

... the same geograph
same cultural tradition: th
of Belgium; a Bantu langua
language; the English lang

trans·la·tion /træns'leiʃn
rendering of something into
language or into one's own
language. 2. a version of su

1. Tradução Assistida por Computador

¹Em toda a Terra, havia somente uma língua, e empregavam-se as mesmas palavras. ²Emigrando do oriente, os homens encontraram uma planície na terra de Chinear e nela se fixaram. ³Disseram uns para os outros: «Vamos fazer tijolos, e cozamo-los ao fogo.» Utilizaram o tijolo em vez da pedra, e o betume serviu-lhes de argamassa. ⁴Depois disseram: «Vamos construir uma cidade e uma torre, cujo cimo atinja os céus. Assim, havemos de tornar-nos famosos para evitar que nos dispersemos por toda a superfície da Terra.» ⁵O Senhor, porém, desceu, a fim de ver a cidade e a torre que os homens estavam a edificar. ⁶E o Senhor disse: «Eles constituem apenas um povo e falam uma única língua. Se principiaram desta maneira, coisa nenhuma os impedirá, de futuro, de realizarem todos os seus projectos. ⁷Vamos, pois, descer e confundir de tal modo a linguagem deles que não consigam compreender-se uns aos outros.»

Génesis 11, 1:7

1.1 A Tradução Manual

Sem dúvida que desde cedo o Homem teve necessidade em comunicar. A língua foi evoluindo em cada ponto da terra, sendo que quando habitantes de diferentes zonas do planeta terra se começaram a cruzar, surgiu a necessidade da comunicação inter-língua. Se por um lado surgiu a necessidade da tradução oral, assim que os primeiros escritos foram encontrados por diferentes culturas, foi necessária a sua tradução.

O documento mais antigo que se conhece que contém a tradução de um texto em mais que uma língua é a Pedra de Roseta (ver figura 1.1). Esta pedra contém um decreto promulgado em 196 a.C., em nome do rei Ptolomeu V. Este decreto é composto por três parágrafos, ou partes, cada uma numa língua distinta, mas com o mesmo conteúdo. Os dois parágrafos superiores estão escritos em egípcio, o primeiro por meio de hieroglíficos, em egípcio antigo, e o segundo em Demótico, uma variante do egípcio mais recente. O terceiro parágrafo está escrito em grego antigo. Se por um lado esta pedra acaba por ser um artefacto curioso, não deixa de ser, também, o mais antigo

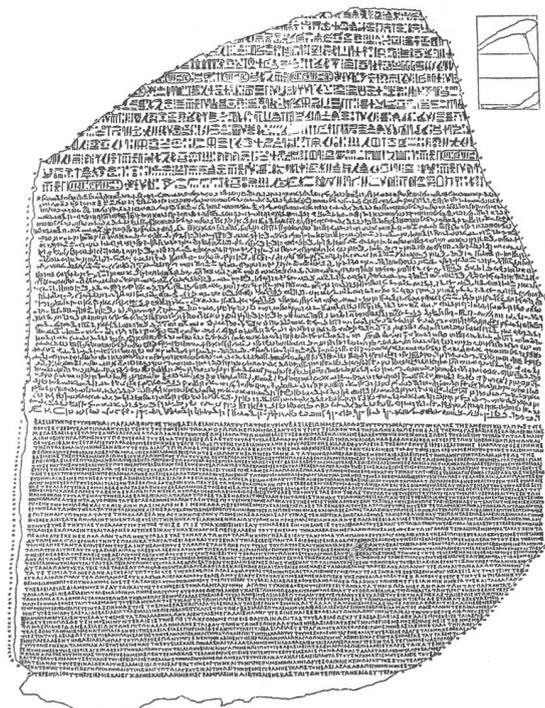


Figure 1.1: Pedra de Roseta.

exemplo de um texto que foi usado para aprender outra, já que foi graças à Pedra de Roseta que os investigadores foram capazes de decifrar a escrita hieroglífica.

