

TM & MT NA TRADUÇÃO TÉCNICA GLOBALIZADA – TENDÊNCIAS E CONSEQUÊNCIAS

Markus J. Weininger
Universidade Federal de Santa Catarina
markus@cce.ufsc.br

Resumo: O presente artigo revisa algumas consequências da aplicação de sistemas de tradução automática (*Machine Translation* – MT) e memórias de tradução (*Translation Memories* – TM) na área de tradução técnica e as tendências de desenvolvimento neste campo. Há convergências entre as abordagens aparentemente opostas de MT e TM no sentido a) que ambas integram ferramentas complexas associadas com a outra, b) na sua aplicação combinada e c) na tendência do uso de linguagem controlada (*Controlled Natural Language* – CNL) em ambas, seja *a priori* e explicitamente, seja *a posteriori* e implicitamente. O trabalho discute as causas iniciais dos problemas no uso de TM constatados por outros autores e defende que há uma tendência à fusão entre original e texto traduzido na área de tradução técnica, seja pela introdução de sistemas sofisticados de CNL, seja pelo uso de memórias de redação (*Authoring Memories* – AM).

Palavras-chave: CAT, tradução automática, MT, memórias de tradução, TM, *Controlled Language*, CNL.

Abstract: This paper reviews some consequences of Machine Translation (MT) and Translation Memories (TM) in the field of technical translation and reports trends and developments in this area. There is convergence between the apparently opposite approaches of MT and TM as far as a) both integrate complex tools associated with their counterpart, b) they are used in combination and c) by the trend to use Controlled Natural Language (CNL) in both, be it *a priori* and explicitly or *a posteriori* and implicitly. The article also discusses root causes for problems in TM use detected by other authors and defends that there is a tendency of fusion between original and translated text in technical translation, be it via introduction of sophisticated CNL systems or by using Authoring Memories (AM).

Keywords: CAT, Machine Translation, MT, Translation Memories, TM, Controlled Language, CNL.

Desde os anos 50, com os primeiros experimentos da IBM com *machine translation*, a tradução por máquina ou tradução automática prenuncia o fim da profissão de tradutor. Mais tarde, já nos anos 80, uma empresa japonesa anunciou uma máquina de xerox de onde o original em inglês sairia numa cópia já traduzida para o japonês. Porém, claramente, o anúncio do fim do tradutor foi até hoje sempre precoce. Pelo contrário, a tradução está se desenvolvendo cada vez mais, não apenas como campo de trabalho dentro da economia globalizada, mas também como área de pesquisa dentro da academia.

MT – a ameaça

Claro que a engenhoca da IBM em 1954, na verdade, não era uma máquina de traduzir. Ela “conhecia” 250 palavras e “sabia” traduzir cerca de 50 frases simples do russo para o inglês. Assim, a apresentação de tais sistemas desde então, na maioria das vezes, documenta mais a virulência do sonho humano de superar definitivamente os resultados da Torre de Babel do que realmente um progresso tecnológico definitivo. Porém, mesmo nestes inícios limitados, a tradução automática chegou a ser usada durante a guerra fria para “traduzir” textos russos, pelo menos a ponto de um leitor humano poder decidir se no texto havia palavras chaves que sugerissem que a tradução por um tradutor humano seria interessante. Ou seja, já naquela época, o objetivo da utilização deste recurso era economizar tempo (= dinheiro) e mão de obra com qualificação mais específica (ou seja, mais uma vez, dinheiro).

Não é preciso entrar em detalhes sobre as inúmeras e imensas dificuldades que mesmo sentenças aparentemente simples podem apresentar para a MT. Já se pode ter uma impressão viva de qualquer serviço de tradução automática online, como o da Google.

Também as publicações acadêmicas nesta área são eloquentes¹. Mesmo assim, investiu-se muito dinheiro em pesquisa teórica para permitir que uma máquina possa realmente processar uma língua natural sintática e semanticamente com o fim de servir como análise para a tradução. Há teorias muito interessantes como a HPSG (*Head-driven Phrase Structure Grammar*) de Carl Pollard e Ivan Sag (1994) que foi utilizada, por exemplo, por pesquisadores do DFKI (*Deutsches Institut für Künstliche Intelligenz* – Instituto Alemão para Inteligência Artificial) em Saarbrücken, nos anos 90 do século passado².

