

O texto adaptado à máquina: estratégias de controle autoral para implementação da tradução automática

(The text adapted to the machine: authoring control strategies for machine translation implementation)

Érika Nogueira de Andrade Stupiello

Instituto de Biociências, Letras e Ciências Exatas – Universidade Estadual Paulista (Unesp)

erika@traducao-interpretacao.com.br

Abstract: Machine translation systems have been increasingly used for translation of large volumes of specialized texts. The efficiency of these systems depends directly on the implementation of strategies for controlling lexical use of source texts as a way to guarantee machine performance and, ultimately, human revision and post-edition work. This paper presents a brief history of application of machine translation, introduces the concept of lexicon and ambiguity and focuses on some of the lexical control strategies presently used, discussing their possible implications for the production and reading of specialized texts.

Keywords: machine translation, lexical control, pre-edition, post-edition.

Resumo: Sistemas de tradução automática têm sido cada vez mais utilizados para tradução de grandes volumes de textos especializados. A eficácia desses sistemas depende diretamente da implementação de estratégias de controle de uso do léxico dos textos de origem, como uma forma de garantir o desempenho da automação e, por fim, o trabalho humano de revisão e pós-edição. Este trabalho apresenta um breve histórico da aplicação de automação em tradução, introduz o conceito de léxico e ambiguidade e enfoca algumas das estratégias de controle lexical empregadas na atualidade, discutindo suas possíveis implicações para produção e a leitura de textos especializados.

Palavras-chave: tradução automática; controle lexical; pré-edição; pós-edição.

Introdução

Desde o início, o pensamento fundamentador de projetos de tradução automática esteve voltado ao desenvolvimento de sistemas que atuassem na tradução de textos de áreas tidas como técnicas e científicas e que, de acordo com o senso comum, apresentariam menos variações contextuais e outras limitações resultantes de diferenças culturais. As primeiras ideias de automatização da tradução remontam à invenção do computador e à crença, fomentada pelos pioneiros trabalhos em engenharia da linguagem, de que as línguas naturais poderiam ser por este processadas. O matemático norte-americano Warren Weaver compôs, na década de 50, um histórico memorando que inaugurou as pesquisas em tradução automática (doravante, TA) e em que desenvolvia associações para o problema da ambiguidade lexical, desde o início, considerada um dos maiores obstáculos à automatização das línguas naturais.

As pesquisas em TA tomaram impulso no início da década de 60 e, já em seu início, dividiram-se basicamente em dois pólos: um que adotava abordagens empíricas de “tentativa e erro” para aplicação computacional e que buscavam desenvolver sistemas para produção de traduções simples, porém úteis; e aquelas de cunho teórico, envolvendo pesquisas linguísticas, que mais tarde inaugurariam a área de estudos da Linguística Computacional.

Durante o período inicial de pesquisas teóricas e práticas em TA, as propostas de projetos para solucionar os problemas gerados com a informatização da língua voltaram-se a aspectos interlinguísticos da automação. Conforme explica Hutchins (2007), alguns dos modelos desenvolvidos baseavam-se em representações “neutras” da língua, com o estabelecimento de códigos independentes para descrever as línguas envolvidas na tradução em dois estágios: da língua fonte para uma interlíngua e dessa interlíngua para a língua meta. Também nessa época foram apresentadas propostas, que supostamente reduziriam problemas de homonímia e ambiguidade, na forma de “dicionários bilíngues simplificados”, em que, para cada palavra da língua de origem, somente um equivalente na língua meta era designado, buscando abranger a maior parte dos sentidos que, idealmente, “não demandariam a análise de contextos (em geral, de palavras imediatamente adjacentes), e permitiriam que a ordem das palavras do texto de origem fosse mantida ao máximo” (HUTCHINS, 2007, p. 3).¹

O otimismo inicial dos projetos da primeira fase de pesquisas em TA acabou sendo substituído pela desilusão, frente à limitação ou à inaplicabilidade dos primeiros resultados, e pela desaprovação, principalmente por parte das agências governamentais norte-americanas, dos altos investimentos até então realizados. No retrospecto de Martins e Nunes,

estabelecia-se, pouco ao pouco, o consenso de que os recursos disponíveis, fossem linguísticos (como dicionários e gramáticas), fossem computacionais (como memória e processadores), eram não apenas insuficientes, mas inadequados para prover o tipo de demanda criado pelo processamento automático de línguas. (2005, p. 4)

