

TRADUÇÃO, COGNIÇÃO E TECNOLOGIA: INVESTIGANDO A INTERFACE ENTRE O DESEMPENHO DO TRADUTOR E A TRADUÇÃO ASSISTIDA POR COMPUTADOR

Fabio Alves
Universidade Federal de Minas Gerais
fabio.ufmg@gmail.com

Resumo: Baseando-se em estudos sobre segmentação cognitiva em tradução e no surgimento de novos modos de produção decorrentes da criação de tecnologias de tradução (Dragsted, 2004), este artigo apresenta a estrutura e o desenho de um projeto de pesquisa que tem por objetivo investigar o impacto das novas tecnologias de tradução na cognição humana. Através da revisão da literatura sobre segmentação cognitiva em tradução e tecnologias de tradução, o artigo discute possíveis implicações que o uso de sistemas de memória de tradução possa ter no desempenho de tradutores.

Palavras-chave: Segmentação cognitiva, tecnologias de tradução, sistemas de memória de tradução, unidade de tradução, competência em tradução.

Abstract: Building on studies of cognitive segmentation in translation and on the emergence of new modes of production with the advent of translation technologies (Dragsted, 2004), this article presents the structure and design of a research project which aims at investigating the impact of translation technology on human cognition. By reviewing the literature on cognitive segmentation in translation and on translation technology, it discusses the possible implications that the use of translation memory systems may have on the performance of translators.

Keywords: Cognitive segmentation, translation technologies, translation memory systems, translation unit, translation competence.

Introdução

Embora ainda seja considerado um campo de pesquisa incipiente, os estudos sobre o processo de tradução avançaram bastante nos últimos quinze anos. A partir de um quadro multifacetado em que havia uma notável carência de rigor metodológico e pouca familiaridade com a prática de experimentação (cf. Fraser, 1996, para uma revisão bibliográfica), os últimos anos foram testemunha de esforços feitos por pesquisadores na Alemanha, Brasil, Canadá, Dinamarca, Espanha e Finlândia, entre outros, para desenvolver desenhos experimentais sobre o processo de tradução, passíveis de serem replicados em diferentes pares lingüísticos e em distintas condições de produção (cf. Alves (ed.) 2003; Alves & Gonçalves 2003, Hansen (ed.) 1999, 2002; Jakobsen 2002, 2003; Lauffer 2004; Lorenzo 2004; Lörscher 2004; PACTE 2003, Tirkkonen-Condit & Jääskäleinen (ed.) 2000). Com base nos resultados dessas pesquisas, sabe-se hoje que o processo de tradução é uma atividade cognitiva complexa caracterizada por um ritmo cognitivo específico que requer a utilização de múltiplas habilidades. Trata-se de um processo que tem uma natureza não-linear, de caráter recursivo, e que pode ser mapeado através de experimentos desenhados para fins específicos.

Embora esses trabalhos tenham contribuído de forma significativa para o avanço das pesquisas sobre o processo de tradução e seus resultados sejam significativos, o advento de tecnologias de tradução acabou por introduzir modificações significativas na prática de tradução que repercutem na forma como o processo tradutório ocorre. Se, por um lado, conhece-se hoje muito mais sobre os processos cognitivos subjacentes ao ato tradutório, por outro lado, novas variáveis, surgidas a partir de mudanças concretas na prática da tradução, interferem e modificam de maneira significativa as condições de processamento mental dos tradutores. Atualmente, tradutores experientes, além de incorporarem ferramentas tecnológicas, tais como consultas em tempo real e

editoração eletrônica, à prática da tradução, passam, cada vez mais, a se valer dos sistemas de memória de tradução como forma de otimizar suas condições de trabalho. Os sistemas de memória de tradução configuram, portanto, uma variável adicional recente que aporta mudanças significativas na forma como seres humanos segmentam e processam o texto para fins de uma tradução. Investigar de forma empírico-experimental como o uso de sistemas de memória de tradução afeta o processo de tradução cresce, portanto, em importância para os pesquisadores interessados na vertente processual dos Estudos da Tradução.

Buscando uma melhor compreensão desta nova modalidade de fazer tradutório, este artigo se propõe a analisar pesquisas em dois campos disciplinares distintos mas complementares: o campo dos processos cognitivos e do processamento de informação em tradução e o campo das tecnologias de tradução. A partir da revisão da literatura, o artigo procura destacar a relevância de se estudar de forma mais detalhada a interface entre o homem e a máquina no campo da tradução e propõe a estrutura de um projeto de pesquisa para tal fim.

Quadro teórico

Faz parte do senso comum que a tradução de um texto é feita por partes, ou seja, por unidades de tradução. A idéia da palavra como constituinte desta unidade é tida por leigos praticamente como condição *sine qua non* do processo de tradução. Sabe-se, contudo, que a segmentação cognitiva em tradução é um processo complexo que não pode ser delimitado exclusivamente por características morfológicas, lexicais ou sintáticas. A segmentação em tradução depende de uma série de fatores cognitivos e contextuais que abarcam, entre outros, o papel da memória humana, dos processos inferenciais, da solução de problemas e da tomada de decisão (cf. Alves (ed.) 2003; Alves (org.) 2004 e Alves 2005).

