

QUAH, Chiew Kin. *Translation and Technology*. Basingstoke etNew York, Palgrave Macmillan, 2006.



Marc Pomerleauⁱ
(Doutorando em Tradução - Département de Linguistique et de Traduction -
Université de Montréal - Montreal, Quebec, Canadá)
marc.pomerleau@umontreal.ca

Traduzido por:

Patrícia Rodrigues Costaⁱⁱ
(Doutoranda em Estudos da Tradução – PGET/UFSC, Brasil)
prcosta1986@gmail.com

Rodrigo D'Avila Braga Silvaⁱⁱⁱ
(Mestrando em Estudos da Tradução, POSTRAD/UnB)
rodrigodavilabraga@gmail.com

225

Chiew Kin Quah nos apresenta aqui o que poderíamos considerar um manual do tipo *guia* sobre tecnologias da tradução, tudo para o benefício das partes interessadas no mundo da tradução. Desde o início, Quah diz aos leitores que a obra visa preencher o vazio na área uma vez que há pouca literatura disponível sobre as ferramentas de apoio à tradução. Contudo, ela assegura que seu livro pretende apresentar uma visão simples, em nossa opinião grande o suficiente, a respeito da área das tecnologias da tradução. Os especialistas da área talvez estejam de acordo sobre esta questão, mas para a grande maioria dos leitores, que são familiarizados ou não com essas tecnologias, este livro será muito mais que uma simples visão. De fato, a autora apresenta um vasto panorama acerca das tecnologias da tradução, começando por um histórico da pesquisa e das descobertas neste campo. Em seguida, ela aborda a definição dos termos, a integração das tecnologias no campo de estudos da tradutologia, os diferentes estilos, as ferramentas de avaliação e o desenvolvimento atual e futuro. Quah conclui o livro propondo um novo modo de classificar as diversas ferramentas informáticas disponíveis no mercado.

No primeiro capítulo, a autora nos conta que a primeira terminologia utilizada para se referir às tecnologias da tradução não era uniforme; os tradutólogos usavam determinados termos e os tradutores utilizavam outros. Além disso, os termos usados pelos profissionais da indústria e da localização e os da informática nem sempre eram os mesmos, ou, mesmo que fossem eles divergiam em alguns pontos. Em suma, Quah afirma no início que ela utilizará essencialmente o termo genérico *computer-aided translation*, isto é, tradução *assistida por computador*.

Em relação ao breve histórico, a autora afirma que no início as pesquisas sobre tradução automática objetivavam essencialmente conceber um sistema que permitisse traduções de qualidade, sem qualquer intervenção humana. Contudo, na década de 1950, os pesquisadores perceberam que este objetivo era inalcançável. Eles começaram a procurar soluções intermediárias entre a tradução completamente automática e a tradução humana, esta última referindo-se ao bom e velho método de tradução “à mão”. Estas soluções intermediárias – ou paliativas – são a “tradução automática assistida por um humano” e a “tradução humana assistida por um computador”. No primeiro caso, o computador está em primeiro plano e realiza a maior parte da tradução; o ser humano, por outro lado, é responsável pela preparação do texto e a revisão após a produção da tradução inicial. No caso da “tradução humana assistida por computador”, o tradutor usa as ferramentas informáticas necessárias, por exemplo, os dicionários eletrônicos, dos corretores, os bancos de dado e os softwares de ajuda à tradução do tipo SDL Trados.

No segundo capítulo, Quah faz uma retrospectiva eloquente sobre a história da tradução e uma reflexão sobre a área, desde Cícero à atualidade, passando por São Jerônimo, Viney e Darbelnet, Benjamim, Chesterman e outros. Esta observação, baseada no esquema de Holmes e na versão reformulada deste esquema pela autora, permite o leitor compreender a inserção e posição atual da tradução automática na tradutologia. A autora também tem o cuidado de incluir uma breve descrição da relação desenvolvida entre a linguística e a informática, que são a bases da pesquisa sobre o processamento automático da linguagem humana. Na verdade, os profissionais da tecnologia da linguagem tentaram primeiramente estabelecer regras que permitiam analisar as línguas naturais a fim de introduzi-las em programas de computador. Alguns pesquisadores têm utilizado uma abordagem formal com base na descrição da estrutura gramática das línguas, enquanto outros adotaram uma abordagem funcionalista enfocando em como combinar palavras e frases para produzir textos compreensíveis.