Mesmo nas diferentes estruturas de projeto apresentadas nos anos oitenta, em que predominou a orientação por sistemas de interlíngua, e noventa, em que se iniciaram as primeiras pesquisas baseadas em *corpus* e as investigações em inteligência artificial, constatou-se uma tendência, na maior parte dos casos, de os defensores de novas abordagens exagerarem as contribuições por elas proporcionadas, ou de divulgarem soluções definitivas para o processamento da língua, com base em demonstrações de projetos desenvolvidos em pequena escala.

Um relatório elaborado pela Academia de Ciências dos Estados Unidos – ALPAC, em 1966, tornou pública a constatação dos altos custos e das restrições dos sistemas totalmente automáticos. As iniciativas em TA perderam seus subsídios mediante denúncias de falta de conhecimento e de tecnologia para execução das propostas. A maior crítica emitida pelo referido relatório seria a de que as pesquisas até então desenvolvidas teriam ignorado a possibilidade de alcançar sistemas que não almejassem à perfeição, mas que pudessem alcançar bons resultados com participação humana.

O desenvolvimento de um sistema reunindo trabalho mecânico e humano também foi advogado por Kay (1980/1997),² em um artigo considerado seminal à

¹ Esta e as demais traduções de citações deste trabalho de obras em língua estrangeira, e sem tradução para o português do Brasil, foram feitas por mim.

² O referido artigo foi primeiramente publicado em forma de um relatório para a Empresa Xerox em 1980. Este trabalho utiliza a referência publicada no periódico *Machine Translation*, de 1997, conforme consta nas referências bibliográficas.

discussão sobre a aplicação da automação em tradução intitulado *The proper place of men and machines in language translation*. Nele, Kay condena o investimento em pesquisas que visassem a um “ideal distante” de implementação, ao mesmo tempo em que defende a aplicação de recursos informatizados com a finalidade exclusiva de ampliar a produtividade humana. Conforme argumenta, “a eficiência de um sistema de tradução, como qualquer outro, deve ser avaliado em todos os seus componentes: humanos e mecânicos” (KAY, 1997, p. 11).

Em consequência às críticas, emitidas por órgãos financiadores e renomados pesquisadores na área como Kay (1997), Hutchins (1997, 2007), Melby (1997) e Somers (2003), sobre a limitação de aplicação dos projetos de TA e a exclusão de envolvimento humano no processo, muitas das pesquisas, a partir dos anos oitenta, voltaram-se ao desenvolvimento de sistemas para fins específicos, que requeressem trabalho humano de pós-edição. A mudança de concepção buscava atender à crescente demanda por traduções de documentações produzidas pelas áreas da ciência e da tecnologia e, posteriormente, à necessidade de intercâmbio de informações e de comunicação rápida entre usuários da internet falantes de diferentes línguas.

Atualmente, a imensa quantidade de documentação gerada pela necessidade de divulgar e promover produtos e serviços no mercado global, que muito se expandiu com as possibilidades de comunicação e acesso a informações promovidas pela internet, depende de traduções que possam ser elaboradas em prazos cada vez mais exíguos, acompanhando o lançamento, por vezes mundial, desses bens. Para a tradução de documentação promocional e técnica, assim como de extensos manuais, muitas indústrias estão investindo no desenvolvimento e no aprimoramento de sistemas de TA para o tratamento de textos com conteúdo lexical previamente padronizado com a finalidade tanto de favorecer a aplicação da automação para tradução em diferentes línguas, como de promover a coesão e a padronização terminológica da produção final. A aplicação da automação nesses casos tem sido considerada bem-sucedida devido à introdução de estratégias de pré-condicionamento dos textos de origem a determinados padrões de uso do léxico e a construções gramaticais fixas que teriam por finalidade limitar a ambiguidade lexical e, desse modo, viabilizar a aplicação da automação em grandes volumes de texto.

Para entendermos algumas das implicações dessas estratégias para a prática tradutória e a composição textual, tanto do original como da tradução, apresentamos, no próximo item, a conceituação de léxico, abordando a inevitabilidade da ambiguidade tanto no léxico geral como no especializado.