Nesse sentido, no âmbito da investigação sobre as interfaces entre processos cognitivos e sistemas automáticos de memória de tradução, faz-se necessário explorar a segmentação humana e eletrônica do processo de tradução, discutir o papel da memória humana e computacional, e identificar as características básicas dos sistemas de memória de tradução em sua interface com o tradutor humano. Dada sua prática ainda incipiente, são poucos os estudos feitos até o momento sobre a influência desses sistemas de memória de tradução sobre o comportamento cognitivo de tradutores. Dragsted (2004), em tese de doutorado defendida na Escola de Administração de Copenhague (*Copenhagen Business School*), revela que existem diferenças significativas na forma como tradutores segmentam um texto a ser traduzido e como isto é feito quando esses mesmos sujeitos trabalham assistidos computacionalmente por sistemas de memória de tradução. As conclusões de Dragsted apontam para a necessidade de modificações na forma como os atuais sistemas de memória segmentam o texto a ser traduzido a fim de que tradutores possam, de fato, se beneficiar da interface com a máquina. Em outras palavras, somente com um melhor conhecimento dos processos de segmentação cognitiva observado entre tradutores será possível emulá-los em sistemas virtuais na forma de uma tradução assistida por computador.

As próximas seções deste artigo buscarão delimitar, em separado, características específicas a cada uma dessas áreas e enfatizar a necessidade de um melhor entendimento da interface entre elas.

Segmentação cognitiva

Definir a unidade de tradução (daqui em diante, UT) é um problema bastante complexo. Na literatura sobre o tema, nota-se uma linha divisória entre uma corrente de orientação cognitiva e outra de orientação lingüística. Em uma revisão recente sobre o conceito de UT, Dragsted (2004) sugere ser necessário levar em considera-

ção tanto a perspectiva lingüística quanto aquela cognitiva, buscando-se uma convergência entre elas. Para Dragsted, a UT é identificada como sendo um segmento lingüístico que pode ser processado pela memória de trabalho do tradutor. Trata-se de uma delimitação conceitual que pode ser identificada com base nas pausas observadas no decorrer do fluxo do processo de produção textual e que pode variar de acordo com a percepção do nível de dificuldade de tarefa por parte do tradutor.

Lingüisticamente, para Dragsted, a UT pode ser entendida como um item flexível abaixo do nível da sentença, como um item focado na sentença ou oração, ou como sendo o parágrafo ou o texto. As duas primeiras categorias não são excludentes já que UTs abaixo do nível da sentença podem, às vezes, se constituir no nível da oração. Por outro lado, cognitivamente, a UT é considerada como sendo a unidade efetivamente processada pelo tradutor em oposição à uma unidade ideal, qual seja, aquela de natureza lingüística. Concordo com Dragsted que estudos empírico-experimentais sobre o processo de tradução em geral devam privilegiar sobretudo uma orientação cognitiva em suas definições de UT. Proponho, portanto, que a UT deve ser identificada inicialmente com base nos processos cognitivos observáveis indiretamente em um conjunto de dados. Somente após esta identificação, a UT deve passar por uma classificação de natureza lingüística. Nesse sentido, uma definição de UT processualmente relevante deve se afastar de uma perspectiva normativa na qual a UT é descrita como o segmento mais apropriado para fins de estabelecimento de uma equivalência entre os textos de partida e de chegada. A UT deve, portanto, ser definida como o segmento efetivamente processado pelo tradutor e em consonância com a proposta de Alves (2000:128) citada abaixo:

A UNIDADE DE TRADUÇÃO é um segmento do texto de partida, independente de tamanho e forma específicos, para o qual, em um dado momento, se dirige o foco de atenção do tradutor. Trata-se de um segmento em constante transformação

que se modifica segundo as necessidades cognitivas e processuais do tradutor. A UNIDADE DE TRADUÇÃO pode ser considerada como a base cognitiva e o ponto de partida para todo o trabalho processual do tradutor. Suas características individuais de delimitação e sua extrema mutabilidade contribuem fundamentalmente para que os textos de chegada tenham formas individualizadas e diferenciadas. O foco de atenção e consciência é o fator direcionador e delimitador da UNIDADE DE TRADUÇÃO e é através dele que ela se torna momentaneamente perceptível.

A partir da definição de Alves, assume-se, de forma complementar, que os segmentos processados cognitivamente terão sempre uma natureza idiossincrática, mas podem ser investigados através da aferição de parâmetros de intersubjetividade entre os sujeitos constituintes de uma determinada amostra. Nesse sentido, a investigação de aspectos da segmentação cognitiva em tradução deve se pautar pelo estabelecimento de parâmetros comparativos entre os sujeitos tradutores que participem de um determinado experimento. Para tanto, sugiro que à definição proposta por Alves (2000), seja acrescentada a seguinte definição proposta por Dragsted (2004:78):

A UT pode ser definida como a compreensão simultânea/consecutiva na língua de partida e a produção na língua de chegada de um segmento textual cujo tamanho é limitado pela capacidade da memória de trabalho e cujas fronteiras podem ser identificadas através de pausas.

Conseqüentemente, seguindo a sugestão de Dragsted, a fim de tornar observável — ainda que de forma indireta — um fenômeno existente apenas na mente do tradutor, o conceito de UT e sua segmentação cognitiva precisam ser operacionalizados em termos de tempo, incluindo-se as pausas observadas ao longo do processo de tradução. Nesse sentido, os padrões de segmentação cognitiva

observados entre tradutores serão decorrentes do tempo real investido pelo tradutor na execução da tarefa de tradução e delimitados por intervalos de pausa durante o tempo de produção textual.

Segmentação automática

Em contraposição à segmentação cognitiva observada em tradutores, a segmentação computacional existente nos sistemas de memória de tradução tem uma natureza completamente distinta. Um sistema de memória de tradução (daqui em diante, SMT) é um banco de dados que armazena e recupera traduções prévias que compõem a base desse sistema. Esta recuperação é feita na forma de sentenças cuja fronteira preferencial é o ponto final. A partir desta marca ortográfica, um SMT destaca uma sentença determinada na tela do computador e disponibiliza possíveis correspondências para que uma delas venha a ser selecionada pelo tradutor. Os SMTs têm, portanto, um foco explícito na sentença enquanto instância de segmentação e delimitação da UT. Eventualmente, esses sistemas são capazes de executar outros tipos de processamento de natureza sintática, mas a segmentação é sempre feita através de marcadores ortográficos, sejam eles o ponto, o ponto e vírgula, a vírgula e outros sinais ortográficos previamente definidos. Contudo, não é de conhecimento público os princípios de programação computacional que norteiam esta segmentação.

