

A NOVA TORRE DE BABEL
– QUE FUTURO PARA A TRADUÇÃO AUTOMÁTICA?

Sara Cerqueira

E o Senhor disse: «Eles constituem apenas um povo e falam uma única língua. Se principiarem desta maneira, coisa nenhuma os impedirá, de futuro, de realizarem todos os seus projectos. Vamos, pois, descer e confundir de tal modo a linguagem deles que não se compreendam uns aos outros». E o Senhor dispersou-os dali para toda a face da terra, e suspenderam a construção da cidade. Por isso, lhe foi dado o nome de Babel, visto ter sido lá que o Senhor confundiu a linguagem de todos os habitantes da terra.

Génese, 11, 1-9

Poderá, um dia, a tradução automática suprimir ou dissipar as diferenças que bloquearam e obstruíram a comunicação na mítica torre babeliana? Será a tradução automática capaz de diluir as diferenças linguísticas que separam as comunidades humanas? Estas perguntas são, provavelmente, as que ocorrem ao público em geral, quando ouve falar de uma nova descoberta em Inteligência Artificial ou de uma nova aplicação informática para tradução. A curiosidade e o entusiasmo do público não-especialista contrasta, não raras vezes, com algum ceticismo e desconfiança de tradutores e teóricos da tradução¹.

A proliferação do acesso a tradutores automáticos disponíveis na Internet conduziu, por um lado, à disseminação da obtenção de informação e obriga, por outro, a uma requalificação e re-orientação de formação da profissão de tradutor, que terá, necessariamente de passar por uma abordagem destas novas tecnologias. Este imperativo, nem sempre é bem aceite pelos profissionais de tradução. Considerar, porém, a TA utópica, arrogante e insolente, só porque os resultados que se obtiveram são nitidamente incipientes e até mesmo anedóticos, não nos deve fazer voltar as costas a uma realidade que não só existe, como tem também provas dadas em vários campos. Ver a TA como uma ameaça para o futuro dos tradutores e a virtualização do trabalho de tradução como um monstro desumanizador, só encurta os horizontes de conhecimento e, logo, uma das qualidades imperativas de um bom tradutor, a assimilação de conteúdos e a adaptação a uma realidade em constante mudança. “It is clear that machine translation will play an important role in the new millennium, helping to bring down the communication barriers in the newly interconnected

world. It is up to us translators to explain to the general public what machine translation is, what are its strengths and weaknesses, and what is its likely role in the future development of our civilization” (Vitek, 2000: 7). Um uso profissional e metodologicamente correcto da TA requer, aliás, as competências de um tradutor humano e especializado, uma vez que, como adiante se dirá, as mais recentes aplicações de TA incluem avanços qualitativos, tais como a pré e pós-edição.

Hoje em dia, não se pode descurar o papel das novas tecnologias aplicadas ao processo tradutivo, nomeadamente as mais recentes aplicações informáticas, arriscando-se o tradutor, se o fizer, a tornar-se obsoleto e inadaptado. A principal tarefa do tradutor só poderá, doravante, ser compreendida num contexto de globalização, especialização e digitação e requer um novo paradigma teórico-prático onde deverá, obrigatoriamente, constar a incorporação dos recursos da linguagem electrónica e das novas ferramentas de tradução assistida por computador. O uso da Tecnologia de Informação e Comunicação (TIC) é, nos nossos dias, um *fait accompli*. E não são somente os prazos e as pressões editoriais que convencem os tradutores a um uso extensivo das Tecnologias da Informação e Comunicação. A exposição a uma metodologia correcta no uso de ferramentas electrónicas acarretará uma automatização do trabalho de tradução, que apoia os tradutores na incrementação exponencial da qualidade e eficiência dos seus serviços.

São, pois, dois os factores a considerar na abordagem do uso da TIC:

- melhoria da qualidade;
- aumento da produtividade.