TM – a resposta

Os sistemas de memórias de tradução (*translation memories* – TM) que abordam a mesma questão de maneira menos pretensiosa querendo facilitar e acelerar o serviço do tradutor humano foram pesquisados em universidades a partir do final dos anos 70. As primeiras versões comerciais foram lançadas 10 anos depois, ainda muito caras e com aplicabilidade reduzida. A partir do final dos anos 90 começaram a se tornar padrão em algumas áreas específicas de tradução técnica.

O princípio de TM é quase o oposto da MT, ou seja, ao invés de tentar “interpretar” e “traduzir” o texto original, o funcionamento de TM baseia-se na força bruta, de ter um banco de dados de traduções anteriores semelhantes suficientemente grande que, a partir da repetição e reutilização de elementos parcialmente ou totalmente idênticos, faz surgir uma economia de tempo ao traduzir. Assim, ainda é o tradutor humano, agora “turbinado” pelo banco de dados, que traduz e decide sobre a usabilidade dos segmentos anteriormente traduzidos por ele ou outros tradutores humanos e é só ele quem adapta o material do banco de dados ao novo contexto. Já pelo seu princípio de funcionamento, TM parecem amedrontar menos a categoria dos tradutores, ou até valorizá-la e libertá-la do

trabalho monótono e repetitivo. Porém, apesar de parecerem abordagens opostas, há uma aproximação entre MT e TM em alguns sentidos.

Convergências entre MT e TM

A primeira convergência evidente é o campo de aplicação. MT e TM são úteis para todas as situações onde há volumes de textos grandes com grau de standardização elevado, repetição considerável de elementos textuais e uma terminologia mais ou menos controlável. Maior vantagem se dá ainda onde estes textos precisam ser traduzidos para muitas línguas simultânea e rapidamente. Somente nestas situações o investimento em hardware/software, treinamento, configuração, customização e alimentação / revisão dos sistemas é rentável. Assim, temos como área de uso principalmente a tradução de documentação técnica em grandes empresas multinacionais ou para empresas menores que agora estão competindo no mercado global e precisam oferecer a documentação técnica atualizada em todas as línguas por motivos legais ou ainda em grandes organismos administrativos como a União Européia, onde existem condições semelhantes.

Uma outra convergência se dá pela combinação dos procedimentos dos dois métodos em grandes corporações. Assim, Boehme & Svetova (2001) apresentam um sistema típico de integração de vários processos para alcançar o objetivo principal, a redução de custos (suposto aumento de produtividade de 40%), através do uso de MT e TM, em combinação com garimpagem semi-automática de terminologia, incluindo o estágio de pós-edição para adaptação do banco de terminologia, dos dicionários e da memória de tradução depois de cada tradução. Blatt (1998:5) descreve um processo semelhante da sua área de atuação no serviço de tradução da União Européia.

Depois, recursos semelhantes são integrados tanto aos sistemas de MT quanto de TM, principalmente dicionários internos e bancos de terminologia. A alimentação de ambos pode ser efetuada manualmente ou de forma semi-automática. Rodolfo Raya da IBM discute as possibilidades de integração entre diferentes TM com TMX / TBX (TM Exchange / Term Base Exchange) no seu artigo Raya (2005). Chanod (1998) descreve exemplos do procedimento da Xerox em seus sistemas de MT neste sentido. Um exemplo da sofisticação de um sistema de TM é o conceito MetaMorpho que o grupo de seus autores húngaros chama de “modelo lingüisticamente enriquecido” que integra “rule-based machine translation” (Hodász, Gröbler & Kis, 2004: 83). O gráfico a seguir ilustra a integração de instrumentos avançados de processamento automático de língua natural com base em regras e descrições lingüísticas.