Para fins de exemplificação, o manual do usuário do *Translator's Workbench*, um SMT de uso bastante disseminado, produzido pela empresa alemã Trados não informa qual é o método utilizado pelo programa para recuperar os segmentos relevantes da sua base de dados. Aparentemente, o objetivo de um SMT é alcançar o nível de recuperação mais alto e preciso possível. Para tanto, esses sistemas calculam a semelhança das ocorrências ortográficas identificadas entre o insumo fornecido pelo texto de partida e as correspondências existentes no banco de dados do sistema. O valor

de recuperação (*match value*) é, portanto, calculado com base no número de caracteres idênticos que ocorrem em uma ordem específica no insumo e na base. (cf. Reinke, 1999).

Constata-se, assim, uma diferença fundamental entre o processo de segmentação cognitiva natural e a segmentação automática. Sabe-se que a sentença não constitui a unidade básica do processo de segmentação cognitiva em tradução. Pelo contrário, os tradutores parecem alternar o foco da UT com predileção por segmentos micro-textuais abaixo do nível sentencial. Na busca de uma correlação entre parâmetros de segmentação cognitiva natural com os parâmetros da segmentação utilizada pelos SMT, faz-se necessário abordar adicionalmente as características da memória humana e daquela utilizada por sistemas automáticos de tradução com o intuito de verificar se o processo tradutório e o produto da tradução são afetados – positiva ou negativamente – quando se traduz utilizando SMTs.

Características da memória humana

Nas pesquisas sobre a memória humana parece ser consenso que a informação processada pelo cérebro humano é armazenada em diferentes tipos de memória, cada qual com diferentes capacidades e características de acesso, armazenamento e recuperação. Esses diferentes tipos de memória são classificados como sistemas sensoriais, memória de curto prazo, memória de trabalho, memória de longo prazo, etc.

Miller (1956/1967), em um artigo seminal intitulado *The magical number seven, plus or minus two: some limits on our capacity for processing information* sugeriu que a capacidade limite para processamento da memória de trabalho estaria restrita entre cinco e nove itens com uma frequência predominante em torno do número sete. Este parâmetro quantitativo vigorou até Baddeley (1986) propor que o foco principal do conceito de memória de trabalho

não deveria se restringir à capacidade limite do número de segmentos passíveis de armazenamento, mas deveria ser observado a partir da variável tempo, diretamente atrelada à capacidade de armazenamento e recuperação de informações. Obviamente, há restrições quantitativas no que diz respeito à quantidade de tarefas que um dado indivíduo pode executar simultaneamente, mas o tempo utilizado para a sua execução em correlação com o número de atividades desempenhadas simultaneamente parece ser um fator mais relevante na investigação dos processos de memória. A noção de memória de trabalho sugere, portanto, uma correlação entre a recuperação de uma determinada informação e o tempo gasto nesta execução. Mesmo que esse tempo de execução seja extremamente rápido, será considerado como um processo de elaboração cognitiva e não como um processo decorrente de automatismos prévios.

Conseqüentemente, para fins dos interesses de pesquisa aqui relatados, a tradução, vista como uma atividade de produção textual que envolve a recuperação de informações (lingüísticas, culturais, contextuais, etc.), deve ser considerada como um processo não automático no qual a memória de trabalho do tradutor está constantemente envolvida, ainda que em graus de magnitude diferenciados. Para que possa construir significado a partir de um determinado insumo lingüístico veiculado pelo texto de partida, o tradutor necessita retextualizá-lo na língua de chegada por intermédio de um processo que requer o monitoramento cognitivo consciente das suas atividades. Os trabalhos recentes de Rothe-Neves (2002), Gonçalves (2003) e Dragsted (2004) corroboram experimentalmente a noção de que a tradução é um processo não automático com alto grau de monitoramento cognitivo consciente. Esta visão está de acordo com Gile (1995) para processos de interpretação consecutiva e simultânea, atividades cognitivas de certa forma correlatas àquela da tradução escrita. No caso das pesquisas sobre o processo de tradução, há evidências de que, à medida que tradutores se tornam mais experientes, cresce sua capacidade de monitorar de for-

ma consciente o processamento cognitivo de unidades maiores (Jakobsen, 2003). Isto pode ser mais facilmente observado no caso de tradutores que se especializam em determinadas áreas do conhecimento. Rothe-Neves (2002) encontrou evidências de que a memória de trabalho não é necessariamente uma condição cognitiva para assegurar o bom desempenho do tradutor; antes, a experiência e especialização em tradução contribuem para o aperfeiçoamento desse tipo de memória. Conclui-se, portanto, que os padrões de segmentação do texto de partida de uma tradução em UTs maiores ou menores podem ser correlacionados a padrões de segmentação cognitiva envolvendo informações de natureza geral.