Supõe-se que a tradução continuou tal como a própria reprodução de livros. Se na época medieval a reprodução era feita pelos *copistas*, que meticolosa e cuidadosamente copiavam os livros, estes mesmos copistas transformavam-se em *tradutores* quando o livro era copiado de uma língua para outra.

Este tipo de tradução não foi completamente extinta, e ainda há muito quem use esta abordagem na tradução. Especialmente quando falamos da tradução literária, a abordagem totalmente manual ainda poderá fazer algum sentido. Já na tradução técnica e especializada, a quantidade de conhecimento necessária para a realização de uma boa tradução é tal que a tradução manual, com consulta manual a diferentes fontes de conhecimento, é demasiado lenta, propensa a erros e a traduções inconsistentes.

1.2 Tradução Automática

Recentemente temos assistido ao uso massivo de tradutores automáticos, ferramentas que, dado um texto numa língua, são capazes de produzir traduções numa qualquer outra língua. Se por um lado esta afirmação não é uma mentira, por outro, também não é a verdade completa, já que a tradução produzida pelas aplicações de tradução automática ainda é bastante pobre.

A evolução da tradução automática está intrinsecamente ligada à evolução da capacidade computacional dos dispositivos eletrónicos e à quantidade de recursos disponibilizados. As técnicas recentes de tradução automática baseiam-se no processamento de documentos paralelos (pedras de Roseta), pares de textos, em que um é a tradução do outro. Isto leva a que a tradução entre línguas para as quais existem mais recursos seja mais fiável do que para as restantes. Por exemplo, para obter um tradutor entre a língua portuguesa e a língua árabe seria necessário obter documentos paralelos entre estas duas línguas, em grandes quantidades. É claro que este tipo de recurso é

escasso. Mas se pensarmos num sistema de tradução entre a língua portuguesa e a inglesa, sabemos que será muitíssimo mais fácil encontrar documentos paralelos entre essas duas línguas.

Concordamos, no entanto, que a tradução automática não está pronta (e possivelmente nunca estará) para ser usada como uma ferramenta certa, capaz de traduzir um documento oficial num outro, pronto para ser enviado sem qualquer revisão. Mas pode ser útil para muitas outras coisas!

Um bom uso da tradução automática é denominada de *tradução exploratória*, e tem como principal objetivo a compreensão de um texto. Por exemplo, se precisarmos perceber determinado texto que só nos foi entregue em norueguês, é possível que o consigamos compreender traduzindo-o com um tradutor automático. Em princípio, a tradução de norueguês para português não será a melhor. Mas caso não seja de todo perceptível, é possível traduzir de norueguês para inglês e, com ajuda da tradução portuguesa e inglesa, perceber a mensagem. Por vezes, a própria triangulação de traduções (de norueguês para inglês e posteriormente para português) pode ser suficiente para que se consiga compreender a mensagem original.

1.3 Tradução Assistida por Computador

A tradução assistida por computador, habitualmente designada por CAT (Computer Assisted Translation) pode ser vista como a tradução manual com recurso às tecnologias da informação (e, mais recentemente, também com recurso a ferramentas de tradução automática).

Habitualmente a CAT é realizada com software próprio, que integra de algum modo diferentes recursos que são úteis ao tradutor enquanto desempenha as suas funções. Por exemplo, permite a integração com dicionários ou terminologias, ou o acesso a bases de memórias de tradução (sobre as quais falaremos em breve).

No mundo em que vivemos, um tradutor nunca receberá um trabalho que consista na tradução de um livro, sendo que esse livro lhe é entregue em papel, e em que a tradução também deverá ser entregue nesse mesmo formato. Os pedidos de tradução vêm nos mais diferentes formatos, mas todos eles eletrónicos. Um tradutor tanto poderá ter de traduzir uma página da Internet (habitualmente armazenadas num formato denominado HTML), como poderá ter de traduzir um panfleto (que poderá, por exemplo, estar armazenado num ficheiro InDesign) ou a tradução de um livro (que provavelmente virá em formato Word). O que o tradutor terá de fazer é criar um documento em tudo semelhante ao recebido, mas em que o texto está numa outra língua.

