de traducción que hay que tener en cuenta a la hora de realizar un presupuesto. También hemos presentado algunas técnicas y herramientas de ayuda al recuento de palabras y al análisis de los proyectos. Recordemos una vez más las dos características imprescindibles de todo presupuesto: se tiene que poder hacer rápidamente y tiene que ser esmerado.

6. Para ampliar conocimientos

Actualmente se está elaborando un borrador de norma ISO 639 - 3 de codificación de lenguas que incluya todas las lenguas del mundo. Os recomendamos la lectura del artículo de Wikipedia http://es.wikipedia.org/wiki/ISO_639-3 y la consulta del borrador en http://www.sil.org/iso639-3/codes.asp