Ericsson & Kintsch (1995) apontam que o desempenho de atividades complexas exige maior memória de trabalho. A fim de explicar o funcionamento da memória de trabalho em condições de desempenho de alto nível, os autores propõem a expansão do conceito de memória de trabalho no conceito de memória de trabalho de longo prazo (daqui em diante, MT-LP). Este tipo de memória seria resultado de prática deliberada, um conceito proposto por Ericsson (2002) para explicar instâncias de desempenho de alto nível entre atletas, artistas e outros profissionais, incluindo intérpretes e tradutores, com alto rendimento em suas atividades. Nesses casos, a memória de trabalho utilizada por esses indivíduos adquire status de memória de longo prazo em termos de capacidade de armazenamento e recuperação de informações enquanto mantém características da memória de trabalho em termos do tempo necessário para sua implementação. Em outras palavras, as principais características da MT-LP são a armazenagem rápida e eficiente e a recuperação de informação seletiva relevante para a tarefa que está sendo executada. No caso da tradução, isto significa a capacidade especializada por parte do tradutor de recuperar de forma rápida e eficiente e a partir de determinadas pistas linguísticas e comunicativas segmentos armazenados na MT-LP que sejam correlatos à UT sendo processada em um determinado momento no decorrer do processo de tradução.

Os resultados dos experimentos de Dragsted (2004) fornecem evidências conclusivas em favor de uma MT-LP para tradutores experientes. Trata-se de resultado de certa forma corroborado pelos dados de Rothe-Neves (2002) e Gonçalves (2003). Estas conclusões são relevantes porque: (1) explicitam que a memória de trabalho tem uma capacidade limitada determinada pelo tamanho da UT processada de forma idiossincrática por tradutores; (2) revelam que a tradução, vista de uma perspectiva cognitiva, envolve o processamento consecutivo de uma determinada UT nas línguas de partida e de chegada e exige esforços de compreensão e produção gerenciados pela capacidade da memória de trabalho; (3) indicam que essa natureza dual leva tradutores a desenvolver um tipo específico de memória especializada, qual seja, a MT-LP; (4) demonstram que esse tipo diferenciado de memória pode ser mapeado através de pausas registradas no fluxo do processo de produção textual; (5) evidenciam que a segmentação cognitiva natural em tradução muda de acordo com o grau de especialização da memória do tradutor e gera UTs mais longas à medida que aumentam a eficiência e a rapidez da MT-LP.

Sistemas de memória de tradução

Uma vez constatadas as principais características da memória de trabalho e um possível desenvolvimento em direção a uma MT-LP, parece-me pertinente avaliar as características da memória existente nos SMT para, posteriormente, discutir implicações para os processos de segmentação cognitiva natural em tradução, tecer analogias e fazer comparações entre a segmentação cognitiva natural e aquela realizada em ambiente assistido por sistemas semi-automáticos.

A partir da premissa, atualmente não hegemônica, que existe uma linguagem universal baseada em princípios lógicos e símbolos icônicos, as primeiras tentativas para mecanizar a tradução de textos

foram feiras na França e na Rússia nas décadas de 1930 e 1940. Pouco mais tarde, Shannon & Weaver (1949) propuseram a utilização de computadores para executar traduções automáticas. Esta proposta resultou em uma série de pesquisas que buscavam impulsionar a tradução automática e implementá-la de forma dissociada do tradutor (cf. Hutchins & Somers, 1992). A partir de então, como propõe Dragsted (2004), a história da tradução automática pode ser dividida em três gerações.

Os trabalhos da primeira geração eram baseados em dicionários e envolviam o desenvolvimento de sistemas de substituição completa por meio dos quais a tradução era executada palavra por palavra. O processo de tradução estava sujeito a alguns arranjos de natureza morfológica e manipulações na ordem vocabular do texto de partida, mas consistia, quase exclusivamente, na busca de equivalentes lexicais por meio de entradas dicionarizadas (cf. Somers, 1998). O objetivo era alcançar uma tradução automática total de alta qualidade, conhecida pela sigla inglesa FAHQT (*fully automatic high quality translation*). Uma vez que a maioria dos pesquisadores envolvidos nessas pesquisas não eram lingüistas e/ou tradutores, não é de se admirar que os trabalhos prescindissem de uma análise da estrutura interna do texto de partida. A preocupação dos cientistas da computação, proponentes da maioria desses projetos da primeira geração de sistemas de tradução automática, era voltada diretamente para a construção de algoritmos de base lógica. Como consequência, os produtos das primeiras FAHQT apresentavam limitações que geravam inconsistências de base lexical, sintática e semântica. Aspectos discursivo-pragmáticos da tradução não chegavam sequer a ser considerados. Apesar de um otimismo inicial, os trabalhos de pesquisa desta primeira geração de sistemas de tradução automática ficaram severamente comprometidos com a divulgação do relatório Bar-Hillel em 1960. Esta publicação do pesquisador do MIT recomendava que a FAHQT tivesse objetivos menos ambiciosos e reconhecia suas limitações de natureza semântica — e implicitamente discursivas — (cf. Hutchins & Somers,

1992). Esta primeira fase em busca de FAHQT terminou com a publicação do relatório ALPAC em 1966. Este trabalho concluiu que a tradução automática era mais lenta, menos precisa e muito mais onerosa que a tradução executada por seres humanos. Foram drásticas as repercussões deste relatório. Em decorrência da conclusão que não havia um futuro imediato para a tradução automática, os investimentos de pesquisa na área de FAHQT foram congelados nos EUA.

A segunda geração de pesquisas sobre tradução automática foi desenvolvida sobretudo no Canadá e na Europa. A principal característica desta segunda fase é o reconhecimento por parte dos pesquisadores da necessidade de uma abordagem lingüística mais sofisticada. A idéia básica consistia na proposição de uma abordagem indireta através da qual o texto de partida era transformado em um texto de chegada através de uma representação intermediária (cf. Somers, 1998; Lehrberger & Bourbeau, 1988). Os trabalhos de pesquisa pautavam-se, em sua grande maioria, por um método de transferência por meio do qual estruturas sintáticas da língua de partida eram transformadas em estruturas sintáticas na língua de chegada. Apesar de se trabalhar com a idéia de uma representação intermediária, o foco continuava a recair no processamento sintático da linguagem. No decorrer da década de 1980, os pesquisadores no campo da tradução automática começaram a perceber as limitações desta segunda geração de FAHQT. Chegou-se ao consenso de que a tradução automática teria que abordar outras questões além daquelas meramente sintáticas e semânticas. Deveria ser levado em conta que o processo de tradução pressupõe a compreensão do texto a ser traduzido e sua produção no contexto diferenciado da língua de chegada, sugestões feitas pelo relatório Bar-Hillel em 1960.

