

Universidade de São Paulo - USP
Universidade Federal de São Carlos - UFSCar
Universidade Estadual Paulista - UNESP

Noções Gerais de Tradução Automática

Ronaldo T. Martins
Maria das Graças V. Nunes

NILC-TR-05-12

NOTAS DIDÁTICAS DO ICMC-USP (No.68)

Outubro 2005

Série de Relatórios do Núcleo Interinstitucional de Linguística
Computacional
NILC - ICMC-USP, Caixa Postal 668, 13560-970 São Carlos, SP, Brasil

Noções Gerais de Tradução Automática¹

Ronaldo Teixeira Martins^{1,3}, Maria das Graças Volpe Nunes^{1,2}

¹Núcleo Interinstitucional de Linguística Computacional (NILC) – Universidade de São Paulo (USP) - Caixa Postal 668 – 13560-970 – São Carlos – SP – Brazil

²Instituto de Ciências Matemáticas e da Computação (ICMC) – Universidade de São Paulo (USP) - Caixa Postal 668 – 13560-970 – São Carlos – SP – Brazil

³Universidade Mackenzie, São Paulo, SP

rtmartin@uol.com.br, gracac@icmc.usp.br

***Resumo.** O objetivo deste texto é introduzir o leitor no domínio da tradução automática, seus pressupostos, seus objetivos e suas técnicas. O texto recupera, em linhas gerais, a história das tentativas de automação do processo de tradução humana, e detém-se sobretudo na definição de alguns dos problemas centrais ao campo. Reportam-se particularmente as dificuldades que conduziram os pesquisadores e os desenvolvedores a abandonarem a perspectiva inicial de construção de sistemas robustos de tradução completamente automática com qualidade, e avaliam-se as estratégias alternativas, hoje predominantes, e bem mais modestas, relativas à pesquisa e ao desenvolvimento de sistemas auxiliados por humanos ou, principalmente, de sistemas para auxiliar os humanos a realizar, de forma mais rápida e menos dispendiosa, o processo de tradução.*

1. Introdução

O Livro VII da *República* de Platão traz das parábolas filosóficas talvez a mais conhecida: a alegoria da caverna. Platão reporta um diálogo (imaginário?) entre Sócrates e Gláucon, sobre uma caverna subterrânea onde alguns homens teriam sido aprisionados desde meninos, amarrados pelas pernas e pelo pescoço, de tal forma que tivessem de permanecer imóveis durante todo o tempo, e não conseguissem olhar para trás ou para os lados, mas apenas para frente. Atrás deles e num plano superior, arderia, a certa distância, fora da caverna, um fogo, que projetaria, na parede para onde os homens estavam condenados a olhar, as sombras dos fatos que se passariam às suas costas. Obrigados durante toda a vida a manter imóveis as cabeças, os homens amarrados não veriam outra coisa de si mesmos, de seus companheiros e de tudo que se passaria no mundo, senão as sombras projetadas pelo fogo sobre a parede da caverna. Os prisioneiros tomariam as sombras pela realidade. E seu drama não se reduziria à ignorância - de que, aliás, os prisioneiros não teriam consciência. Sócrates afirma que, se um deles fosse libertado, e conduzido para fora da caverna, continuaria, a princípio, a sustentar que as sombras que antes via seriam mais verdadeiras do que os objetos reais que, com muita dificuldade, em virtude do excesso de luz, conseguiria passar a perceber. Mas, acostumando-se, pouco a pouco, à luz do dia, o prisioneiro liberto reconheceria finalmente a verdade, e sentiria piedade de seus companheiros de cárcere. No entanto, se voltasse à caverna, veria com muito mais dificuldade do que os que ali

¹ Este texto foi produzido em 2002

ficaram, por ter sido subtraído da luz do sol. Além disso, ao tentar convencer os companheiros de que eles veriam apenas sombras, seria tomado por louco ou por imbecil, e seu testemunho de muito pouco ou nada valeria.

O uso que Platão faz da alegoria da caverna está diretamente relacionado ao estado de ignorância a que estariam acorrentados os que confiariam excessivamente no "visível" (em oposição ao "inteligível"), sem que percebessem que, como os prisioneiros da caverna, tomam sombras por realidade. A mesma parábola pode ser utilizada, com alguma utilidade, para que se possa entender o que significa "tradução" e, particularmente, o que significa "tradução automática".