A questão da ambiguidade lexical no léxico geral e no especializado

Sendo um sistema léxico “a somatória de toda a experiência acumulada de uma sociedade e do acervo de sua cultura através das idades” (BIDERMAN, 2001, p. 179), pode-se dizer que, a todo o momento, o léxico de uma língua passa por contínuas mudanças ao longo do tempo que engendram novas conotações aos significados que dele fazem parte.

As mudanças lexicais sempre foram sentidas nas diversas áreas do conhecimento, mais notadamente na contemporaneidade, em que o ritmo acelerado de descobertas e avanços nas áreas técnica e científica têm impulsionado a criação de palavras e termos para descrever novas realidades que, por não serem formalmente registradas em meios de ampla divulgação, muitas vezes passam despercebidas pelo

grande público não familiarizado com uma determinada área de estudo. Para Biderman (2006, p. 36), “à medida que fabrica novas realidades, o homem cria novas palavras em um processo incessante. E o léxico vai assumindo dimensões gigantescas sendo praticamente impossível registrá-lo e descrevê-lo por meio de um dicionário”. Biderman chama a atenção para a complexidade de selecionar e fixar o gigantesco número de unidades lexicais, criadas com uma frequência muito maior na atualidade, em obras impressas, em muitos casos consideradas “limitadas”, por não conseguirem acompanhar o acelerado ritmo contemporâneo das mudanças e criações léxicas.

No campo do léxico geral, mudanças seriam sentidas no uso do léxico por falantes de uma língua, como parte de situações cotidianas, que estimulariam o crescente emprego de palavras em contextos diferentes. Esse fato, somado à arbitrariedade linguística e à “pressão cultural”, responsável por aceitar ou repelir um novo uso de uma palavra, constituiria uma das causas da polissemia. Segundo Rehfeldt, “muitas palavras mudam de significado só pela razão de os falantes desconhecerem o sentido primitivo ou de confundi-lo com outro; ou, ainda, por quererem dar novo sentido à palavra, adaptando-a ao momento da comunicação” (1980, p. 78).

A flexibilidade do signo linguístico permite que a ele sejam acrescentados um número considerável de sentidos novos sem que essa ocorrência necessariamente resulte na perda do significado original. Segundo Barbosa (1996), “a polissemia é a regra e a monossemia, a exceção” no léxico de uma língua natural. O caráter polissêmico dos significados de uma língua não seria promotor de confusão, visto que o contexto em que um signo ocorre se incumba de a ele conceder um significado em uma situação específica.

Idealmente, um significado só teria um sentido na mente do interlocutor em um determinado contexto extralinguístico que, para Rehfeldt (1980), nunca pode deixar de ser considerado na comunicação, “pois nem sempre o contexto linguístico, a frase, é suficiente, por si só, para esclarecer o significado em jogo, como, por exemplo, no caso das ambiguidades” (p. 83). Rehfeldt distingue a ambiguidade sintática da lexical, esta de interesse direto deste trabalho. A ambiguidade sintática ocorreria quando a frase permite duas interpretações diferentes devido não à polissemia das palavras, mas às diversas estruturas sintáticas resultantes delas. Já a ambiguidade lexical aconteceria a partir de uma ou mais interpretações possibilitadas por uma mesma estrutura sintática, com polissemia nas palavras. Conforme exemplifica a autora, a frase “*A mãe tira o café de seu filho*” pode conduzir a duas interpretações, a saber: o fato de a mãe levar embora o café do filho e de lhe ter suprimido a bebida (p. 84).

A ambiguidade lexical não se reduz ao campo considerado como do léxico geral, mas também ocorre no uso técnico-científico, afinal, não há como impedir a influência de palavras de uso corrente em textos tidos como técnicos. Conforme aponta Zavaglia (2002), “uma palavra ambígua de uso comum, quando introduzida em um contexto técnico e científico, também ocasiona confusões e equívocos em seu uso” (p. 36). Embora certa precisão conceitual seja uma das condições necessárias ao intercâmbio comunicativo entre especialistas de diferentes áreas do conhecimento, não se pode garantir a univocidade do léxico descritivo de uma dada especialidade em todas as circunstâncias em que ele ocorre. Para a efetivação da comunicação, entretanto, buscase, em todas as especialidades, o estabelecimento de um determinado número de termos especializados, cujo papel principal seria reduzir a polissemia da língua empregada naquela área, evitando-se, assim, equívocos considerados previsíveis.