A terceira geração de pesquisas sobre tradução automática surge nos anos 1990 como o desenvolvimento de uma abordagem empírica baseada em textos paralelos, na lingüística de corpus e no tratamento estatístico dos dados. Segundo Somers (1998), esta abor-

dagem empírica pode ser dividida em dois sub-grupos: um grupo baseado em exemplos e outro com base estatística. No primeiro caso, a tradução era feita através da comparação de um determinado insumo da língua de partida com uma base bilíngüe em um corpus com exemplos típicos de tradução; as correspondências mais freqüentes são utilizadas como modelo para a produção do texto traduzido. Acreditava-se que este método geraria traduções menos literais que aquelas baseadas na estrutura sintática dos textos a serem traduzidos. Críticos desta vertente argumentavam que os corpora bilíngües utilizados eram restritos e, conseqüentemente, não tinham validade em termos das possibilidades reais de se vir a traduzir um determinado texto. No segundo caso, os sistemas estatísticos trabalhavam com base nas probabilidades de ocorrência de segmentos com base num conjunto de milhões de palavras armazenadas em um corpus de textos paralelos. Nesse caso, a escolha de equivalentes lexicais e a estrutura sintática do texto traduzido eram feitas com base em combinações probabilísticas. Apesar de reconhecer o potencial de generalização desta segunda alternativa, seus críticos argumentavam que a ela faltava a naturalidade da tradução observada em sistemas mais fechados como aquele baseado em exemplos.

Em decorrência dessas discussões, chegou-se à conclusão que sistemas automáticos de tradução, pelo menos em sua presente forma, necessitavam ser monitorados por seres humanos. Surgem, assim, os primeiros textos paralelos em formato eletrônico preparados para serem trabalhados por tradutores, os primeiros SMT. Inicialmente, em vez de se trabalhar com a proposta de uma FAHQT, o enfoque passa a recair sobre uma tradução automática assistida por seres humanos, conhecida pela sigla inglesa HAMT (*human aided machine translation*), na qual o computador executa a tradução e é auxiliado por seres humanos em diversos estágios, tais como, pré-edição, intervenções durante a produção e/ou revisão. Contudo, com o desenvolvimento da proposta de uma interface entre a máquina e o homem, observou-se que a tradução automáti-

ca ainda apresentava problemas básicos que exigiam a intervenção constante do tradutor. Verificou-se que a atividade do ser humano tem proeminência em relação à atividade do computador. Em consequência, surgiu uma nova vertente denominada tradução humana assistida por computador, conhecida pela sigla inglesa MAHT (*machine aided human translation*). Nesta nova forma, cabe ao tradutor o papel central no processo de tradução e o computador aparece como ferramenta auxiliar para a recuperação de informações previamente armazenadas a serem disponibilizadas para um possível uso pelo tradutor. Este artigo faz uma opção preferencial por MAHT por acreditar que um corpus formado por traduções já existentes conterà mais soluções para problemas de tradução que outros recursos eletrônicos disponíveis (cf. Isabelle et al, 1993).

Um SMT que trabalha dentro das bases propostas para uma MAHT pode ser definido, segundo Freigang (1998), como um sistema que permite o processamento computacional de um texto de partida em formato eletrônico para fins de produção de uma tradução tendo o tradutor como gerenciador de todos os estágios do processo e responsável pela atividade intelectual de produção textual. Este sistema também empregará o conceito de UT analisado anteriormente. Contudo, um SMT não pode prescindir de uma UT definida *a priori*. Neste sentido, a UT em um SMT é conceitualmente distinta daquela definida a partir de um enfoque cognitivo.

Um SMT pode ser alimentado de duas maneiras. Na primeira delas o tradutor interage com o sistema e o insumo lingüístico vai sendo armazenado à medida que o tradutor avança no trabalho de tradução. Desta forma, assim que um determinado segmento é armazenado, este se torna disponível para fins de recuperação caso seja identificado como correspondente a um insumo lingüístico subsequente. A segunda forma de armazenamento de um SMT consiste em alimentá-lo com fontes textuais disponíveis sobre um assunto específico. Antes que tais recursos possam ser utilizados pelo tradutor, é necessário alinhá-los com os segmentos correspondentes de um dado texto de partida. É possível fazer um alinhamento

automático por meio de ferramentas fornecidas pelo *Translator's Workbench* (cf. Trados, 2003). Contudo, este alinhamento automático precisa ser verificado pelo usuário e importado para o SMT. Para fins de pesquisas experimentais, este artigo defende o uso do segundo procedimento. Esta decisão não descarta o uso da primeira modalidade quando da prática de experimentação, mas procura garantir que o usuário terá uma resposta do SMT quando estiver desempenhando a tarefa de tradução. Isto se faz necessário, pois no caso de não encontrar uma correspondência para determinado segmento do texto de partida, o SMT registrará a informação “*NO MATCH*” (sem correspondente) na janela utilizada pelo programa para fornecer sugestões ao tradutor.