Urge, pois, descobrir as várias aplicações existentes, bem como compreender as suas utilizações actuais e os seus desafios futuros. Para este efeito, ir-se-á, num primeiro momento, recuar no tempo e conhecer um pouco da história e evolução da TA, para, de seguida, se proceder a uma tipologização das diferentes ferramentas de TA.

BREVE HISTORIAL DA TRADUÇÃO AUTOMÁTICA

O primeiro marco da história da tradução automática (TA) deve datar-se no início dos anos 40, quando a recém-fundada ciência informática se propôs como primeira tarefa a de desenvolver aplicações de tradução automática, potenciada pela explosão na transmissão de informação e pela errónea facilidade de decalcar uma técnica humana aparentemente simples: a tradução.

O início da guerra-fria iria, em 1946, dar o primeiro impulso a este processo, sustentado pela necessidade de obter informações soviéticas à distância, da forma mais rápida e eficiente possível. O inglês Booth e o americano Warren Weaver iriam entrar na história como os percussores da tradução automática, devendo-se-lhes a criação de uma calculadora científica com dados suficientes para fazer tradução palavra por palavra, alheada de qualquer tipo de consideração sintáctica ou da ordem lexical (Alfaro, 1998:05). Weaver, antigo presidente da Fundação Rockefeller, concebia a TA de forma bastante simplista, ao sustentar que o processo de tradução seria basicamente análogo ao processo de descodificação de códigos para fins militares. Estava largamente convencido de que esta modalidade era um objectivo facilmente alcançável, dando origem a um movimento de crescente interesse pela TA, bem como à fundação de diferentes grupos de investigação nos EUA e na Europa.

Em 1948, o inglês Richens introduziria melhorias na máquina desenvolvida por Booth e Weaver, nomeadamente informações relativas à análise gramatical das desinências russas. Em 1950, Weaver proporia a exploração automática do contexto terminológico, visando solucionar as ambiguidades semânticas. Desde esta data, com reflexões como as de Reifler, que insistia na necessidade de preparação prévia dos textos submetidos a tradução automática, começa a delinear-se um paradigma de reflexão que torna imperativos a pré-edição, o auxílio humano durante a tradução e/ou a revisão de textos.

Ainda durante a década de 50, várias investigações envidam esforços no sentido de desenvolver a tradução automática. A primeira conferência sobre TA, realizada no Massachusetts Institute of Technology, dava grande azo ao optimismo e ao entusiasmo característicos desta primeira fase de pesquisas. Os trabalhos concluiriam da necessidade primordial de orientar os paradigmas metodológicos lexicalmente, enquanto as análises sintáctica e semântica do texto eram relegadas para um plano secundário. Até aos anos 50 e inícios dos anos 60, os investigadores estavam plenamente convencidos que seria possível desenvolver sistemas que produzissem “Fully automatic high-quality machine translation” (FAHQMT). Rapidamente se perceberia que nem tudo seria tão fácil.

Chegados aos anos 60, já o desânimo e o cepticismo se tinham apoderado mesmo dos que se mostraram mais optimistas. “As aplicações práticas não correspondiam às previsões teóricas e a linguística formal não conseguia explicar problemas ligados a estruturas, processos, funções e formas que se multiplicavam” (Alfaro, 1998:07). As traduções palavra por palavra não conseguiam produzir resultados inteligíveis e a formalização das regras