Para Krieger (2000), o recurso à terminologia constituiria uma forma de promover a univocidade na comunicação especializada, porém a terminologia técnico-científica não deixa de ser uma “questão de linguagem”, não constituindo exclusivamente “um constructo ideal e homogêneo a serviço de uma comunicação restrita ao âmbito de especialistas e isento de polissemia e de ambiguidades conceituais” (p. 211). Essa afirmação de Krieger baseia-se em sua defesa de que as terminologias são unidades lexicais e, como tais, “componentes naturais dos sistemas linguísticos” e na tendência, observada na contemporaneidade, de aproximação entre as terminologias e o léxico comum, em especial, pelo crescente “número de sentidos terminológicos nos verbetes da lexicografia da língua comum” (p. 211).

A constatação de Krieger nos lembra que o léxico científico usado na comunicação não escapa da intervenção de empréstimos, analogias e ocorrências de polissemia e, ainda que esforços sejam feitos para preservar a monossemia do léxico especializado, não se pode garantir que uma palavra ou termo técnico terão sempre o mesmo sentido em todos os contextos de que podem fazer parte. Ela remete também à tenuidade dos limites entre o léxico comum e aquele considerado como da especialidade, cuja cientificidade e suposta transparência constituiriam a mentalidade fundamentadora das propostas de controle do léxico especializado com fins de tratamento por sistemas de tradução automática. Essa questão será abordada no próximo item.

Estratégias de controle lexical em tradução automática para produção de traduções especializadas em grande escala

O desenvolvimento de sistemas de tradução automática para aplicação em textos especializados e elaborados segundo técnicas de uso controlado da língua teve início na década de noventa. Essas técnicas, denominadas na literatura de “autoria controlada”, envolvem restrições de utilização do léxico, com uso exclusivo de termos autorizados e de sentido supostamente único, que teriam por objetivo reduzir os problemas de ambiguidade durante a tradução. Essa estratégia é, segundo o enfoque deste trabalho, denominada de “controle lexical”, por não envolver somente a adoção de uma terminologia técnica e específica a uma determinada área, mas também a determinação de como o léxico da língua de origem, em geral, o inglês (língua de origem da maior parte dos manuais e textos técnicos traduzidos de forma automática), deve ser empregado em construções de frases de estrutura gramatical simplificada que, idealmente, produziriam uma única interpretação quando traduzidas para outras línguas de forma automática.

Uma das empresas pioneiras a implementar técnicas de controle lexical para aplicação de tradução automática em documentação técnica foi a norte-americana Xerox, que normalizou a elaboração de documentos de origem segundo previsto no manual de estilo da empresa denominado *Inglês Customizado Multilíngue*, em que estabeleceu os termos específicos a serem utilizados e as regras para construções frasais de seus textos. Entre o conjunto de regras a serem seguidas pelos autores técnicos, responsáveis pela preparação do material de origem a ser traduzido para diversas línguas, estão aquelas que delimitam o número de palavras de uma frase (25 palavras) e a determinação do campo conceitual dos termos para que sejam elaboradas frases simples, que “expressem uma só ideia” e, portanto, permitam somente uma tradução.

A vantagem das normas de controle do léxico original de manuais seria tanto a padronização de suas traduções para diversas línguas, como a possibilidade de produção

dessas traduções quase simultaneamente à disponibilização internacional dos respectivos bens manufaturados que acompanham essas traduções. A ampliação das relações comerciais entre os países, promovida pela globalização e pela facilidade de comunicação internacional atualmente disponível, demanda a produção concomitante de manuais em diversas línguas, como no caso da empresa Caterpillar, fabricante e exportadora de máquinas e equipamentos. Cada produto fabricado é constituído por uma série de partes (elétrico, hidráulico, motor) que requerem registro de suas especificações técnicas (montagem, desmontagem, manutenção e testes) para prestação de assistência nos países em que serão comercializados. Conforme relata Cronin (2003, p. 114), para dar conta de cerca de 800 páginas de manuais técnicos produzidos diariamente, essa empresa adota o *Inglês Técnico Caterpillar*, um sistema de uso controlado da língua inglesa, que emprega cerca de 70.000 palavras e frases fixas na elaboração de construções gramaticais padronizadas.