Em vista das similaridades operacionais entre o *Translator's Workbench* da Trados e outros SMTs (Dejá Vue, ISDLX, WordFast, Transit, etc.), assume-se que os resultados experimentais de um SMT possam ser generalizados em relação a outros sistemas similares. Para que uma sentença (ou um outro segmento) possa ser recuperada por um SMT é necessário que ela tenha uma semelhança de 30% a 100% com o insumo disponibilizado como texto de partida. Cabe ao tradutor decidir o grau de similaridade com que deve operar. O *Translator's Workbench* trabalha com uma referência padrão de 70% de correspondência entre sentenças (ou segmentos) e ocorrências com similaridade inferior a 50% são consideradas irrelevantes (cf. Trados, 2003).

Em busca de hipóteses de trabalho

Considerando-se que um SMT opera a partir de um foco explícito na sentença, qual seja, sua UT padrão, tem-se como consequência que o tradutor assistido por um SMT recebe em uma janela na tela do computador uma sentença do texto de partida. Isto, segundo afirma Dragsted (2004), pode levar o tradutor a priorizar o processamento da UT no nível da sentença quando, se

pudesse agir espontaneamente, o tradutor poderia segmentar o mesmo trecho de forma distinta e, provavelmente, mais próxima aos padrões naturais de segmentação cognitiva. Este forte enfoque na sentença por parte dos SMT pode ter efeitos no processo e no produto da tradução.

À luz dessas reflexões e a fim de investigar se, de fato e de que maneira, os SMT afetam o processo de tradução, este artigo levanta as seguintes hipóteses de trabalho:

- 1) A integração de um SMT ao processo de tradução terá impacto no tempo de produção global bem como no tempo dedicado especificamente à edição e revisão da tradução;
- 2) O processo de segmentação padrão utilizado por um SMT (com foco na sentença) levará tradutores a mudar o padrão de segmentação cognitiva observado em circunstâncias naturais;
- 3) Dependendo do nível de experiência, os tradutores serão mais ou menos afetados pelo processo de segmentação imposto por um SMT.

Uma vez que sejam mensurados os padrões de segmentação cognitiva natural e aqueles norteados por um foco em sentenças, poder-se-á avaliar qual o impacto de um SMT no processo de segmentação cognitiva de tradutores e em que medida este impacto interfere nas decisões do tradutor. Diante dessas constatações, serão levantadas duas outras hipóteses:

- 4) A modificação no processo de segmentação padrão utilizado por um SMT levará tradutores a trabalharem mais próximo ao padrão de segmentação cognitiva observado em circunstâncias naturais;
- 5) A segmentação cognitiva de tradutores novatos será mais afetada que aquela de tradutores experientes devido à presença e/ou ausência de um SMT no processo de tradução.

Por uma metodologia exploratória

Nas ciências sociais existe uma tradição consolidada que defende o uso de métodos quantitativos e qualitativos como formas complementares de se levantar dados a partir de pontos de observação diferenciados. Chama-se esta opção metodológica de triangulação. Jakobsen (1999) explica que a metáfora da triangulação tem sua origem nas técnicas de navegação e na estratégia militar. Ambas utilizam pontos de referência múltiplos para localizar a posição exata de um determinado objeto no espaço. Por analogia, investigar um mesmo objeto por meio de dados coletados e interpretados através de métodos diferentes aumenta as chances de sucesso do pesquisador em sua tentativa de observação, compreensão e explicação de um determinado fenômeno.

O caso da tradução não é diferente. Jakobsen (1999) comenta que nos estudos da tradução, sobretudo dentro das universidades européias, existe muito pouco espaço para pesquisas experimentais de laboratório que priorizem uma abordagem quantitativa. Muitos pesquisadores têm receio de serem rotulados de behavioristas e, por isso, evitam a inclusão de metodologias quantitativas em seus trabalhos de pesquisa. Contudo, observa-se continuamente nas investigações sobre a natureza do processo de tradução uma necessidade concreta de se tabular e sistematizar dados através de análises estatísticas.

A triangulação apresenta-se, pois, como uma alternativa metodológica para pesquisas empírico-experimentais em tradução que almejam esclarecer com objetividade características processuais do processo de tradução sem, porém, desprezar sua natureza subjetiva. Procura-se, desta forma, identificar convergências e divergências nas investigações de natureza quantitativa e qualitativa e, através do cruzamento dos dados obtidos por intermédio de abordagens metodológicas múltiplas, chegar a resultados mais confiáveis, mais generalizáveis e, por conseguinte, mais capazes de elucidar questões cruciais para os estudos da tradução.

Em Alves (2003) apresento uma proposta de triangular dados processuais e criar um banco de dados que permita, a partir de estudos de caso, um maior poder de generalização para as pesquisas processuais em tradução. O CORPRAT – Corpus Processual para Análises Tradutórias – é um sub-corpus do CORDIAL – Corpus Discursivo para Análises Lingüísticas e Literárias – desenvolvido conjuntamente por pesquisadores do Núcleo de Estudos da Tradução (NET) da Universidade Federal de Minas Gerais. A metodologia utilizada no âmbito do CORPRAT pauta-se pelo paradigma da triangulação e busca maior rigor na sistematização de critérios para fins de desenho experimental, coleta e análise de dados. No cenário internacional, esta metodologia é atualmente desenvolvida e utilizada conjuntamente por pesquisadores da Universidade Autônoma de Barcelona, da *Copenhagen Business School* e da *York University* (cf. Alves (ed.), 2003 e Alves (org.), 2004).