Quah nos fala sobre várias intervenções humanas que podem facilitar o processamento computacional da linguagem, a primeira sendo a pré-edição. Isso limita o vocabulário a mais simples unidade vocabular e gramatical. Esta operação é particularmente útil quando a tradução é multilíngue. Em seguida, vem a pós-edição que consiste em melhorar o texto produzido por um sistema computacional. Ela pode ser rápida e visar à produção de um texto correto no qual o estilo não precisa o mais refinado, o produto final é essencialmente “apresentável”. Nos demais casos, a pós-edição pode objetivar a produção de um texto de alta qualidade. O discurso mais interessante que apresentado por Quah diz respeito sobre o controle da linguagem. Entendemos por “linguagem controlada” um subconjunto de uma língua natural, no qual limitamos delimitamos voluntariamente o vocabulário, a gramática e o estilo. Em geral, um editor treinado para este fim redige o texto em função dos critérios pré-determinados, por exemplo, quanto ao seu tamanho (curto) e à precisão das frases, a eliminação da polissemia e da sinonímia. O objetivo deste controle é reduzir as ambiguidades e as dificuldades do texto de partida a fim de melhorar o resultado da tradução automatizada. Por outro lado, isto permite reduzir o volume de trabalho na pós-edição. Mais uma vez, Quah fornece ao leitor tabelas práticas que permitem comparar textos em inglês padrão e sua versão controlada. Isto constitui um forte complemento útil às descrições. Assim, podemos pensar que a linguagem controlada é uma nova tendência, mas a autora nos informa que diversos projetos deste gênero têm sido desenvolvidos desde a década de 1930, incluindo o *BASIC English*, o *Caterpillar Fundamental English* e o *AECMA Simplified English*, alguns destes projetos utilizam cerca de 1% da língua inglesa. Embora a maior parte dos projetos de linguagem controlada relacionadas à língua inglesa, outras línguas também foram controladas deste mesmo modo, com o alemão, o espanhol e o chinês. No final do capítulo, a autora adverte os tradutores: eles devem trabalhar mais com esse tipo de linguagem.

O terceiro capítulo também tem um curto histórico sobre a tradução automática, que inicia com a proposta formulada em 1949 por Warren Weaver para o uso de técnica de criptografia por mecanismos de tradução, porém não foi bem sucedida. A autora nos informou que no auge da Guerra fria, isto é, entre a década de 1950 e 1960, as pesquisas da área foram financiadas principalmente pela CIA e pela KGB. A década de 1970 foi um período de calmaria para o setor. Os avanços na linguística computacional da década de 1980 deram um novo impulso às pesquisas em tradução automática, apoiadas por grandes empresas como IBM e Apple. Desde então, pesquisas e avanços em tradução automática têm se multiplicado.

Após ter concluído o panorama histórico, Qual inicia a descrição de diferentes tipos de estilo. Ela se concentra em dois componentes principais dos sistemas de tradução automática, sejam os dicionários ou o analisador sintático. Enquanto que a descrição do primeiro destes componentes seria supérflua, exceto por dizer que os sistemas automáticos contêm em geral dicionários monolíngues e bilíngues, sendo que o segundo requer um pouco mais de esclarecimentos. Um analisador sintático, conhecido em inglês por *parser*, é uma ferramenta que analisa os textos a serem traduzidos e os dividem em unidades sintáticas. Estas unidades são geralmente constituídas por uma sequência de palavras ou de uma frase. Portanto, a partir desta sequência, o analisador determina os lugares ou relações entre as palavras e seus representantes estruturais, que permitem ao sistema obter a informação gramatical necessária para obter o significado e produzir posteriormente uma tradução de qualidade.

Quah nos informa, em seguida, que os primeiros sistemas de tradução automática funcionaram por tradução direta, isto é, essencialmente palavra por palavra, o que a priori não surpreende. Conforme com a autora, esta metodologia é, sobretudo, útil para pares de línguas relativamente próximas, por exemplo, o francês e o italiano. A tradução direta baseia-se em duas abordagens fundamentadas em regras, seja a abordagem intralingual seja a abordagem por transferência. A abordagem intralingual consiste uma utilizar uma representação abstrata da língua fonte que codifica as informações sintáticas e semânticas de modo a transcodificar no texto-alvo. Esta abordagem visa produzir uma intralíngua, uma espécie de passagem obrigatória entre as línguas envolvidas na tradução. Evidentemente, esta é uma abordagem idealizar que dificilmente pode ser aplicada a todas as línguas, em particular devido à quase impossibilidade de chegar a uma representação interlingual que considera todas as possibilidades sintáticas e semânticas de todas as línguas do mundo. Notamos, porém que esta abordagem ainda é explorada por muitos pesquisadores. A abordagem de transferência visa também converter o texto fonte em uma representação abstrata, mas sem passar por uma interlíngua. O texto fonte é mais comumente transferido diretamente para a língua de chegada com a ajuda de um dicionário bilíngue acompanhado de uma boa gramática. Segundo esta abordagem, cada par de línguas possui seu próprio modelo de transferência. O principal desafio encontra-se no uso de dicionários bons o suficiente para conseguir superar a ambiguidade do significado. Além disso, deve-se elaborar um modelo para cada par de línguas, o que exige recursos consideráveis.