A aplicação mais óbvia talvez consistisse em perceber que os monoglotas estão aprisionados à caverna-prisão de sua própria língua, e das imagens nela projetadas. O tradutor seria o prisioneiro liberto que, tendo contemplado o texto original à luz do dia, se veria na difícil tarefa de vertê-lo para o vocabulário das sombras para que os demais prisioneiros pudessem entendê-lo. No entanto, segundo Sócrates, isso seria inútil, seja porque ninguém confiaria no tradutor, seja porque sua linguagem não poderia ser entendida por nenhum outro que não tivesse estado, também, à luz do dia. Por isso, o uso que se faz aqui mais pertinente deve tomar como ponto de partida o lugar-comum e afirmar que traduzir é, sobretudo, trair. Toda tradução, na medida em que materializa (apenas) uma entre as várias leituras possíveis de um texto, na medida em que empreende forçosamente escolhas lexicais, sintáticas e semânticas, é apenas uma pálida sombra do texto original, projetada na parede da caverna a que estão aprisionados os que desconhecem a língua em que o texto foi produzido. E o tradutor - longe de ser o prisioneiro liberto - é a própria luz que, no mesmo movimento, fabrica a imagem e produz o equívoco.

A traição se torna mais nítida - é bem verdade - quando considerado o texto literário, principalmente em verso. Não são poucas as vezes em que a musicalidade de um poema ou as ambigüidades intencionais do texto poético são sacrificadas no processo de tradução, porque irrepetíveis em outra língua. Mas a falta de visibilidade da traição em textos técnicos e prosaicos, por mais exatos e unívocos que possam parecer, não deve trair os sentidos: a aparente identidade entre o traduzido e o original é também, neste caso, uma sombra, que não escapa aos que podem perceber o que efetivamente se passa fora da caverna. E a máquina, por sua talvez excessiva honestidade, há de prová-lo bem.

O objetivo deste capítulo será, em grande parte, o mesmo e inglório objetivo do prisioneiro liberto: tentar estabelecer uma clara diferença entre o visível e o inteligível, entre as sombras projetadas pela visão espontânea (e acrítica) do senso comum, e a luz do dia que banha todos aqueles que estão efetivamente mergulhados na tarefa de construir máquinas falantes. Por nossa prévia experiência com a tradução automática, e os problemas dela derivados, julgamo-nos autorizados a derramar um pouco de luz sobre os pressupostos, os métodos e os objetivos deste domínio, tão problemático quanto essencial para o processamento automático das línguas naturais.

Este capítulo está organizado como se segue: na segunda seção, será delimitado o campo da tradução automática, com a identificação dos pressupostos e objetivos em torno dos quais ela se tem estruturado nos últimos cinquenta anos; na terceira seção serão discutidas algumas das diferenças entre a tradução humana e a tradução

automática, necessárias para que o campo possa ser convenientemente delimitado tanto por profissionais oriundos da área de Letras (ciosos, quase sempre, das diferenças entre línguas naturais e linguagens artificiais), quanto por profissionais da área das Ciências da Computação (que muitas vezes tomam o conceito de "tradução" por algo mais simples do que em verdade é); a quarta seção discute algumas das estratégias de desenvolvimento de sistemas de tradução automática, e recupera as principais técnicas adotadas; a quinta seção apresenta algumas conclusões gerais; e a sexta e última parte traz algumas sugestões para leituras complementares.

2. Delimitação do campo

A tradução automática (TA), ou tradução mecânica, ou, simplesmente, MT (*Machine Translation*), tem seu registro de nascimento normalmente situado em 1949. Embora tenha havido iniciativas anteriores (nomeadamente o experimento desacreditado de Smirnov-Trojanskij, em 1933), o célebre *memorandum* do matemático Warren Weaver é considerado o marco inicial da investigação no domínio. Weaver havia sido desafiado para que fizesse uso das calculadoras numéricas, então recém-inventadas, para fins de tradução. Vivia-se o início da Guerra Fria, e a demanda, no Ocidente, por traduções rápidas, sigilosas e baratas de documentos militares russos era crescente, e a sua possibilidade de automação, estratégica. As técnicas de criptografia, de que Weaver era profundo conhecedor, haviam tido grande desenvolvimento durante a Segunda Guerra Mundial. A primeira iniciativa sistemática de desenvolvimento de uma ferramenta de tradução automática partiu, pois, da hipótese de que um texto em russo não seria senão um texto em inglês codificado com símbolos ainda desconhecidos. Weaver investigou particularmente as invariâncias estatísticas observáveis entre as línguas naturais (especialmente entre o inglês e o russo), considerando a língua tão-somente como um sistema formal de codificação de informações (não muito diferente, portanto, dos outros códigos com os quais já havia trabalhado).