sintácticas a partir das gramáticas não era suficiente para abarcar todos os aspectos linguísticos observáveis. O cientista norte-americano Bar-Hillel será uma das vozes mais críticas das orientações das pesquisas da época, sublinhando o facto de a resolução das ambiguidades semânticas patentes nas diferentes línguas só ser possível graças à introdução de quantidades inestimáveis de conhecimentos enciclopédicos. É famoso o exemplo dado por Bar-Hillel, ao apontar os problemas que um programa de TA encontraria ao traduzir frases como: “Little Peter was looking for his toy box. The box was in the pen”. Bar-Hillel argumenta que esta passagem será correctamente interpretada somente quando se tem o conhecimento do tamanho típico de canetas e caixas para que se reconheça a impossibilidade de se colocar uma caixa dentro de uma caneta. Assim, recorre-se a um contexto infantil, onde “pen” se refere a “playpen”. Segundo Bar-Hillel, um computador nunca poderia ser provido de conhecimento suficiente para lidar com este tipo de problema; os objectivos das pesquisas em TA deveriam, por consequência, ser mais modestos (Bar-Hillel, 1964).

A desilusão definitiva virá com o relatório ALPAC (*Automatic Language Processing Advisory Committee*) – encomendado pelos principais investidores norte-americanos – que, em linhas gerais, concluiu da ausência de necessidade da TA, dada a relativa inexistência de procura, negando, igualmente, a futura redução efectiva dos custos e a improbabilidade imediata de a TA produzir traduções de textos de linguagem geral sem a intervenção humana. Estas conclusões, embora parciais e tendenciosas, conduziram ao descrédito da TA e as verbas governamentais para o desenvolvimento de investigação sofrerão cortes radicais. Efectivamente, facilmente se rebaterá qualquer um destes pontos, como manifesta Frank Austermühl:

In view of what we know about the constantly growing volume of texts to be translated, the first point made by the ALPAC Report particularly seems quite short-sighted. Although widely condemned as being narrow and biased, the ALPAC Report had considerable influence on MT research in the 1960's. It led to the virtual end of US government funding and most MT projects were stopped.
(2001:156)

Nos EUA, apenas alguns cientistas e esforços isolados persistem nos seus estudos, como é o caso de Peter Toma, responsável pelo desenvolvimento do Systran (<http://www.systransoft.com>).

Os anos 80 trarão novo fôlego à investigação em tradução automática. O interesse recrudesciente por parte de diversas instituições, nomeadamente da CEE, que em 1976 comprará o Systran, o aumento exponencial da

informatização e o desenvolvimento da linguística formal esboçariam um quadro de circunstâncias ideais para que a inteligência artificial e a TA recebessem novo alento e apoios financeiros. É de destacar o papel preponderante desempenhado pela UE, que desde sempre se notabilizou e distinguiu como bastião contra a uniformização linguística, garantindo a cada um dos seus actuais 15 países membros o direito de usar a sua língua oficial nas instituições europeias. O Systran é hoje em dia, aliás, usado extensivamente pela Comissão Europeia que, já no início dos anos 80, decidiu fundar um projecto ambicioso com vista a desenvolver um sistema multilíngue para as línguas dos países membros – o EUROTRA (<http://www.ccl.kuleuven.ac.be/about/EUROTRA.html>). Equipas de cada país membro, num total de cerca de cem linguistas, desenvolvem métodos e paradigmas de análise da sua língua. Paralelamente, os investimentos da indústria privada abririam novos horizontes ao desenvolvimento de projectos de TA e de Tradução Assistida por Computador (TAC).

Com efeito, a partir da década de 80, os estudos e investigações em TA passam a orientar-se segundo objectivos mais realistas e modestos, instigados pelos fracassos sucessivos de um projecto demasiadamente ambicioso. A focalização passou então a ser, não a de produzir um sistema capaz de gerar uma tradução correcta sem intervenção humana, mas, ao invés, a de desenvolver aplicações informáticas que auxiliassem a tradução e programas de tradução automática que permitissem a intervenção humana. A principal meta da pesquisa em TA passou a ser a oferta de instrumentos eficazes, que num espaço de tempo razoável possam ajudar o tradutor humano, tornando o seu desempenho mais rápido, menos repetitivo e menos monótono.