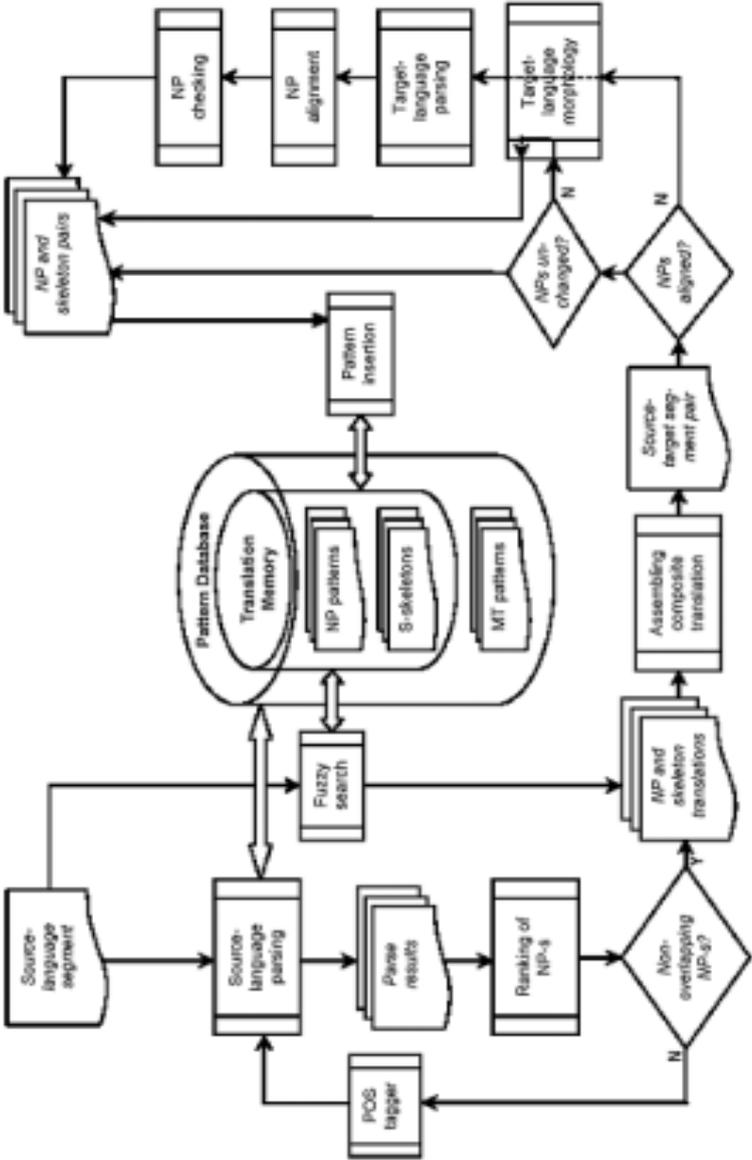


Gráfico 1: MetaMorpho, um sistema de TM com RBMT integrada (Hodász, Gröbler & Kis, 2004: 84)

A multinacional alemã de software SAP introduziu MT já em 1996 para o gerenciamento das traduções de sua documentação técnica. Como a necessidade de correção posterior dos resultados da MT é um fator considerável, é natural que as empresas começaram relativamente cedo a pensar em reduzir os problemas da MT já no início do processo, ou seja, no passo da criação do texto original. Já desde 1990 existe o “MultiLingual Technologies Group” na SAP para este fim³. O objetivo principal, da redução de custos, é perseguido através da redução dos problemas encontrados na MT já no texto original, mais exatamente, reduzir a ambigüidade semântica / complexidade sintática e controlar a consistência terminológica antes do início do processo de tradução.

Resultado: a linguagem controlada

Mais recentemente, esta abordagem ficou conhecida como *Controlled Language*. À primeira vista, é uma abordagem de um pragmatismo estonteante, no sentido de simplesmente adaptar a complexidade dos textos a serem traduzidos ao nível (limitado) da capacidade de processamento dos sistemas de MT, já que as tentativas inversas mostraram-se extremamente dispendiosas e, mesmo com os avanços da pesquisa em NLP (*Natural Language Processing*) e AI (*Artificial Intelligence*), resultados mais satisfatórios são esperados apenas a médio ou longo prazo. Assim, já que as máquinas (ainda) estão com dificuldade de adaptarem-se às línguas naturais humanas, os humanos precisam mudar a sua língua, ou, melhor, usar um subconjunto limitado e definido dela, para não exceder a capacidade de processamento dos sistemas de MT. Ou seja, para poder aplicar o corte de cabelo automatizado sem haver mutilações, rostos, orelhas, crânio, nariz e testa do freguês são sujeitos a uma cirurgia plástica profilática, no sentido de enquadrar a fisionomia e aparência nos padrões aceitáveis da máquina. Porém, segundo os defensores de *Controlled Language*, a

adaptação proativa à máquina não traz apenas perdas no texto. Supostamente, a legibilidade dos textos originais mesmo para falantes nativos aumentará. Claro que a linguagem controlada perde as nuances e a expressividade, igual aos rostos uniformizados acima mencionados, mas é inegável que ela pode garantir o conteúdo básico e reduzir problemas técnicos e econômicos da tradução assistida por computador.

John Hutchins (1998: s.p.) da universidade de East Anglia em Norwich (Reino Unido) cita um exemplo deste tipo de adaptação:

When fasteners are removed, always reinstall them at the same location from which they were removed.