Esta forma de trabalhar levanta um grande problema: os tradutores não podem gastar o seu tempo a aprender todas as diferentes tecnologias que a cada momento vão surgindo. Nesse sentido, a ferramenta de CAT deverá ser capaz de reconhecer estes formatos, processá-los e compreendê-los, e escondê-los do tradutor. Este último, deverá concentrar-se na tradução do texto e, finda essa tarefa, a ferramenta de CAT deverá ser capaz de reconstruir o documento final, aplicando os mesmos estilos do documento original. Na figura 1.2, que representa os vários componentes de um sistema CAT, este processo de lidar com os formatos dos documentos é representado pelas setas *Strip Markup* e *Rebuild Markup*.

Nessa mesma figura vemos que a tradução tira partido de duas bases de dados: uma terminológica e outra de memórias de tradução.

A necessidade de uma ou mais bases de dados terminológicas é relativamente fácil de defender. Imagine-se o que seria a tradução de um documento médico com referências a termos como “parênquima”, “ecogenicidade”, “esteatose”, “ectasia”, “endoluminal” ou “esplenomegalia”. Claro que na pior das hipóteses o tradutor iria consultar dicionários, textos técnicos e, se necessário fosse, um especialista da área (neste caso, medicina). Com a informação obtida iria saber realizar a tradução e, como tradutor previdente, iria apontar a tradução usada em algum sítio para que mais tarde, quando voltasse a ter a necessidade de traduzir essa mesma palavra, não tivesse de gastar todo o mesmo tempo a consultar as diferentes fontes de informação. Estes apontamentos, em que

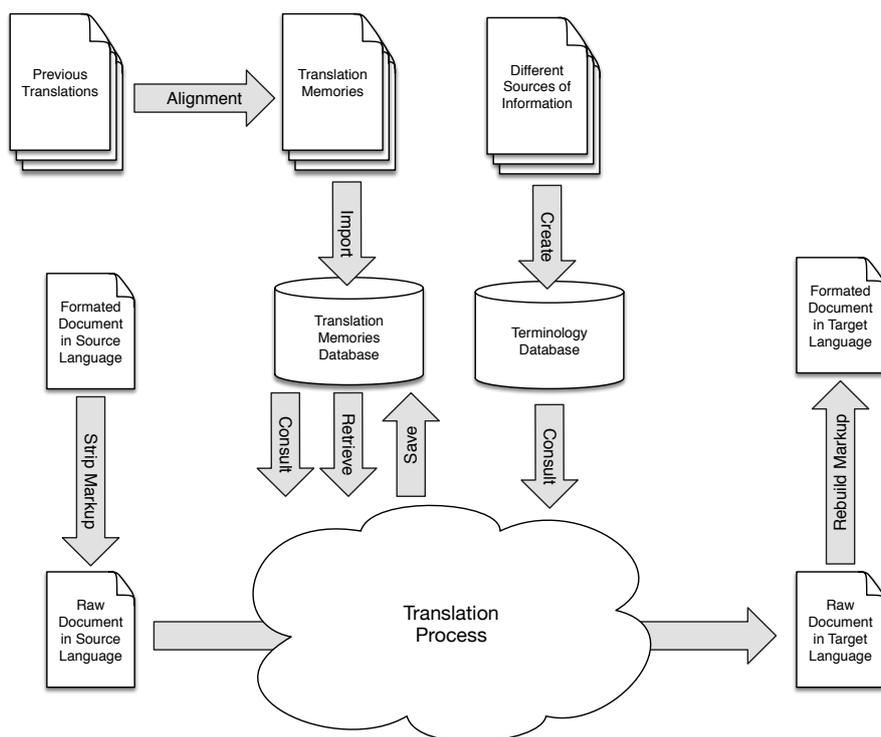


Figure 1.2: Componentes de um sistema CAT

o tradutor guarda uma palavra, sua(s) tradução(ões), exemplos de uso, etc, constituem uma base terminológica.