O *Inglês Técnico Caterpillar* originou-se a partir de normas anteriormente implementadas para redação em língua inglesa simplificada, uma iniciativa adotada pela empresa na década de oitenta que almejava dispensar a necessidade de tradução para outras línguas. Naquela época, acreditava-se que o inglês simplificado usado na redação dos manuais seria mais preciso e idealmente inteligível para distribuidores, mecânicos e operadores de máquinas em diferentes partes do mundo em que elas eram comercializadas. Essa abordagem na redação de documentação técnica foi logo abandonada por uma série de razões, entre elas, a insuficiência do vocabulário fixado para tratar das complexidades dos equipamentos, o conhecimento limitado da língua inglesa por representantes dos países importadores do maquinário Caterpillar e, especialmente, pelo fato de a empresa ter enfrentado resistência na aceitação de seus produtos em países que exigem que a documentação de bens importados esteja disponível na língua local (KAMPRATH et al., 1998).

Constata-se atualmente que um grande número de empresas e organizações está investindo no desenvolvimento de seus próprios sistemas de tradução automática, muitas vezes em departamentos criados na empresa para dedicação exclusiva à implementação desses sistemas e outras ferramentas de auxílio à tradução de documentações produzidas. O sistema Météo, desenvolvido pelo Serviço Ambiental do Canadá, para traduzir automaticamente previsões e boletins meteorológicos do inglês para o francês e vice-versa, constitui um exemplo de *software* projetado especialmente para lidar com a língua especializada da meteorologia. Empresas como a Caterpillar, por sua vez, investem em projetos em conjunto com departamentos de pesquisa em universidades, como a Carnegie-Mellon University, em Pittsburgh, Estado da Pensilvânia, cujo centro de pesquisas em tradução automática acumula anos de experiência na arquitetura e na produção desses sistemas para clientes específicos, muitos dos quais em línguas consideradas incomuns na comunicação internacional, como o sérvio-croata.

As estratégias adotadas para o controle lexical de textos que serão traduzidos automaticamente para diversas línguas prescrevem a padronização da terminologia especializada, do emprego do léxico geral e da construção gramatical de orações. De acordo com Hutchins (2005), com relação à terminologia, em geral, são aceitos no texto de origem somente os termos previamente “autorizados” que produzam maior consistência nas traduções. Quanto ao léxico, a atenção volta-se ao uso de palavras que não sejam ambíguas em um determinado contexto, como, por exemplo, a palavra em inglês *replace*, que pode significar “substituir”, no sentido de “remover um componente e colocá-lo de volta” ou “retirar um componente e colocar um novo” (p. 13). Palavras

consideradas ambíguas são substituídas por aquelas com sentido mais restrito, portanto de mais eficiente processamento automático.

Outro procedimento utilizado para controle do léxico está na elaboração de orações que não empreguem recursos como o de coordenação, mas optem pelo uso de orações simples (geralmente fixadas em, no máximo, 20 palavras), em que são sempre mantidos os sujeitos para auxiliar a identificação automática. O exemplo oferecido por Hutchins sobre a divisão dos períodos em frases curtas, contendo a descrição de somente um comando, pode ser constatado na redação controlada da seguinte orientação: *It is very important that you keep all of the engine parts clean and free of corrosion* [É muito importante manter todas as peças do motor limpas e sem corrosão], que, mediante aplicação de estratégias de controle lexical, passa a ser lida como *Keep all of the engine parts clean. Do not let corrosion occur* [Mantenha limpas todas as peças do motor. Não deixe ocorrer corrosão] (HUTCHINS, 2005, p. 13).