A partir da década de 1990, os pesquisadores começaram a usar as abordagens com base em corpus. Isto é, recolher uma quantidade exponencial de textos já traduzidos e os inserir em bases de dados em função de pares de línguas específicos. Este sistema utiliza ainda de traduções-modelo feitas por humanos de modo a produzir novas traduções. Mais uma vez, há duas abordagens, esta baseada em corpus: a abordagem baseada em estatísticas e a abordagem baseada em exemplos. A abordagem baseada em estatística se fundamenta no teorema de Bayes, o qual utiliza as probabilidades e os conhecimentos adquiridos em eventos anteriores. No caso da tradução, utilizamos as traduções antigas juntamente com os textos de partida e a combinação de segmentos para produção de novos textos de chegada. Um dos sistemas estatísticos mais conhecidos, o projeto *Candide*^{iv}, que contém mais de dois milhões de frases coletadas nos debates parlamentares do governo do Canadá, conhecido por *Hansards*. Essa abordagem requer, obrigatoriamente, uma grande quantidade de textos previamente traduzidos no mesmo par de línguas. A abordagem baseada em exemplos fundamenta-se também em um corpus bilíngue de textos traduzidos. De acordo com esta abordagem, um algoritmo é utilizado para localizar segmentos traduzidos, por exemplo, que se assemelham a segmentos a serem traduzidos. Portanto, pode ser que o sistema tenha que combinar diversos segmentos de modo a produzir uma tradução completa do segmento desejado. Esta abordagem é semelhante àquela utilizada por ferramentas de ajuda à tradução como o *SDL Trados*. Contudo, esta abordagem é totalmente automática, enquanto que ferramentas do tipo *SDL Trados* requerem intervenções do tradutor. Além disso, também existem os sistemas híbridos que combinam estas abordagens.

No quarto capítulo, Quah nos fala sobre ferramentas, recursos e normas para a tradução assistida por computador. Enquanto algumas ferramentas são destinadas a usos comerciais, o autor nos informa a respeito das ferramentas que podem ser úteis aos tradutores profissionais, indicando que essas ferramentas podem permitir-lhes aumentar a sua produtividade, melhorando a qualidade seus textos. A maioria dos softwares de tradução assistida por computador destinada aos profissionais funciona a partir de um pacote integrado a um mesmo software de processamento de texto, como no caso de *Trados Workbench*. Muitos destes sistemas, muitas vezes chamados de "memória de tradução" foram desenvolvidos na década de 1990. Eles permitem aos tradutores arquivar textos traduzidos e reutilizá-los para outras traduções. Os textos são armazenados em segmentos alinhados, e posteriormente, recuperados pelo software, quando um segmento novo a ser traduzido é idêntico ou quase idêntico a um segmento já traduzido e

armazenado no banco de dados. O usuário pode geralmente configurar o sistema para recuperar os segmentos antigos com base em uma porcentagem de semelhança com os novos segmentos a serem traduzidos (por exemplo, 100% idêntico, 90% ou mesmo 50%). Obviamente, quanto menos antigo o segmento em relação ao novo, mais o usuário deve fazer alterações nos segmentos de texto-alvo. O tradutor também pode modificar um segmento idêntico ao antigo, dependendo do contexto ou necessidades, ou simplesmente para fazer correções.