Existem, então, programas que permitem que o tradutor construa uma ou mais bases terminológicas, sobre um ou mais assuntos, com as línguas que bem entender. E estas bases terminológicas podem ser acedidas sempre que o tradutor bem entender, para saber como traduzir um termo. Os sistemas CAT integram com bases terminológicas para que, mal surja a necessidade de traduzir uma frase que contenha um termo constante da base terminológica, surja também a informação sobre qual a tradução convencional. Isto garante, ainda, que o tradutor seja coerente, e traduza sempre do mesmo modo os mesmos termos.

Na figura 1.2, o processo de recolha de dados, seja pela consulta de bibliografia especializada ou a simples conversa com alguém da área dos documentos a traduzir, é representada pelas diferentes fontes de informação (*Different sources of information*). Os dados recolhidos devem ser introduzidos na base terminológica, processo que está representado nessa mesma figura por *Create*. Finalmente, durante a tradução, essa base terminológica é consultada (*Consult*).

A tradução técnica é, por um lado, desafiante, já que obriga o tradutor a perceber um pouco de tudo. Por outro, é muitas vezes aborrecida, já que um mesmo documento pode, por vezes, repetir-se, ou conter frases de outros documentos semelhantes. Por exemplo, o tradutor que seja encarregue de traduzir os manuais de uso de veículos automóveis de determinada marca, vai deparar-se com grandes porções de texto que são repetidas entre manuais, já que embora os veículos possam ser diferentes, alguns procedimentos são semelhantes.

Os sistemas CAT também ajudam a minimizar este problema. Sempre que o tradutor acaba a tradução de uma frase ou segmento¹, o sistema CAT guarda esse segmento e a sua tradução numa base de dados, chamada de memória de tradução. Isto permite que, sempre que uma nova frase

¹Iremos usar o termo *segmento*, já que muitas vezes se trata de um título, legenda de uma figura, célula de uma tabela, ou outro qualquer excerto que não seja propriamente uma frase.

tiver de ser traduzida, o sistema CAT possa consultar as memórias de tradução e tentar encontrar qualquer outra tradução que já tenha sido feita e que seja semelhante àquela que se pretende realizar. Note-se que não é necessário que o sistema encontre uma frase exatamente igual, já que muitas vezes se traduzem frases muito semelhantes em que apenas uma ou duas palavras variam. Este processo é representado no diagrama pelas setas *Retrieve* (que representa uma tradução que foi encontrada na memória de tradução) e *Save* (que representa o armazenamento de uma nova tradução realizada pelo tradutor).

Para além deste uso direto, de simplificar a vida do tradutor usando traduções já realizadas, as memórias de tradução também podem ser usadas para consultar, por exemplo, como é que determinada palavra foi traduzida em situações anteriores. Claro que sendo um termo de uma área específica, o tradutor poderia ou deveria ter adicionado esse termo na base terminológica. Mas tanto pode não o ter considerado, como pode ter escapado. E durante a altura da tradução, ao verificar a existência de uma palavra estranha que já foi traduzida numa qualquer outra altura, o tradutor poderá usar as memórias de tradução para encontrar exemplos de frases em que essa palavra tenha aparecido, e assim perceber como é que, nessa altura, realizou a tradução. Este mecanismo dos sistemas CAT está representado na figura pela seta *Consult*.

Voltando a olhar para a figura 1.2 poderemos ainda verificar que existe um processo de alimentação das bases de memórias de tradução de que ainda não falamos. Este processo corresponde ao reaproveitamento de documentos paralelos. Suponhamos, por exemplo, a necessidade de traduzir textos da área económica. Antes de proceder à tradução, o tradutor poderia realizar algumas buscas na Internet para obter textos dessa área que já tivessem sido traduzidos. É natural que encontrasse documentos da Comunidade Europeia sobre economia, traduzidos nas línguas em que vai ter de trabalhar. Esses documentos podem ser recolhidos, e utilizado uma aplicação própria (designada por alinhador) capaz de extrair associações entre segmentos nas duas línguas. Poderemos considerar este processo como o reaproveitamento de traduções que já foram feitas, seja pelo próprio tradutor seja por outro, que queremos importar na base de memórias de tradução. Os textos paralelos encontrados (traduções prévias) são alinhados e obtidas as memórias de tradução que poderão ser importadas numa base de memórias de tradução, para que o sistema CAT as possa usar durante a tradução.