Algumas empresas, como a Tedopres International, com sede na Holanda, desenvolvem *softwares* específicos de controle lexical para elaboração de textos segundo critérios de *Autoria de Inglês Técnico Simplificado*. Entre os usuários desses tipos de serviços estão indústrias aeroespaciais, de máquinas e equipamentos, de tecnologias de processos, de telecomunicações e de equipamentos médicos. A adoção do inglês técnico simplificado inclui a padronização do vocabulário e do estilo dos textos, com a finalidade de aumentar a coerência, eliminar a ambiguidade e reduzir a complexidade da redação de materiais a serem traduzidos de modo automático. De acordo com as informações oferecidas na página eletrônica da empresa (disponível em www.tedopres.com), as técnicas desenvolvidas para implementação do *software* incluem um dicionário geral, um conjunto de regras de redação e a terminologia própria da empresa contratante. Algumas das regras de redação restringem o uso de gerúndios e verbos considerados complexos, além de controlarem o uso e a frequência com que os termos fornecidos são empregados. O exemplo a seguir, extraído da página eletrônica da Tedopres, apresenta o texto de origem de um manual de operação de maquinário, em que são apresentadas instruções para operação de uma empilhadeira. Temos o mesmo texto, antes e após a aplicação das técnicas de *Autoria de Inglês Técnico Simplificado*:

OPERATING TECHNIQUES

WARNING: Before operating the lift truck, FASTEN YOUR SEAT BELT. There are a number of operations, if not performed carefully, that can cause the lift truck to tip. If you have not read the WARNING page in the front of this Operating Manual, do so NOW. As you study the following information about how to properly operate a lift truck, remember the WARNINGS.

HOW TO OPERATE THE LIFT TRUCK

WARNING: Before you operate the lift truck, fasten your seat belt. If you are not careful, some operations can cause the lift truck to fall over. Read the warning page in the front of this manual now. While you read the procedures below, remember the warnings. (TEDOPRES INTERNATIONAL, 2008)

O texto controlado apresenta frases mais curtas do que a versão não submetida às técnicas de controle lexical. Os sujeitos do texto controlado encontram-se explícitos nas frases para identificação automática dos tempos dos verbos que os acompanham. A adoção das estratégias exemplificadas na elaboração de manuais e outros textos especializados auxiliam na redução do volume dos textos assim produzidos, o que viabiliza não só os custos do processamento automático e do trabalho de pós-edição, como a compilação e atualização dos bancos de dados terminológicos formados,

basicamente, a partir das traduções do material de origem para as diversas línguas em que será distribuído.

Bancos de dados terminológicos e fraseológicos eletrônicos estão, cada vez mais, sendo utilizados como materiais de consulta e referência para a tradução, ocupando o papel antes desempenhado pelos dicionários técnicos impressos que, como mencionado, apresentam limitações não somente para crescer, mas também para atualizar sua nomenclatura. A estruturação dos bancos de dados terminológicos é majoritariamente realizada de forma a facilitar a consulta e o reaproveitamento dos termos e frases nele armazenados e, para tanto, incluem-se informações mais detalhadas sobre as categorias gramaticais dos termos armazenados, suas variantes morfológicas, os contextos sintáticos em que se inserem e as circunstâncias em que cada opção de tradução é aplicável. Segundo Cabré (1999), os bancos de dados desempenham um papel importante como “pilares da normalização das línguas de especialidade, atuando como elementos de referência tanto para a padronização como para a normatização das línguas de especialidade” (p. 254).

Para aplicação da automação, a padronização da língua especializada de origem, realizada mediante a reunião da terminologia autorizada em bancos de dados eletrônicos e a aplicação de técnicas de controle lexical na composição textual, constitui um trabalho fundamental para restringir a ambiguidade lexical, um dos maiores limitantes ao desempenho de sistemas de tradução automática. De acordo com Mitamura (1999), pesquisadora de sistemas de processamento automático de línguas pela Carnegie-Mellon University, as estratégias de controle lexical desenvolvidas para aplicação em tradução automática em textos especializados enfocam a redução da ambiguidade nos textos da língua de origem, por meio da codificação de um “único” significado para cada par de palavra ou parte do discurso. No programa desenvolvido para controle de uso de língua denominado KANT, Mitamura enumera as três etapas seguidas para implementar o controle lexical de um texto de um dado domínio da especialidade: a limitação do significado por par de palavra/parte do discurso; a codificação de significados por meio de sinônimos e a codificação de termos ambíguos para desambiguação interativa, fase esta desenvolvida durante o trabalho de pós-edição.