A este impulso inicial aderiram muitos outros pesquisadores durante toda a década de 1950, nos Estados Unidos e na então URSS, que passaram a investigar outros aspectos centrais ou subsidiários do problema da tradução automática. Dedicavam-se especialmente à tradução do inglês para o russo, ou do russo para o inglês. Realizaram-se conferências locais e internacionais sobre o tema, estabeleceram-se linhas diferentes de pesquisa e investigação, inauguraram-se publicações dedicadas à matéria, e a euforia inicial em pouco tempo contaminou também algumas nações européias (França e Inglaterra, principalmente).

A disseminação do campo trouxe à luz um conjunto progressivamente maior de divergências teóricas e metodológicas entre pesquisadores. Desde cedo se tornou patente, por exemplo, que a concepção de tradução como decifração não poderia ser mantida. Embora pudessem ser efetivamente concebidas como sistemas formais de codificação de informações, as línguas naturais, principalmente em função da ambigüidade, que se revelava nos mais variados níveis de análise, não podiam ser comparadas a sistemas artificiais em que a relação entre símbolo e referente era estabelecida de forma biunívoca. Buscaram-se outras invariâncias interlingüísticas (semânticas, lógicas), propuseram-se outros modelos de tradução (baseados no mapeamento sintático, por exemplo), restringiu-se o processo de automação (com a previsão de pré-edição do texto a ser traduzido ou de pós-edição dos resultados), mas

não se pôde furtar ao óbvio: que a empresa era mais difícil do que a princípio tinha parecido, e que o grau de complexidade da linguagem humana talvez ultrapassasse o limite do razoável (ou do possível). A despeito dos esforços e dos investimentos, os resultados práticos eram pífios, e de muito pouco serviam. Estabelecia-se, pouco a pouco, o consenso de que os recursos disponíveis, fossem lingüísticos (como dicionários e gramáticas), fossem computacionais (como memória e processadores), eram não apenas insuficientes, mas inadequados para prover ao tipo de demanda criado pelo processamento automático das línguas naturais.

Em 1966, o Relatório ALPAC, da Academia de Ciências dos Estados Unidos, encarregada da análise dos resultados dos (muitos) programas subsidiados pelo governo norte-americano, confirmou o sentimento de frustração experimentado pelos desenvolvedores. Todas as iniciativas até então consumadas foram vigorosamente criticadas, e denunciados o seu espontaneísmo, a sua precariedade teórica, e a falta de conhecimento e tecnologia necessários para a execução das propostas. Como resultado, o investimento público no setor refluíu consideravelmente. Segundo Slocum (1985), em 1973 havia apenas três projetos de tradução automática subsidiados pelo governo dos Estados Unidos. Em 1975, nenhum projeto teria sido contemplado. As iniciativas, pelo menos na América do Norte, ficaram bastante circunscritas a experiências isoladas (como o desenvolvimento do sistema SYSTRAN, por Peter Toma). Na Europa, o declínio do campo teria sido menos expressivo.

No início dos anos 1980, com o desenvolvimento tecnológico, o domínio da tradução automática voltou a receber maior atenção dos pesquisadores, principalmente na Europa e no Japão, mas sob novas bases e com outra aceção. A primeira fase da tradução automática havia provado que a automação do processo somente se tornaria viável se a complexidade da tarefa pudesse ser expressivamente reduzida. A idéia de uma tradução completamente automática de qualidade foi definitivamente abandonada. Propuseram-se, como alternativas, a redução do grau de complexidade dos textos a serem traduzidos, a redução do grau de automação do processo, ou mesmo a completa transformação da tarefa.

No primeiro caso, propôs-se que as ferramentas de tradução automática operassem não sobre quaisquer textos em língua natural, mas sobre textos específicos, que fizessem uso controlado da linguagem. Por "uso controlado" deve-se entender aqui uma de duas noções: o controle derivado da seleção (natural) de textos de determinada forma e sobre determinado conteúdo; ou o controle artificial derivado da imposição de uma forma padronizada para os textos a serem traduzidos. Em ambos os casos, restringe-se o escopo da tradução automática, que passaria a operar apenas sobre um subconjunto (otimizado) da língua natural, normalmente chamado "sublíngua" ou "sublinguagem".