Versão controlada:

Always reinstall fasteners in the same location.

O texto controlado é mais curto (16 itens lexicais antes contra 7 depois), o que por si já representa um fator de economia e duas subordinadas (problema potencial para sistemas de MT) foram eliminadas. A redundância diminuiu. A maior legibilidade depende, porém, não apenas destes fatores, pois a versão controlada pode ser menos inteligível em termos pragmáticos se no contexto não há menção da possibilidade da remoção dos fixadores.

Unscrew the plug-in unit from the wall and paint it.

Versão controlada:

Unscrew the plug-in unit from the wall and paint the plug-in unit.

Aqui a relação de economia já se inverte (10 contra 12 itens lexicais) e a eliminação da ambigüidade para um sistema de MT traz uma redundância desnecessária para um leitor humano. O'Brien (2003) discute os vários tipos e níveis de regras necessárias para controlar a linguagem.

CNL: Linguagem controlada *a priori* e sistematicamente

Desde então, a pesquisa na área de CNL (*Controlled Natural Language*) avançou consideravelmente. Rolf Schwitter da Macquarie University (Austrália), iniciador do projeto PENG (*Processable English*), define CNL como segue (Schwitter, 2005: s.p.):

Usually a CNL is defined as a well-defined subset of a natural language that has been restricted with respect to its grammar and its lexicon. Grammatical restrictions result in less complex and less ambiguous sentences, while lexical restrictions reduce the size of the lexicon and the meaning of the lexical entries for a particular domain (...)

Uma frase simples em PENG segue o seguinte esquema⁴:

Sentence -> Subject + Predicate
 Subject -> Determiner
 {+ Pre-nominal Modifier}
 + Nominal Head
 {+ Post-nominal Modifier}
 Subject -> Nominal Head
 Predicate -> { Negation }
 + Verbal Head
 + Complement
 {+ Adjunct}

A codificação do texto em PENG ocorre em RDF / OWL⁵, linguagens de marcação baseadas na linguagem XML⁶, já bastante usada na área de documentação técnica, para manter detalhes da formatação da documentação técnica⁷. Um breve exemplo deve ser o suficiente aqui para fins ilustrativos:

```
< rdf:Description rdf:ID= "nora_yuen" >  
< ex:supervise rdf:resource= "john_smith" >  
< /rdf:Description >
```

Esta descrição simples permite gerar (e traduzir automaticamente) frases como:

Nora Yuen supervises John Smith.

Com a adoção de um controle da linguagem usada na documentação técnica, no mínimo na fase de transição, o ônus é deslocado para a redação original dos documentos, como salientam Van der Eick & Van Wees (1998: 67)

Technical writers often find it hard to create new documents in a Controlled Language (or to rewrite existing documents), especially if a large number of previously acceptable sublanguage constructions can no longer be used. To prevent frustration, they should know how to paraphrase these constructions in the Controlled Language. Apart from training, it is important to provide authors with supporting software to support the authoring process.

A solução proposta por Schwitter (2004: 717) é o uso de um editor de texto que antecipa as possibilidades de continuação das sentenças e ao mesmo tempo oferece listas com opções e a correção de trechos não aceitáveis para as regras da CNL.

Sistemas de autoria para tradução com TM: linguagem controlada *a posteriori*

Mas a necessidade de unificação da linguagem surge não apenas para fins de MT. Mesmo no uso mais convencional de TM, a tendência globalizada de reduzir custos exerce uma pressão quase no mesmo sentido, como documenta Dawn Murphy (Murphy, 2000: 7):

Many companies nowadays expect their translation providers to use translation memory tools to cut translation costs. However, often these same clients are disappointed when the level of text re-use (100% matches) reported by the translation provider is much lower than they had expected. On analysis, it is often found that the lower hit rate is down to the way the source text is written. In effect, the source text is rewritten every time, so not surprisingly, it has to be retranslated every time.

Já Underwood & Jongejan (1999: 147) mencionam um software desenvolvido na universidade de Regensburg (Alemanha) para a avaliação da simplicidade / complexidade de um texto através de parâmetros facilmente avaliáveis de forma automática, tais como tamanho médio de palavra / frase, quantidade de vírgulas por frase, frequência de verbos modais e auxiliares, presença de terminologia pré-existente no dicionário do sistema de TM.