Desenho experimental

Um dos maiores desafios em um desenho de natureza empírico-experimental é o controle de variáveis. Idealmente, devem ser controladas todas as variáveis de um determinado estudo, mantendo-se como variável independente o fator que se deseja investigar. Proponho, neste artigo, que o SMT deve configurar a variável independente para pesquisas que tenham por objetivo investigar as implicações desta tecnologia de tradução na segmentação cognitiva em ambiente natural em comparação com aquela realizada em ambiente assistido por computador. As demais variáveis, incluindo os sujeitos tradutores, os textos, a tarefa e as condições de produção, devem ser consideradas variáveis dependentes afetadas pela presença da variável independente, qual seja, a utilização de um SMT.

Os sujeitos de pesquisa devem ser tradutores de textos especializados haja vista que a utilidade dos SMT se dá majoritariamente em traduções de/para este tipo de texto. A tradução de

outros gêneros textuais, tais como a tradução literária, por sua especificidade, não consegue alimentar adequadamente um SMT e, portanto, não deve ser priorizada. A experiência prévia com um SMT deve configurar também pré-requisito para a seleção dos sujeitos, haja vista o fato que não haverá treinamento deliberado para esta modalidade cuja prática também deverá estar controlada. Os tradutores deverão traduzir textos em ambiente natural e em ambiente assistido por computador. Desta forma, poder-se-á medir e comparar entre sujeitos os padrões de segmentação cognitiva observados em cada um dos ambientes.

No que se refere aos textos para os experimentos, observa-se aqui um problema crucial com relação ao controle das variáveis do experimento. Se os tradutores traduzirem textos diferentes em cada um dos ambientes, natural e assistido por computador, não será possível mensurar qual é a influência dos textos no processo de segmentação. Em outras palavras, não será possível afirmar com segurança se diferenças de processamento ocorrerão devido aos ambientes diferenciados ou a diferenças textuais. Por outro lado, se forem mantidos os mesmos textos, não será possível afirmar com segurança se diferenças observáveis são resultado da segmentação em ambientes diferenciados ou devido a efeitos da memória (natural e virtual) no processo de tradução. Uma solução para este impasse pode ser encontrada na preparação dos textos para fins de experimentação. Proponho, portanto, que sejam utilizados pares de textos de forma a que venham a constituir uma variável dependente. Nesse sentido, os textos utilizados na primeira fase de um desenho experimental que avalia a segmentação em ambiente natural deverão ser correlatos aos outros textos utilizados na segunda fase que avalia a segmentação em ambiente assistido por um SMT.

A coleta de dados deverá ser conduzida de acordo com os procedimentos metodológicos especificados em Alves (2003) por meio da triangulação de dados processuais. Para registrar padrões de segmentação cognitiva em ambiente natural será utilizado o programa *Translog*, um programa de informática que grava os movi-

mentos registrados no teclado do computador e permite ao pesquisador avaliá-los em termos de tempo de duração e de pausas e, com base nessas análises, verificar como tradutores com diferentes níveis de competência tradutória segmentam textos semelhantes em ambiente natural. Para registro dos padrões de segmentação cognitiva em ambiente assistido por computador será utilizado o *Translator's Workbench*, um SMT comercial desenvolvido pela empresa Trados. Como explicitado anteriormente, o princípio básico deste programa, assim como de qualquer outro SMT, consiste no armazenamento em um banco de dados de pares de estruturas lingüísticas (preferencialmente sentenças) e na recuperação automática de ocorrências similares disponíveis neste banco de dados.

Ainda de acordo com a metodologia proposta em Alves (2003), logo após a conclusão da tradução nos dois ambientes (natural e assistido por computador), serão coletados relatos retrospectivos, gravados digitalmente, obtidos por intermédio da visualização do processo de tradução através da função replay do *Translog* com aceleração de 500% e/ou das gravações feitas com o *Translator's Workbench* da Trados. As consultas feitas em tempo real pelos tradutores serão monitoradas e gravadas através do programa *Camtasia* e aquelas oriundas de fontes impressas identificadas quando dos relatos retrospectivos. Finalmente, em consonância com a proposta feita em Jakobsen (2002) e Alves (2005), os textos de chegada serão utilizados como parâmetros de aferição das diferentes etapas do processo de tradução e diferenciados entre textos concluídos quando do término da fase de redação e textos modificados durante a fase de revisão. Nesse sentido, de acordo com Lorenzo (2004), a revisão é considerada como uma fase diferenciada do processo de tradução e analisada em separado.

Posteriormente, os dados coletados por intermédio do *Translog* e do *Translator's Workbench* serão tratados estatisticamente por meio de análises correlacionais para aferir parâmetros de significância e variância. As análises qualitativas serão complementadas por análises qualitativas com base nos relatos retrospectivos relacionados a cada uma das tarefas de tradução.

Considerações finais

Como mencionado no início deste artigo, dada sua prática ainda incipiente, são poucos os estudos feitos até o momento sobre a influência dos sistemas de memória de tradução sobre o comportamento cognitivo de tradutores. Estudos preliminares revelam que existem diferenças significativas entre a modalidade natural e aquela assistida por computador. Esses estudos parecem apontar para a necessidade de modificações na forma como os atuais sistemas de memória segmentam o texto a ser traduzido a fim de que tradutores possam, de fato, se beneficiar da interface com a máquina. Ganham, portanto, importância pesquisas que visam a um melhor conhecimento dos processos de segmentação cognitiva observado entre tradutores e na comparação com formas de tradução assistida por computador. São perspectivas que este artigo aponta como promissoras para fomentar o diálogo entre a teoria e a prática e, de forma complementar, qualificar programas de formação de tradutores no país.

Referências

ALVES, F. "Um modelo didático do processo tradutório: a integração de estratégias de tradução". In Alves, F., Magalhães, C., Pagano A. *Traduzir com autonomia: estratégias para o tradutor em formação*. São Paulo: Contexto. 2000.