O conceito de sublíngua natural aposta na idéia de que a especialização da forma e do conteúdo dos textos tratados é necessária e suficiente para a eliminação da ambigüidade. Está amparado na hipótese de que textos parentes (por afinidade temática, formal ou funcional) comportam uma série de invariâncias (de vocabulário, de estruturas sintáticas) que facilitaríam o processo de análise. Neste caso, quanto mais próximos e fixos (enlatados, padronizados) os textos, mais eficazes as traduções. É o que tem sido observado, com sucesso, no sistema Météo, que traduz boletins meteorológicos canadenses do inglês para o francês. Como a estrutura e o vocabulário

dos textos são (naturalmente) muito restritos e repetitivos, o sucesso da automação do processo de tradução amplia-se consideravelmente.

No entanto, é forçoso observar que essa delimitação implica o desenvolvimento de sistemas excessivamente especializados, de utilidade bastante localizada. Este é o caso, aliás, do próprio Météo, cujas tentativas de extensão (para o domínio dos manuais de aviação, por exemplo) foram todas malsucedidas. Além disso, a especialização da forma e do conteúdo dos textos tratados não tem podido funcionar como panacéia para muitos dos problemas encontrados, já que o próprio traçado das fronteiras temáticas e formais tem estado freqüentemente em discussão. Os gêneros textuais, por exemplo, têm revelado maior variabilidade interna do que o esperado: fracassaram todas as tentativas, na Lingüística do Texto, de construção de "gramáticas textuais", ou seja, de formalização de conjuntos finitos de regras que estabelecessem, de forma inequívoca, os princípios de formação de todos e apenas dos textos pertencentes a um determinado gênero (como o jornalístico, por exemplo) ou a uma determinada tipologia (como a narrativa).

Não se afirmará, evidentemente, que cada gênero textual não tenha suas próprias preferências vocabulares e sintáticas, cuja identificação poderia evitar muitos dos casos de ambigüidade presentes na língua do dia-a-dia. Mas têm sido reportados experimentos em que, apesar da restrição da forma e do conteúdo dos textos, a ambigüidade não vem sendo expressivamente reduzida. O Projeto Verbmobil, por exemplo, apesar de restringir o universo do discurso dos textos a serem tratados para o de uma conversa entre dois interlocutores sobre o lugar e o momento de um próximo encontro, tem deparado inúmeras ambigüidades residuais, que apenas o contexto de enunciação poderia resolver [Kay 1995]. Além disso, a ambigüidade não se reduz ao nível lexical apenas, mas envolve também o nível gramatical, cuja ambivalência nem sempre está relacionada a um domínio específico. Há muitas formas lingüísticas cuja ambigüidade transcende a forma ou o conteúdo do texto. É o caso, no português, de alguns morfemas flexionais, como {Ø}, que pode representar tanto o presente {cant-a-Ø-mos} quanto o passado {cant-a-Ø-mos}, independentemente do universo do discurso. E há itens lexicais que, mesmo no interior de contextos de uso bastante delimitados, podem ser ambíguos. A sumária delimitação do contexto de meteorologia não nos autorizará, sem risco de erro, a traduzir toda e qualquer instanciação da forma do português *tempo* como *weather*, em vez de *time*, por exemplo. Por mais que restrinjamos a forma ou a temática, é pouco razoável supor que não continuemos a enfrentar problemas relativos à ambigüidade nestes casos.

A proposição de uma sublíngua artificial traz a vantagem de evitar a busca de invariâncias textuais e o desenvolvimento de sistemas excessivamente especializados, mas passa a exigir a pré-edição, por um especialista, do texto-fonte. O controle artificial é provocado por uma de duas formas: pode ser derivado do estabelecimento de regras rígidas de redação dos textos, cuja forma passaria a ser induzida e padronizada (os textos a serem traduzidos deveriam submeter-se a alguns protocolos de redação, como o preenchimento de todas as elipses ou a explicitação de todas as relações anafóricas, por exemplo); ou pode consistir na sinalização, nos textos, por meio de marcações, etiquetas e outras formas de anotação, de informações de natureza metalingüística, que pudessem reduzir seu nível de ambigüidade. Restringe-se novamente o escopo de atuação da ferramenta, mas em outro sentido. Trata-se agora da tentativa de, na linha inaugurada