2. SmartCAT

“SmartCAT drew our company’s eye right away thanks to its rich technology set. Until recently, consolidating all translation tools with role-based access control was just a dream! Traditional methods without translation memory simply don’t cut it anymore for the linguistic needs of business. We look forward to an integrated, high-tech approach that ensures consistent terminology in all output, centralized task management, and real-time monitoring of task performance”.

Igor Shevshuk, Volkswagen Group

O SmartCAT é um sistema de tradução assistido por computador da empresa Abbyy, que funciona na Internet, como aplicação *web*. É gratuito, sendo apenas necessário o registo prévio. Desse modo, para seguir as instruções deste capítulo, é necessário o acesso a <http://smartcat.pro/> e o registo. Este pode ser feito usando um e-mail e um palavra chave específicos para este sítio *web*, ou utilizando informação de autenticação de outras aplicações como o *Facebook*, *Google* ou *LinkedIn*.

2.1 Primeiro Projeto de Tradução

Depois de registado no SmartCAT, é possível a autenticação. Feita a autenticação o utilizador é direcionado para uma zona de controlos, denominada *dashboard*. Aqui é possível a alteração da informação do seu perfil (como por exemplo o seu nome) ou a definição de serviços de tradução, permitindo que o tradutor use esta aplicação de forma comercial.

Para experimentar o nosso primeiro projeto de tradução iremos utilizar o menu *My Projects*, no topo, junto ao logótipo do SmartCAT. Surgirá uma página semelhante à apresentada na figura 2.1.

Um projeto pode compreender a tradução de um ou mais documentos, entre uma ou mais línguas. Poderá considerar-se que cada projeto corresponde a cada tarefa independente que o tradutor terá de executar, seja para clientes diferentes, ou para o mesmo cliente, em alturas diferentes.

A figura 2.2 apresenta o primeiro passo na criação de um novo projeto de tradução (depois de usar o botão *Create Project*). Neste passo a nossa primeira tarefa é a adição dos documentos que

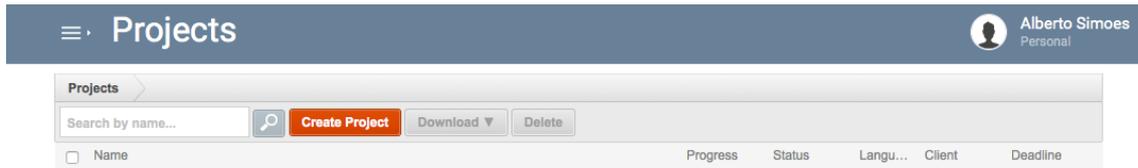


Figure 2.1: Janela de Projetos do SmartCAT.

pretendemos traduzir. Mais tarde será possível adicionar mais documentos, mas o sistema dá-nos já a possibilidade de os seleccionar.

Figure 2.2: Criação de um novo projeto (passo 1).

Seguem-se dois campos menos úteis do ponto de vista académico, mas bastante relevante para o tradutor: a data limite em que a tradução deverá estar pronta, e o nome do cliente que solicitou a tradução.

Mais relevante é a escolha das línguas. Em primeiro lugar é necessário indicar a língua de origem (e note-se que um projeto só pode ter uma única língua de origem), e em segundo lugar o conjunto de línguas de destino.

Seguem-se outros dois campos descritivos, um com o título do projeto de tradução e um outro com uma descrição. Mais uma vez estes campos não são úteis propriamente para a realização da tradução, mas são muito importantes para a gestão de projetos de tradução.