A eficiência de sistemas de controle lexical como o KANT é majoritariamente atribuída à sua aplicação em domínios restritos, como aqueles de documentação técnica. Essa condição pressupõe a univocidade de sentido dos textos especializados ao basear-se na possibilidade de manipulação da semântica lexical dos materiais textuais desse domínio. Todavia, conforme argumenta Cabré (1999), não seria possível garantir que os recursos propostos para promover a desambiguação de sentidos terão validade necessariamente devido à natureza intrínseca pressuposta dos textos especializados, pois,

a pluralidade de situações na comunicação especializada abre a porta a uma diversificação de possibilidades discursivas dentro de uma mesma fronteira geral. Assim, os textos especializados produzidos em um processo de comunicação direta entre especialistas diferem daqueles produzidos entre um especialista e um estudante da especialidade, e mais ainda entre o especialista, direto ou mediado por um periodista, e o público em geral. Os textos são distintos e a terminologia também pode apresentar vários níveis de opacidade, imprecisão conceitual, especialização, variação formal ou abertura estrutural segundo o tema, a perspectiva pela qual se trata e o nível de especialidade do texto. (CABRÉ, 1999, p. 210)

As diferentes situações comunicativas de que um texto especializado pode fazer parte exercem influência nos trabalhos de desambiguação lexical e de revisão da tradução automática, realizados na fase de pós-edição. Entretanto, a atenção à definição e à implementação de controle do uso do léxico nos textos de origem para aumentar a eficiência da aplicação de automação da tradução acabaria dissimulando o trabalho do tradutor de reconstrução e adaptação do texto traduzido durante a etapa de pós-edição, gerando a impressão de esse trabalho ser passível de previsão e controle mecânico.

No tocante à produção de textos especializados para tradução em diversas línguas, as estratégias de controle lexical empregadas podem, ademais, estar influenciando a criatividade da língua, (afinal, mesmo em um domínio especializado, ela não é totalmente técnica) além de promover profundas mudanças nos textos traduzidos. Mogensen (2000) argumenta que um texto traduzido automaticamente é mais preso ao original e, com frequência, supersimplificado. Como exemplo, Mogensen cita o caso de manuais técnicos redigidos em dinamarquês os quais, devido ao amplo emprego de estratégias de controle lexical para aplicação de tradução automática, estariam substituindo o emprego de construções complexas impessoais comuns nessa língua por comandos curtos, na forma verbal imperativa e sem as conexões explicativas. Conforme defende, a partir de sua experiência como tradutora técnica,

aparentemente, a língua controlada melhora a legibilidade e a compreensibilidade ao reduzir a complexidade lexical e estrutural (como a ambiguidade) e ao especificar os padrões de estilo. Isso gera economia na manutenção, uma vez que um texto mais fácil de ler e entender é também mais fácil de atualizar. Tal fato torna o processo computacional mais simples devido à redução da complexidade léxica e estrutural (a ambiguidade e outros) e a prescrição de normas de estilo. Em resumo, torna a tradução automática praticável. (MOGENSEN, 2000, p. 30)

Cada vez mais, o planejamento de meios para compensar as limitações da tradução automática vem influenciando a maneira como são elaborados textos para divulgação e tradução. Os recursos aplicados para padronização lexical dos textos de origem (*softwares* e programas para uso simplificado da língua e regras de estilo para controle autoral) encobrem a complexidade da língua e podem estar extinguindo o estilo individual de textos de diferentes especialidades, especialmente ao implementar regras para prever e conter a expressividade do léxico em uso, que assume características próprias dependendo do público e do fim a que o texto se destina.

Considerações finais

A adoção de estratégias de controle lexical de textos de especialidade é, atualmente, uma exigência para o funcionamento eficaz de sistemas de tradução automática. O desenvolvimento de técnicas que envolvam a seleção do léxico e o pré-estabelecimento de seu emprego nos textos de origem demonstra ser tão importante quanto o contínuo aprimoramento dos sistemas de tradução automática.

O exemplo extraído de uma empresa especialista no desenvolvimento de programas de autoria controlada ilustra a posição submissa à máquina que o tradutor assume ao realizar modificações e ajustes na língua de origem e na tradução, nos

estágios de pré e pós-edição, respectivamente. Embora reconhecida a imprescindibilidade do tradutor na pós-edição do texto traduzido automaticamente, conclui-se que seu trabalho acaba sendo minimizado e restrito a duas frentes: à adequação de conteúdos já recuperados, e ao preenchimento das lacunas (trechos não traduzidos do original) deixadas pela máquina.