DA TRADUÇÃO HUMANA À TRADUÇÃO AUTOMÁTICA: TIPOLOGIAS

Uma primeira abordagem da TA deverá, por conseguinte, incluir uma tipologização das várias ferramentas electrónicas disponíveis, de forma a que se distinga o largo espectro de aplicações existentes, funcionando esta categorização como ponto de partida para uma reflexão epistemológica sobre os seus usos, possibilidades e avaliação. O modelo mais divulgado é provavelmente o que diz respeito ao grau de automatização do processo de tradução, fazendo apelo aos diversos acrónimos em inglês, usados para descrever este processo:

- HT (Human Translation): Tradução Humana

- CAT (Computer-Assisted Translation): Tradução Assistida por Computador
- HMT (Human-Aided Machine Translation): Tradução Automática Com Pós-Edição
- MAHT (Machine-Aided Human Translation): Tradução Humana Assistida por Computador
- MT (Machine Translation): Tradução Automática
- FAHQMT (Fully Automatic High Quality Machine Translation): Tradução Automática de Alta Qualidade

Os termos MAHT e HMT estão normalmente agrupados no acrónimo CAT (Tradução Assistida por Computador). Esta listagem apresenta-nos um grau crescente de automatização do processo tradutivo, que vai da Tradução humana à Tradução inteiramente Automática de Alta qualidade, passando pelos sistemas de tradução Assistida por Computador.

Embora a FAHQMT – o mais extremista dos conceitos – esteja longe dos objectivos sonhados nos anos 40 e 50, ela tem numerosas provas dadas em três sectores de vital importância para a economia. Referimo-nos aos sectores aeronáutico, farmacêutico e meteorológico, que conseguiram alcançar um dos objectivos fundamentais perseguidos pela TA, o de reduzir os custos e aumentar a rapidez de execução, sem descurar a qualidade. Como foi isso possível? Mediante a limitação da terminologia empregue, alicerçada numa extrema simplicidade sintáctica. No caso da indústria aeronáutica, por exemplo, os limites terminológicos são de cerca de três mil palavras².

É óbvio, tal como defende Melby, que no momento actual das investigações, os sistemas de TA bem sucedidos só podem ser compreendidos num contexto de especialização. “Current techniques in machine translation produce fully-automatic high quality translation only when applied to a body of similar texts which are all restricted to the same domain. The texts must be static in that they do not contain new metaphors, allusions or grammatical constructions.” (Melby, Abril, 1999). O facto é que a maioria dos sistemas de TA depende do auxílio humano, prevendo quer a pós quer a pré-edição, o recurso a dicionários electrónicos ou a memórias de tradução, ou seja, cada vez mais se deve entender a TA como uma Tradução Assistida.

Na última década, foram gastas inúmeras verbas no desenvolvimento e marketing de produtos capazes de cobrir de forma completa o processo tradutivo. Os produtos, idealizados para automatizar a tradução de forma

(quase) completa são, a saber: as memórias de tradução; as ferramentas de software de localização; e os sistemas de tradução automática:

- As memórias de Tradução: tais como o IBM Translation Manager (<http://www-4.ibm.com/software/ad/translat/>), o *Déjà Vu* (<http://www.atril.com>) e o TRADOS – *Translators Workbench* (<http://www.trados.com>), são arquivos de texto multilíngues contendo textos segmentados, alinhados e classificados, que permitem o armazenamento e/ou busca de segmentos textuais, através do alinhamento de textos de partida com textos de chegada.
- O Software de Localização: como o Corel Catalyst, que permite adaptar um produto à situação específica do seu mercado alvo, retratando a máxima “think globally, act locally”. Os sistemas de localização traduzem os textos adaptando-os às normas culturais do mercado local. Numa só aplicação, combinam-se várias funções relevantes numa tradução: extracção terminológica, busca automática de glossários, memórias de tradução e editores WYSIWYG (*What you see is what you get*).
- A Tradução Automática: uma máquina que converta a língua de partida numa versão perfeita da língua de chegada, sem qualquer tipo de interferência humana. Como já vimos, após uma fase de grande investimento em pesquisa em tradução automática, os investigadores voltaram-se para projectos menos ambiciosos, usando textos de estrutura sintáctica simples e campos enciclopédicos ou terminológicos extremamente reduzidos; são exemplos o Systran, ou o TAUM-Meteo.