Mais abaixo é possível configurar o sistema para usar tradução automática quando não encontra uma tradução humana nas memórias de tradução. Dependendo do tradutor e do tipo de documento, a tradução automática tanto poderá servir para agilizar o processo de tradução, e portanto ser benéfica, como poderá apresentar sistematicamente traduções erradas, tornando-se assim contra-protutiva.

Antes de terminar a janela, existe ainda um botão para configurar diferentes tipos de validação da qualidade de tradução. Basicamente, o SmartCAT consegue detetar um conjunto de situações que podem indicar a má qualidade da tradução, e é possível configurá-lo para que faça avisos sempre que alguma tradução não cumpra os requisitos mínimos de tradução.

Para este exercício vamos usar um documento bastante simples disponível em <http://natura.di.uminho.pt/~ambs/tti2015/2translate.txt>. Armazene-o no seu computador, e adicione-o ao SmartCAT usando o botão *Add*, no topo da janela. Selecione a língua inglesa como língua

de origem, e a língua portuguesa como língua de destino. Dê um nome ao projeto, por exemplo, “Primeiro Projeto”. Ative, também, a tradução automática. Embora existam mais opções que veremos mais tarde, vamos terminar a criação do projeto usando o botão *Finish*. A figura 2.3 mostra a lista de projetos após a criação do novo projeto, depois de expandido (clcando no botão ‘+’ junto ao nome do projeto).



Figure 2.3: Janela de Projetos do SmartCAT com o primeiro projeto.

Nesta posição é possível aceder à página de configuração do projeto (clcando no nome dele), adicionar novos ficheiros a serem traduzidos, descarregar algum dos documentos, ou analisar ou apagar um documento. Para iniciar a tradução deveremos clicar no nome do documento em causa (neste caso, o *2translate*). Será iniciado o interface de tradução numa nova janela, semelhante ao apresentado na figura 2.4.

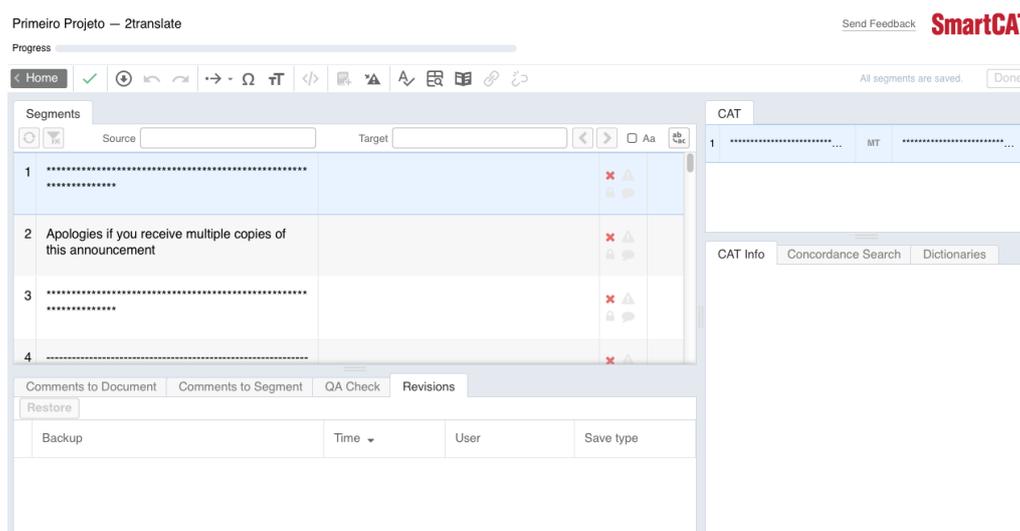


Figure 2.4: Interface de tradução SmartCAT.

A janela principal, denominada *segments*, apresenta cada um dos segmentos que deve ser traduzido. A forma como o sistema dividiu o texto em segmentos não é propriamente relevante, e grande parte das vezes pode ser completamente ignorada. O processo de tradução pode ser começado, colocando na coluna que se encontra em branco a tradução de cada um dos textos da coluna à esquerda. Como se pode ver na figura 2.4, o primeiro segmento é uma sequência de asteriscos que, como se pode imaginar, não tem tradução. No entanto, é importante que o documento resultante da tradução mantenha a estrutura do original, e portanto, que estas linhas sejam mantidas. Existe, então, um conjunto de diferentes formas para copiar a linha de asteriscos, desde contar quantos são; selecionar os da esquerda, copiar e colar; ou mais simplesmente, usar o botão da barra de ferramenta. Este botão também pode ser ativado usando a combinação de teclas **Ctrl+Insert**. 