Ao mesmo tempo que se tenta reduzir a complexidade da linguagem, existe a opção para o autor do texto original de utilizar o banco de dados do próprio sistema de TM. No início, isso era visto como medida para aumentar a consistência terminológica através das versões da documentação geradas por grandes times de tradutores e muitas vezes terceirizados. Uma vantagem imensa para as empresas, logicamente, é o “fator colateral” do ganho de tempo, pois a plataforma XML permite não apenas que o tradutor de documentação técnica utilize segmentos já anteriormente traduzidos. Nada mais evidente do que o autor usar este recurso para compor

um novo texto usando os segmentos originais já existentes para aumentar os desejados *100% matches* na hora da tradução do texto.

Conseqüentemente, Allen (2000: s.p.) defende que é “most probable that an AM [= Authoring Memory; MJW] tool would improve technical writing environments just as much as Translation Memory (TM) has done for translation environments. The exact match and fuzzy match principles are usually language independent”. Ou seja, usam-se os segmentos originais de textos preparados para a tradução mediante o uso de um sistema de TM para compor a partir deles um novo original, com o benefício de já ter a tradução pronta de ao menos partes do texto. Isso é enormemente facilitado pelo uso da mesma linguagem de codificação de ambos os textos, original e traduzido (SGML, XML, etc.).

Zydroñ (2005: s.p.) declara:

hat are the benefits of using xml:tm for authors? Firstly, you have a detailed record of your authoring process. You also have all of your sentences segmented for you. You can then use this information to store all sentences in a phrase reuse database. This allows the creation of a phrase reuse system for authoring. If a sentence has been authored and then translated, encouraging reuse allows for a much higher percentage of leveraged memory in the future. Recent internal studies of automotive manuals have shown that up to 80% of a repair manual can be created from reused phrases.

Na SAP, o projeto de linguagem controlada recebeu o acrônimo eufemístico de SKATE (SAP Knowledge Authoring – Text Enhancement)⁸. Sem dúvida, o ganho em velocidade sugerido pela imagem dos patins é garantido. Este artigo, porém, quer focalizar também outro aspecto conhecido da patinação no gelo: o de derrapar ou cair de forma normalmente não muito elegante.

Quem controla a qualidade?

Um aspecto de deslizos e erros até maiores, quase que programados pelo uso de sistemas de TM (com ou sem combinação com MT e AM), está no fato do cliente muitas vezes não pagar pela tradução de segmentos 100% idênticos, supostamente já pagos em outra ocasião, ou até proíbe que estes segmentos sejam alterados, supostamente para manter o padrão de qualidade e terminologização já alcançado. Michael Benis (em Waller, 2003: s.p.) descreve este dilema:

When you are being asked not to touch 100% matches, you still have to read them and translators should be paid for the time involved. The problem is that the accountants have no idea how translation works. A text may have perfect terminological consistency, but it may not make any sense to the target audience. This may impact the perception of product quality because people cannot use it. That will have both an economic impact and a knock-on effect on brand image. The individuals responsible for such practices have not thought through the question of how communications work and the potential impact of poor documentation.

Ou seja, o tradutor não recebe nada para a revisão e adaptação do material encontrado apesar de ser muitas vezes bastante necessária, principalmente por dois fatores: a) o segmento idêntico no nível de frase provém de outro contexto, e b) há erros de todos os níveis (ortográficos, sintáticos, semânticos, terminológicos). Adriana Rieche (representada com um artigo sobre o tema neste volume) demonstrou estes problemas de maneira muito clara na sua pesquisa de mestrado e postula que: “o trabalho de revisão e manutenção nunca deve ser considerado definitivo ou encerrado. Uma fase de revisão elimina os erros encontrados naquele estado específico da memória, mas não impede que haja outros erros.” (Rieche: 2004: 166). Também Schütz (1998: 61) defende a revisão

integrada de tradução e original para a eliminação de erros de terminologia. Porém, na prática, isso nem sempre acontece. Por um lado, muitos clientes não aceitam uma cobrança extra, pois seria admitir que o trabalho anterior estava com falhas e, neste sentido foi cobrado indevidamente.