Os softwares de tradução assistida por computador utilizam etiquetas invisíveis na tela para codificar o texto. No final da sua tradução, o usuário deve realizar uma etapa de "limpeza" para remover a codificação. Isso geralmente é feito com um único comando. Se precisar, o usuário também pode adicionar os textos já traduzidos ao seu banco de dados com a ajuda de um comando de alinhamento de textos, que é útil quando o cliente fornece textos de referência já traduzidos ou se o tradutor descobre textos bilíngues sobre o mesmo tema na internet ou em outro lugar. Há também sistemas de gestão terminológica que permitem extrair termos a partir de textos de referência, que podem constituir uma verdadeira mina de ouro para o tradutor. Este último pode realmente extirpar os termos de uma área de especialidade a partir de textos previamente traduzidos para usá-los em uma nova tradução. Outras ferramentas incluem os corpora paralelos, que são coleções de textos eletrônicos, e as concordâncias, que mostram termos pesquisados em seu contexto, também a partir de uma base de textos pré-selecionados. Os softwares de reconhecimento ótico de caracteres (OCR), por sua vez, permitem converter em formato eletrônico documentos digitalizados. O reconhecimento nem sempre é perfeito, mas evita que o tradutor tenha que recopiar textos longos. Segundo Quah, essas ferramentas não são amplamente utilizadas pelos tradutores profissionais que ganhariam em conhecê-las em alguns casos. Neste sentido, a indústria deveria implantar esforços de sensibilização. Notemos que algumas ferramentas são relevantes apenas para os tradutores que trabalham no campo da localização. Estas facilitam a tradução de softwares, sites e qualquer coisa que você encontrar na web.

Parece também que a indústria tem feito progressos consideráveis na harmonização das normas de intercâmbio de dados. De fato, um dos principais problemas está na ampla variedade de formatos utilizados pelos sistemas operacionais. Este problema também é comum em muitos subcampos da Informática. Em suma, nos últimos anos, a indústria tem procurado fornecer alguma interoperabilidade, graças à introdução de normas, elas caem no âmbito da Organização Internacional de Normalização (ISO) e de outras organizações industriais ou comerciais. A

harmonização dos formatos permite aos usuários, trocar (importar e exportar) bancos de dados, como memórias de tradução, qualquer que seja o sistema de apoio à tradução utilizado.

O quinto capítulo aborda o tema da avaliação das ferramentas de tradução automática e de tradução assistida por computador. Os métodos de avaliação variam consideravelmente em função dos intervenientes. De fato, os pesquisadores, desenvolvedores, tradutores e outros usuários não avaliam os sistemas necessariamente da mesma forma. Além disso, devido à grande quantidade de ferramentas no mercado, é difícil padronizar qualquer tipo de avaliação. Por um lado, os desenvolvedores querem demonstrar que o seu sistema é o melhor e que os usuários devem comprá-los. Por outro lado, os usuários querem saber qual sistema melhor atende às suas necessidades, melhorar o rendimento e a qualidade do seu trabalho.

Existem várias maneiras de avaliar os sistemas e as modalidades dependem principalmente do objetivo perseguido, elas podem ser feitas automaticamente ou pelos usuários. As avaliações efetuadas exclusivamente por seres humanos requerem a implantação de muitos recursos, incluindo o investimento de tempo e dinheiro. Além disso, toda avaliação humana é necessariamente subjetiva, pelo menos em parte, mesmo que as regras e instruções de avaliação sejam rigorosas. Os pesquisadores estão à procura de métodos para realizar avaliações, parciais ou completas, a partir dos sistemas computacionais. Dentre os vários métodos de avaliação propostos, notemos a norma ISO 9126, que baseia a avaliação em seis características chaves: funcionalidade, fiabilidade, usabilidade, eficácia, manutenção e portabilidade. Muitos outros projetos e programas de avaliação têm surgido nos últimos anos, incluindo EAGLES, ISLE, FEMTI e NIST, mas de acordo com o autor nenhum deles conseguiu ainda estabelecer um quadro avaliação completo.

No sexto capítulo, Quah nos informa sobre os últimos desenvolvimentos e tendências na área das tecnologias de tradução. No caso da tradução automática, o autor acredita que os sistemas serão cada vez mais numerosos, isso por causa dos avanços tecnológicos e o aumento da demanda por traduções. O aumento do uso da Internet é a principal razão: existem cada vez mais textos a traduzir e os prazos de tradução são cada vez mais curtos. Além disso, a variedade de línguas na internet hoje é muito mais importante que há alguns anos atrás. Para atender a essas necessidades, Quah propõe uma pirâmide que leva em conta os tipos de textos e a qualidade requisitada. Nós, portanto, encontramos na parte inferior da pirâmide os documentos menos "importantes", como artigos da web, que requerem menos ajustes, dado que na maioria dos casos,