Cada tradução realizada tem de ser confirmada. Para isso pode ser usado o botão com um visto ou usando a combinação de teclas `Ctrl+Enter`. Assim que o pressionar irá reparar que o terceiro segmento, também composto apenas por asteriscos, será automaticamente traduzido. Isto deve-se ao facto de, ao confirmar a tradução, a termos gravado na memória de tradução, permitindo à aplicação o seu uso para automatizar o processo de tradução.

Se o mecanismo de tradução automática tiver sido ativada, aquando da criação do projeto, no momento em que confirmamos a tradução do primeiro segmento, uma janela do lado direito irá mostrar o segundo segmento com uma tradução. Entre as duas aparecerá a sigla MT (Machine Translation) que indica que aquela tradução foi obtida usando tradução automática e portanto merece cuidados extra de revisão. Para a podermos usar como base da nossa tradução basta fazer duplo clique sobre o segmento ou usar o atalho `Ctrl+1`. Posteriormente o tradutor deve rever a tradução, tornando-a correta, e só depois a deve confirmar.

Repare-se que, depois de confirmar o segundo segmento, o tradutor é colocado no terceiro segmento, que supostamente já estaria traduzido. O funcionamento de um sistema CAT é, habitualmente, este, fazendo com que o tradutor percorra todo o documento, mesmo para aquelas frases que foram encontradas na memória de tradução. Isto permite que o tradutor possa fazer pequenas alterações que possam ser necessárias (por exemplo, no sentido de tornar a tradução menos repetitiva).



Neste momento vamos dar atenção a uma pequena imagem de alerta, amarela, que surge junto da tradução do primeiro e do terceiro segmento. Este alerta estará aceso sempre que o sistema achar que a tradução é duvidosa. Neste caso, ao passar o rato por cima do sinal, poderemos reparar que a tradução foi sinalizada pela tradução ser semelhante ao texto original.

Podemos confirmar a tradução e continuar com a tradução seguinte. O segmento seguinte é, mais uma vez, uma linha de símbolos repetidos, que deve ser copiado para a língua de destino. Já o quinto segmento é um título, e aparece todo em maiúsculas. Note também que a própria sugestão do sistema de tradução automática também se encontra com todos os caracteres em maiúsculas. O tradutor deve ter em atenção estes aspetos, mantendo sempre que possível o aspeto gráfico do documento original no documento traduzido.

Realize a tradução dos segmentos 5, 6 e 7, confirmando-as. O oitavo segmento é um endereço de uma página *web*. Como regra geral, nunca se deve editar um endereço (seja de página web ou de e-mail). Em alguns casos específicos pode haver necessidade de alterar o endereço, para redirecionar para uma página numa língua diferente. No entanto, nessas situações o tradutor será devidamente avisado. Copie, pois, o endereço para a língua de destino e confirme a tradução.

Continue a tradução até ao segmento 16, inclusive, confirmando cada uma delas. Quando chegar à tradução do segmento 17 irá reparar que o sistema irá apresentar duas traduções possíveis, na janela do lado direito. Se por um lado sugere a tradução obtida automaticamente (marcada com MT), também sugere uma outra, obtida pela consulta da memória de tradução (marcada com TM, abreviatura de Translation Memory, e com uma percentagem de precisão). Esta sugestão surge porque já foi realizada uma tradução semelhante (segmento 15), embora com uma palavra diferente. Nestes casos o tradutor pode olhar para a percentagem para ter uma ideia de quão útil será a tradução obtida. Repare-se que neste caso a tradução obtida pela memória de tradução está parcialmente correta: apenas o mês está errado.