ALVES, F. "Tradução, cognição e contextualização: triangulando a interface processo-produto no desempenho de tradutores novatos". *D.E.L.T.A.*, Campinas, v.19, p.71-108, 2003.

ALVES, F. (ed.). *Triangulating translation: perspectives in process oriented research*. (Benjamins Translation Library 45). Amsterdam: John Benjamins, 2003.

ALVES, F. (Org.) *Cadernos de Tradução X/2: O processo de tradução*. Florianópolis: NUT. 2004. Volume especial sobre o processo de tradução organizado por Fabio Alves.

ALVES, F. "Ritmo cognitivo, meta-reflexão e experiência: parâmetros de análise processual no desempenho de tradutores novatos e experientes". In Pagano, A., Magalhães, Alves, F. (Org.). *Competência em tradução: cognição e discurso*. Belo Horizonte: Editora UFMG. 2005.

ALVES, F. & GONÇALVES, J.L.V.R. A relevance theory approach to the investigation of inferential processes in translation. In Alves, F. (Ed.). *Triangulating translation: perspectives in process oriented research*. Amsterdam: John Benjamins, 2003.

BADDELEY, A. *Working memory*. Oxford: Clarendon Press. 1986.

DRAGSTED, B. *Segmentation in translation and in translation memory systems. An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process*. (Tese, Doutorado em tradução). Copenhagen Business School. Copenhagen. 2004.

ERICSSON, K.A & KINTSCH, W. "Long-term working memory". In *Psychological Review* 102/2. p.211-245. 1995.

ERICSSON, K.A. "Expertise in interpreting. An expert-performance perspective". In *Interpreting* 5/2. p.187-220. 2002.

FRASER, J. "The translator investigated: Learning from translation process analysis". In *The Translator* 2/1. p.65-79. 1996.

FREIGANG, K.-H. "Machine-aided translation". In Baker, M. (Ed.), *Routledge encyclopedia of translation studies*. Londres: Routledge. p.134-136. 1998.

GILE, D. *Basic concepts and models for interpreter and translator training*. Amsterdam: John Benjamins. 1995.

GONÇALVES, J.L.V.R. *O desenvolvimento da competência do tradutor: investigando o processo através de um estudo exploratório-experimental*. Faculdade de Letras da Universidade Federal de Minas Gerais. Belo Horizonte. 2003. (Tese, Doutorado em Estudos Lingüísticos). Inédito.

HANSEN, G. *Probing the process in translation: methods and results*. Copenhagen: Samfundslitteratur. 1999. (Copenhagen Studies in Language Series, 24).

HANSEN, G. (Ed.). *Empirical translation studies: process and product*. Copenhagen: Samfundslitteratur. 2002. (Copenhagen Studies in Language Series, 27).

HUTCHINGS, W.J. & SOMERS, H.L. *An introduction to machine translation*. London: Academic Press. 1992.

ISABELLE, P. et al. "Translation analysis and translation automation". In *Proceedings of the fifth international conference on theoretical and methodological issues in machine translation, Kioto*. 1993.

JAKOBSEN, A.L. "Logging target text production with *Translog*". In Hansen, G. (Ed.) *Probing the process in translation: methods and results*. Copenhagen: Samfundslitteratur. 1999. (Copenhagen Studies in Language Series, 24).

JAKOBSEN, A.L. "Orientation, segmentation, and revision in translation". In Hansen, G. (Ed.) *Empirical translation studies: process and product*. Copenhagen: Samfundslitteratur. 2002. (Copenhagen Studies in Language Series, 27).

JAKOBSEN, A.L. "Effects of think-aloud on translation speed, revision, and segmentation". In Alves, F. (Ed.) *Triangulating translation: perspectives in process oriented research*. Amsterdam: John Benjamins. 2003.

LAUFFER, S. "The translation process. An analysis of observational methodology". In *Cadernos de Tradução X/2: O processo de Tradução*, p.59-74, 2004.

LEHRBERGER, J. & BOURBEAU, L. *Machine translation, linguistic characteristics of MT systems and general methodology of evaluation*. Amsterdam: John Benjamins. 1988.

LORENZO, M.P. Competencia revisora y traducción inversa. In *Cadernos de Tradução X/2: O processo de Tradução*, p.133-166, 2004.

LÖRSCHER, W. A model for the analysis of translation processes within the framework of systemic linguistics. In *Cadernos de Tradução X/2: O processo de Tradução*, p.97-112, 2004.

MILLER, G.A. *The psychology of communication. Seven essays*. New York/London: Basic Books Inc. 1967.

PACTE. Building a translation competence model. In: Alves, F. (Ed.). *Triangulating translation: perspectives in process oriented research*. Amsterdam: John Benjamins. 2003.

REINKE, U. Evaluierung der linguistischen Leistungsfähigkeit von Translation Memory-Systemen. Ein Erfahrungsbericht. In: *LDV-Forum* 16, p.100-117. 1999.

ROTHER-NEVES, R. *Características cognitivas e desempenho em tradução: investigação em tempo real*. Faculdade de Letras da Universidade Federal de Minas Gerais. Belo Horizonte. 2002. (Tese, Doutorado em Estudos Lingüísticos). Inédito.

SHANNON, C. & WEAVER, W. *The mathematical theory of communication*. Urbana: University of Illinois Press, 1949.

SOMERS, H. Machine translation. In Baker, M. (Ed.) *Routledge encyclopedia of translation studies*. Londres: Routledge. p.136-139. 1998.

TIRKKONEN-CONDIT, S. & R. JÄÄSKELÄINEN, R. (eds.) *Tapping and mapping the process of translation: outlooks on empirical research*, Amsterdam: John Benjamins. 2000.

TRADOS. *Translator's Workbench user guide*. Dublin: TRADOS. 2003.