O tradutor, orientado para não alterar 100%-*matches* e não recebendo remuneração por este trabalho importante de revisão / eliminação de erros e inconsistências, ainda por cima com prazos exíguos e tarifas cada vez menores, simplesmente fecha os olhos e tenta empurrar o problema de volta para a origem. O problema é eliminado rápida e silenciosamente quando foi o mesmo tradutor que cometeu o erro no projeto anterior do seu cliente. Porém, nos contextos de grandes agências de tradução que recebem as memórias do cliente, ou seja, as mesmas foram elaboradas por colegas anônimos ou até pela concorrência, os tradutores não se sentem responsáveis pelo seu conteúdo. O cliente, por sua vez, se sente ludibriado, pois parece ser coagido a pagar duas vezes pelo mesmo serviço, porém, a agência não tem margem para incluir este serviço sem cobrar. Na prática, o empate apenas pode ser resolvido por uma atitude pragmática dos dois lados para viabilizar a manutenção criteriosa das memórias. Na sua conclusão, Rieche (2004:168) resume bem a situação:

(...) para que os sistemas de memória de tradução realmente sirvam ao propósito de ferramentas de auxílio à tradução, é preciso haver um controle sistemático das memórias, por meio de procedimentos regulares de revisão e manutenção. Caso contrário, as memórias podem conter erros variados e servir como fonte propagadora de erros.

Rieche analisa claramente o problema, mas não aborda com a mesma clareza as causas. Uma delas remonta a dimensões quase filosóficas na diferença entre as abordagens de MT e TM (não obstante as convergências técnicas aqui mencionadas). Enquanto

no caso de MT, obviamente, sempre existe uma necessidade de revisão, pois foi “uma máquina” (supostamente burra e que não tem culpa) que traduziu o texto, no caso da TM, isso não é aceito, pois foram tradutores humanos (supostamente inteligentes e, por isso, culpáveis). Apesar do ditado famoso que é humano errar, em matérias de CAT (*Computer Assisted Translation*) o erro é mais facilmente concedido à máquina, inclusive perdoado e corrigido investindo mais recursos. Claro que a causa raiz do problema está no plano econômico, na pressão de custos sobre ambas as partes que foi bem descrita por Nogueira (2003), exatamente dentro do contexto relevante para este artigo. Mas mesmo Nogueira não chega a nomear claramente o risco intrínseco do uso de sistemas de TM: as memórias de tradução permitem e facilitam um uso pouco responsável em três sentidos.

Primeiro: tradutores experientes sob pressão de tempo ou para manter sua renda em épocas de margens cada vez mais reduzidas não revisam os segmentos com 100%-matches, apesar da alerta expressada em Waller (2003: s.p.) “The main objective of TM should be to improve the quality of text. Speed should be a side benefit.”

Segundo: tradutores iniciantes ou não proficientes numa determinada área técnica conseguem traduzir textos que estariam fora da sua alçada se não usassem os recursos da TM. Porém, além dos erros adicionais que eles inevitavelmente acrescentarão à memória, eles não conseguem avaliar a qualidade das entradas existentes, pois são obrigados a confiar quase cegamente no material existente nas memórias, apesar do perigo que isso representa, claramente constatado por Rieche (2004) ou, por exemplo, pelos membros da mesa redonda já citada (Waller, 2003): “TM can lull you into a false sense of security”, ou “It is dangerous to trust 100% matches.” Muito menos ainda estes tradutores são aptos a executarem a revisão contínua e criteriosa no sentido de Rieche.

Terceiro: pior ainda são os resultados quando em empresas relativamente pequenas, que estão exportando equipamentos para

novos mercados, surge a necessidade de produzir rapidamente a devida documentação técnica traduzida. Pessoas sem nenhuma experiência em tradução usam as facilidades de sistemas de TM para traduzir textos com base em itens isolados. Ou então, para economizar na contagem de palavras a serem traduzidas, já processaram o texto original com um sistema de MT / TM que substituiu elementos isolados por material encontrado em glossários, dicionários online, textos anteriores etc., para reduzir o custo da tradução final que depois é encomendada a um tradutor profissional. Porém, este nem consegue mais entender o que era o sentido do texto original, agora total e grotescamente mutilado por tal procedimento.

No futuro: original e tradução se fundem

Como já se esboçou através das abordagens de CNL e AM, o círculo começa a se fechar e a distância entre original e tradução tende a ser mais reduzida. Idealmente, no conceito de *Semantic Web* como usado por Schwitters (2005), o conteúdo sempre é codificado de forma que possa ser processado por máquina, independente do idioma “original”. Com os respectivos módulos de idiomas, este conteúdo pode ser lido nos demais idiomas automática e imediatamente, sem necessidade de tradução explícita e superando a barreira lingüística, ao menos para os conteúdos produzidos dentro das premissas e com os instrumentos do sistema.