Se até há bem pouco tempo, estas diferentes aplicações pareciam irreconciliáveis e as firmas que os produziam e comercializavam mantinham uma postura perante o mercado de forte concorrência, hoje em dia, e nomeadamente em conferências internacionais, as investigações parecem concorrer no sentido de uma concertação de esforços, com o objectivo final de produzir aplicações informáticas, que com maior rapidez e fiabilidade possam traduzir uma língua de partida numa língua de chegada. Com efeito, quer os investigadores em memórias de tradução quer os que desenvolvem aplicações de TA, que até hoje viviam de costas voltadas, podem, num futuro próximo, unir experiências para otimizar as ferramentas já disponíveis, nomeadamente cruzando as regras de simplicidade sintáctica e terminológica de uns, com as milhões de memórias de tradução, de outros. Estaremos, então, a construir uma nova Torre de Babel?

¹ Tal é o caso de autores como Melby ou Vitek que, embora reconhecendo o papel que a Tradução Automática desempenha, negam a possibilidade de, no futuro, ela conseguir produzir resultados de qualidade semelhante à da tradução humana, quando aplicada a textos de linguagem geral.

² É exemplar o caso do sistema TAUM-METEO, que traduz, directa e automaticamente, boletins meteorológicos do inglês para o francês, sem qualquer tipo de revisão ou pós-edição. O seu sucesso deve-se, em grande parte, à limitação do texto de partida numa sublíngua, em que a estrutura sintáctica e a terminologia empregues são restritas, repetitivas e simples.

BIBLIOGRAFIA

ALFARO, Carolina (1998) *Descobrimo, Compreendendo e Analisando a Tradução Automática*. Monografia de Fim de Curso de Especialização em Tradução Inglês/Português, PUC-Rio de Janeiro.

ARNOLD, D. *et al.* (1994) *Machine Translation: an Introduction Guide*, London, NCC Blackwell

AUSTERMÜHL, Frank (2001) *Electronic Tools for Translators*, Massachusetts, Saint Jerome.

BAR-HILLEL (1964) *Language and Information*. Selected essays on their theory and application, Massachusetts, Addison-Wesley Publishing Company.

GERBER, Laurie *Working Toward Success in Machine Translation*. Disponível em: <http://www.elsnet.org/mt2010/gerber.pdf>.

GUESSOUM, A., ZANTOUT, R. *Semi-Automatic Evaluation of the Grammatical Coverage of Machine Translation Systems*, Disponível em: <http://www.eamt.org/summitVIII/papers/guessoum.pdf>.

MELBY, Alan K. (1995) *The Possibility of Language*, Amsterdam & Philadelphia, Benjamins.

(1999) «Machine Translation and Philosophy of Language», *Machine Translation Review*, n^o9, pp. 6-17, Disponível em: <http://www.bcs.org.uk/siggroup/sg37.htm>.

(1999) “Why Can’t a Computer Translate More Like a Person?” – 1995 Barker Lecture, Disponível em: <http://www.ttt.org/theory/mt4me/index.html>.

SPECIA, Lucia, RINO, Maria Helena Machado (2002) *Introdução aos Métodos e Paradigmas de Tradução Automática*, Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional, NILC – ICMC-USP, São Paulo.

VITEK, S.V. (2000) «Reflections of a Human Translator on Machine Translation or Will Translation MT Become the “Deus Ex Machina” Rendering Humans Obsolete

in an Age When “Deus Est Machina”?» *Translation Journal*, Volume 4, N°3, Disponível em: <http://www accurapid.com/journal/13mt.htm>.