o leitor simplesmente pesquisa informações e tem pouco interesse pelo estilo. Uma tradução totalmente automática é então a privilegiada. No topo da pirâmide, encontramos os textos mais criativos ou técnicos, tais como os de publicidade ou de natureza jurídica. Nesses casos, uma grande intervenção humana é necessária. No curto prazo, por outro lado, o autor não acredita que a máquina vem para substituir o ser humano na tradução de textos altamente especializados, seja no campo técnico ou na literatura. Quah também acredita que os sistemas são chamados a se especializar em determinados pares de línguas, em vez de visar o mercado mundial de línguas. Ela também acha que para obter textos de qualidade, os desenvolvedores deverão produzir sistemas que se especializem na tradução de textos de áreas específicas. Em suma, os sistemas de tradução automática são chamados a buscar nichos de mercado. Em termos de tradução assistida por computador, parece que os conjuntos do tipo Workbench manterão o placar, mas os diversos softwares serão cada vez mais concorrentes. Os usuários também terão alternativas à sua disposição, tais como ferramentas de reconhecimento de voz.

232

Para concluir a propósito de este assunto, parece que o futuro reserva boas notícias para as línguas minoritárias, até agora amplamente ignoradas pelos desenvolvedores de ferramentas de tradução. As pesquisas já produziram resultados na Ásia, por exemplo, o sistema Anusaaraka, especializado nas línguas da Índia. Outros sistemas do gênero visam o crioulo haitiano, o galego, o aimará e o quéchua, entre outros.

No capítulo sétimo, sob a forma de conclusão e de retorno aos seis primeiros capítulos, Quah apresenta em forma de quadro uma nova forma de categorizar os diferentes tipos de tradução em uma ruptura do modelo linear de Hutchins e Somers. Num total de vinte, os quadros apresentam o que o autor chama de *strands*, que poderíamos traduzir por *fios condutores*. O autor, em seguida, aplica cada um de seus fios (automação, a intervenção humana, normas, avaliação, etc.) aos quatro tipos de tradução citados por Hutchins e Somers que são: tradução automática, tradução automática assistida por humanos, tradução humana assistida por computador e tradução humana. Os quadros de Quah são certamente exaustivos, mas eles são bastante complexos. Em suma, eles constituem um jeito novo e audacioso de classificar os sistemas que irá agradar aos pesquisadores da área. No entanto, eles são, sem dúvida, demasiado pesados para descorticar para o simples usuário.

Em conclusão, alguns argumentam que o aspecto humano é o mais tradicional da tradução e é ignorado neste livro. Em nossa opinião, este não era o propósito de Quah. Ele fornece um

panorama interessante da área, bem apoiado e enriquecido de noções históricas e terminológicas que permitem ao leitor a seguir o fio de suas afirmações e argumentos. E para tranquilizar os tradutores que ainda veem as máquinas como potenciais concorrentes que poderiam substituí-los num futuro não tão distante, nós tomamos as palavras de Tsujii (citado por Quah), que indica que a automação completa da tradução é um dos sonhos mais difíceis de realizar nas ciências da informação, e que, apesar dos progressos realizados, não há nada que permita acreditar que isso se concretizará num futuro iminente: está muito mais próximo em uma maior colaboração entre homem e máquina.

BIBLIOGRAFIA

Benis, Michael (2008). Compte rendu du livre *Translation and Technology* de Chiew Kin Quah (Basingstoke, Palgrave Macmillan, 2006) dans *The Interpreter and Translator Trainer* (ITT) vol. 2, n° 2, pp. 261-263.

Kenny, Dorothy (2007) Compte rendu du livre *Translation and Technology* de Chiew Kin Quah (Basingstoke, Palgrave Macmillan, 2006) dans *Machine Translation* vol. 20, n° 4, pp. 291-295.

233

RECEBIDO EM 16/06/2014

ACEITO EM 20/07/2014

ⁱ Marc Pomerleau. <http://mpom.ca/>

ⁱⁱ Lattes Patrícia Rodrigues Costa. Disponível em: <http://lattes.cnpq.br/9546437584230118>

ⁱⁱⁱ Lattes – Rodrigo D’Avila Braga Silva. Disponível em: <http://lattes.cnpq.br/1746993519090773>

^{iv} NdeT: **Statistical Machine Translation**. Disponível em: <http://www.cs.cmu.edu/~ab Berger/mt.html> Acesso: julho 2014.