Porém, não apenas teoricamente e em pesquisas ocorre esta fusão entre texto original e texto traduzido na área de tradução técnica. Nas grandes corporações multinacionais que expandiram globalmente através de aquisições de concorrentes menores e fusões com competidores maiores existe a situação que a matriz transfere a tecnologia para as filiais que passam a responder posteriormente em exclusividade por uma parte da produção do grupo. Assim, cada filial nacional contribui com a sua produção específica e

atua como revendedora de todas as demais. Para as empresas isso representa ganhos pela eliminação de muitos empregos “duplicados”, apesar do custo envolvido para o transporte dos produtos e, não por último, para a tradução constante e contínua nos dois sentidos. Assim, por exemplo, uma filial brasileira de uma multinacional primeiramente absorve a tecnologia e a cultura da empresa através da tradução dos manuais técnicos existentes na matriz. Posteriormente, ela passa a assumir a construção e produção de uma determinada linha de produtos, cuja documentação técnica é redigida em português, para ser depois traduzida de volta primeiro para o inglês e a partir do inglês para os demais 27 idiomas dos mercados onde a multinacional atua. Porém, pelo uso de sistemas de AM, estes “originais”, na verdade, são compostos de segmentos anteriormente traduzidos para o português. Para que este sistema funcione, necessariamente é preciso que uma abordagem de CNL mais ou menos rigorosa seja adotada implícita ou explicitamente. Assim, mesmo se não usar as teorias e ferramentas sofisticadas propostas por Schwitter (2005), acaba criando uma *semantic web* dentro da respectiva multinacional, onde muitas vezes não é mais possível dizer o que era texto original e o que era texto traduzido.

Lista das abreviações usadas:

- AI – Artificial Intelligence
- AM – Authoring Memories
- CAT – Computer Assisted Translation
- CNL – Controlled Natural Language
- MT – Machine Translation
- NLP – Natural Language Processing
- OWL – Web Ontology Language
- PENG – Processable English
- TBX – Term Base Exchange
- TM – Translation Memories
- TMX – Translation Memory Exchange

RBMT – Rule-Based Machine Translation
RDF – Resource Description Framework
SGML – Standard Generalized Markup Language
XML – Extensible Markup Language (versão reduzida de SGML)

Notas

1. Como pequena introdução podem servir os anais dos congressos da EAMT (European Association for Machine Translation) dos últimos anos (disponível em <http://www.eamt.org/archive.html>).
2. Um dos resultados ainda pode ser experimentado online: <http://www.cl.uni-bremen.de/~stefan/Babel/Interaktiv/>
3. Veja Boehme & Svetova (2001).
4. Exemplo de Rolf Schwitter, online em: <http://www.ics.mq.edu.au/~rolfs/peng/nutshell.html>
5. OWL – Web Ontology Language, RDF – Resource Description Framework
6. XML – eXtensible Markup Language, derivada da SGML – Standard Generalized Markup Language.
7. É muito interessante que o Consórcio W3 que mantém os padrões de compatibilidade da internet que providenciou um salto no acesso a informações para a humanidade também é responsável pela definição de uma parte destas novas linguagens de codificação da informação.
8. Veja Brundage (2001).

Referências

- ALLEN, J. (2000) *Adapting the concept of "Translation Memory" to "Authoring memory" for a Controlled Language writing environment*. Proceedings of the 21st Conference of "Translating and the Computer". London: ASLIB.
Online: <http://www.transref.org/default.asp?docsrc=/u-articles/allen2.asp> (Último acesso: 30/11/05).
- BLATT, A. (1998) *Translation Technology at the European Commission: Description of a Workflow*. Online: http://europa.eu.int/comm/translation/reading/articles/pdf/1998_01_tt_blatt1.pdf (Último acesso: 30/11/05).
- BOEHME, U. & SVETOVA, S. (2001) *An Integrated Solution: Applying PROMT Machine Translation Technology, Terminology Mining, and the TRADOS TWB Translation Memory to SAP Content Translation*. In *Proceedings of the VIII Machine Translation Summit Santiago de Compostela, 2001*, European Association for Machine Translation (EAMT).
Online: <http://www.eamt.org/summitVIII/papers/boehme.pdf> (Último acesso: 30/11/05).
- BRUNDAGE, J. (2001) MACHINE TRANSLATION - EVOLUTION NOT REVOLUTION. In *Proceedings of the VIII Machine Translation Summit Santiago de Compostela, 2001*, European Association for Machine Translation (EAMT).
Online: <http://www.eamt.org/summitVIII/papers/brundage.pdf> (Último acesso: 30/11/05).
- CHANOD, J.-P. (1998) "Multilingual tools at the Xerox Research Centre". In *Proceedings of the 1998 EAMT WORKSHOP Translation technology: integration in the workflow environment WHO*. Geneva, p. 73-84.
Online: <http://www.eamt.org/archive/geneva.pdf> (Último acesso: 30/11/05).
- HUTCHINS, J. (1998) *Computer-based translation tools, terminology and documentation in the organizational workflow: report from recent EAMT workshops* [Paper presented at 4th Infoterm Symposium 'Terminology work and knowledge transfer', Vienna, August 1998].
Online: <http://ourworld.compuserve.com/homepages/WJHutchins/Vienna.htm> (Último acesso: 30/11/05).