Por fim, o esforço em busca do engessamento do léxico visando à simplificação da produção textual e ao controle da língua para suprir as limitações da máquina pode atuar no sentido de inibir o processo de leitura e interpretação, restringindo excessivamente a noção de contexto construída pelo leitor final da tradução.

REFERÊNCIAS BIBLIOGRÁFICAS

BARBOSA, Maria Aparecida. *Léxico, produção e criatividade: processos do neologismo*. 3. ed. São Paulo: Plêiade, 1996.

BIDERMAN, Maria Tereza Camargo. *Teoria linguística: teoria lexical e linguística computacional*. São Paulo: Martins Fontes, 2001.

_____. O conhecimento, a terminologia e o dicionário. *Ciência e Cultura*, São Paulo, v. 58, n. 2, p. 35-37, abr./jun. 2006. Disponível em: <http://cienciaecultura.bvs.br/scielo.php?pid=S00097252006000200014&script=sci_arttext&tlng=pt>. Acesso em: 20 set. 2008.

CABRÉ, Maria Teresa. *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Linguística Aplicada, 1999.

CRONIN, Michael. *Translation and globalization*. London: Routledge, 2003.

HUTCHINS, John W. Translation technology and the translator. In: CONFERENCE OF THE INSTITUTE OF TRANSLATION AND INTERPRETING, 11th, 1997, *Proceedings...* London, May. 1997. Disponível em: <<http://ourworld.compuserve.com/homepages.WJHutchins>>. Acesso em 10 mai. 2006.

_____. Current commercial machine translation systems and computer-based translation tools: system types and their uses. *International Journal of Translation*, New Delhi, v. 17, n. 1-2, p. 5-38, Jan./Dec. 2005.

_____. *Machine Translation: a concise history*. 2007. Disponível em: <<http://www.hutchinsweb.me.uk/CUHK-2006.pdf>>. Acesso em 15 ago. 2007

KAY, Martin. The proper place of men and machines in language translation. *Machine Translation*, Hingham, Massachusetts, v. 12, n.1-2, p. 3-23, 1997.

KAMPRATH, Christine; ADOLPHSON, Eric; MITAMURA, Teruko; NYBERG, Eric. Controlled language for multilingual document production: experience with Caterpillar Technical English. In: INTERNATIONAL WORKSHOP ON CONTROLLED LANGUAGE APPLICATIONS, 2nd, 1998. *Proceedings...* Pittsburgh, PA, mai. 1998.

KRIEGER, Maria da Graça. Terminologia revisitada. *D.E.L.T.A.*, São Paulo, v. 16, n. 2, p. 209-228, 2000. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010244502000000200001&lng=en&nrm=iso&tlng=pt>. Acesso em 25 set. 2008.

MARTINS, Ronaldo Teixeira; NUNES, Maria das Graças Volpe. *Noções gerais de tradução automática*. São Carlos: NILC-ICMC-USP, 2005. 25 p. Relatório técnico.

MELBY, Alan. Some notes on 'The proper place of men and machines in language translation'. *Machine Translation*, Hingham, Massachusetts, n. 12, p. 29-34, 1997.

MITAMURA, Teruko. Controlled Language for Multilingual Machine Translation. *Proceedings of Machine Translation Summit VVI*, Singapore, set. 1999.

MOGENSEN, Else. Orwellian Linguistics: how using computer-aided translation tools impacts the target language. *Language International*, Filadélfia, p. 28-31, out. 2000.

REHFELDT, Gládis Knak. *Polissemia e campo semântico: estudo aplicado aos verbos de movimento*. Porto Alegre: EDURGS/FAPA/FAPCCA, 1980.

SOMERS, Harold (Ed.). *Computers and translation: a translator's guide*. Amsterdam/Philadelphia: John Benjamins Publishing Co., 2003.

TEDOPRES INTERNATIONAL. *Documentation, communication, software*. Disponível em <www.simplifiedenglish.net>. Acesso em 20 jan. 2008.

ZAVAGLIA, Claudia. *Análise da homonímia no português: tratamento semântico com vistas a tratamentos computacionais*. 2002. v. 1. 360 p. Tese (Doutorado em Linguística e Língua Portuguesa) — Faculdade de Ciências e Letras Campus de Araraquara, Universidade Estadual Paulista, Araraquara, 2002.