MURPHY, D. (2000) "Keeping Translation Technology under Control". In *Machine Translation Review*, Issue No. 11, December 2000, p. 7-10.
Online: <http://www.bcs.org.uk/siggroup/nalatran/mtreview/mtr-11/mtr-11-7.htm> (Último acesso: 30/11/05).

NOGUEIRA, D. (2003) "Translation: A Market in Crisis?" In *Translation Journal*, Jan. 2003 Online: <http://accurapid.com/journal/23crisis.htm> (Último acesso: 30/11/05).

O'BRIEN, S. (2003) "Controlling Controlled English - An Analysis of Several Controlled Language Rule Sets". In *Programme of EAMT-CLAW 03 Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop*. Online: <http://www.eamt.org/archive/dublin/OBRIEN.PDF> (Último acesso: 30/11/05).

POLLARD, C. & SAG, I. (1994) *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics. Chicago: University of Chicago Press.

RAYA, R.M. (2005) *XML in localisation: Reuse translations with TM and TMX*. IBM DeveloperWorks. Online: <http://www.ibm.com/developerworks/xml/library/x-localis3/> (Último acesso: 30/11/05).

RIECHE, A. (2004) *Memória de tradução: auxílio ou impecilho?* Dissertação de mestrado, Rio de Janeiro: PUC.

SCHÜTZ, J. (1998) "MULTIDOC – Controlling Language in multilingual Documentation". In *Proceedings of the 1998 EAMT WORKSHOP Translation technology: integration in the workflow environment WHO*: Geneva, p. 55-64.
Online: <http://www.eamt.org/archive/geneva.pdf> (Último acesso: 30/11/05).

SCHWITTER, R. (2004) "Representing Knowledge in Controlled Natural Language: A Case Study", In M. G. Negoita, R. J. Howlett, L. C. Jain (eds.), *Knowledge-Based Intelligent Information and Engineering Systems*, 8th International Conference, KES2004, Wellington, New Zealand, September 2004, Proceedings, Part I, Springer LNAI 3213, pp. 711-717, 2004.
Online: <http://www.ics.mq.edu.au/~rolfs/papers/KES2004-Schwitter.pdf> (Último acesso: 30/11/05).

SCHWITTER, R. (2005) *Controlled Natural Language as Interface Language to the Semantic Web*, to be presented at the 2nd Indian International Conference on Artificial Intelligence (IICAI-05), Pune, India, December 20-22, 2005.

Online: <http://www.ics.mq.edu.au/~rolfs/papers/IICAI-schwitter-2005.pdf> (Último acesso: 30/11/05).

UNDERWOOD, N. & JONGEJAN, B. (1999) "Profiling Translation Projects: An Essential Part of Routing Translations". In *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 99)* Chester, England, August 1999, pp 139-149.

VAN DER EIJK, P. & VAN WEES, J.(1998) "Supporting controlled language authoring" In *Proceedings of the 1998 EAMT WORKSHOP Translation technology: integration in the workflow environment WHO*: Geneva, p. 65-70.

Online: <http://www.eamt.org/archive/geneva.pdf> (Último acesso: 30/11/05).

WALLER, J. (org.)(2003) *Translation Memory Debate*. Institute of Translation and Interpreting, ITI 2003 Conference 12 September 2003 . University of Warwick, UK.

Online: http://ecolore.leeds.ac.uk/downloads/2003.09_iti_tm_debate.pdf (Último acesso: 30/11/05).

ZYDRONĚ, A. (2005) "How to Leverage the Maximum Potential of XML for Localization". In *The Globalization Insider*, Vol. 5, 2005.

Online: http://www.lisa.org/globalizationinsider/2005/05/how_to_leverage.html (Último acesso: 30/11/05).