por Reifler em 1950, combinar esforço humano e mecânico de forma a reduzir os custos do processo de tradução - o que viria a ser conhecido como tradução automática auxiliada por humanos (*human-aided machine translation*, ou HAMT), conforme definição proposta por Hutchins e Somers (1992) e retomada por Boitet (1995). As técnicas de HAMT investiriam ainda na automação dos processos de tradução, mas prevendo a intervenção humana na edição do texto-fonte (pré-edição), na tradução do texto-fonte para o texto-alvo (interação) ou na edição do texto-alvo (pós-edição). Estaria abandonada a idéia de um processo de tradução completamente automático.

Por fim, uma última alternativa para a redução da complexidade do problema consistiria em alterar significativamente a definição do que seria "tradução automática". Oferecem-se, neste caso, dois caminhos: o de passar a conceber a tradução automática como suporte ao tradutor humano, ou como processo de geração paralela de textos originais, em línguas diferentes, a partir de uma mesma representação da informação.

A primeira opção vem sendo chamada de tradução humana auxiliada por máquina (em inglês: *machine-aided human translation*, ou MAHT, em oposição a HAMT). A opção por técnicas de MAHT é bastante pessimista: o processo de automação da tradução estaria agora circunscrito ao desenvolvimento de ferramentas de apoio ao tradutor humano, como dicionários bilíngües, corretores ortográficos e revisores gramaticais.

A segunda opção procura reduzir o domínio da tradução ao da comunicação multilíngüe, admitindo que a ferramenta deveria tomar, como ponto de partida, não um texto em língua natural, mas um conjunto de informações a respeito da realidade, representado pelo falante de uma língua natural específica, por meio de sistemas de representação (de natureza dialogada, por exemplo) em que os dados seriam registrados de forma não necessariamente lingüística. Admite-se que o objetivo da tradução seria recuperar o conteúdo informativo registrado em um determinado texto-fonte, o que poderia ser facilitado se fosse adotado, em lugar da forma lingüística (normalmente ambígua e indeterminada), um outro tipo de estruturação (e representação) dos dados, como o preenchimento, por exemplo, de campos pré-determinados em um formulário padronizado.

Em todos estes casos, reconhece-se que a grande dificuldade dos sistemas de tradução automática é justamente o processo de análise e interpretação dos enunciados em língua natural. Diferentemente do que ocorre em outras ferramentas computacionais (como a revisão gramatical automática, por exemplo), a tradução automática é particularmente sensível à representação do conteúdo semântico das sentenças e dela profundamente dependente. Como diria Santos (1995):

"A operação de fazer a transição de uma língua para outra - consistindo afinal na tradução de itens lexicais da língua de partida para itens lexicais da língua de chegada - é a parte mais trivial de todo o processo, e o ônus da tarefa de traduzir (pelo menos se for encarada do ponto de vista computacional) recai sobre as competências monolíngües envolvidas".

O processo de análise dos enunciados das línguas naturais constitui, sem dúvida, a grande dificuldade no que concerne ao campo da tradução automática. Acreditou-se, a princípio, que a especificação dos dicionários e das gramáticas das línguas naturais seria suficiente para o equacionamento da estrutura semântica das sentenças. Mas foi

observado, desde cedo, que há muito mais do que simplesmente trocas lexicais e sintáticas no processo pelo qual um tradutor humano processa uma sentença. Bar-Hillel, por exemplo, um entusiasta de primeira hora da possibilidade de tradução completamente automática com qualidade, concluía, em 1960, depois de uma década de investigação, pela impossibilidade da tarefa caso não fosse incorporada, ao modelo, uma enciclopédia que contivesse todo o conhecimento humano disponível.

Ele analisava problemas relativos à tradução de sentenças simples do francês como "Le bois était dans la maison". A palavra "bois" pode ser traduzida, em português, de duas formas diferentes: "lenha" (material combustível) e "bosque" (pequena floresta). Nenhum tradutor humano sugeriria, no entanto, que "O bosque estava dentro da casa" seria uma tradução possível para a sentença indicada. A rigor, durante a tradução da sentença, é mesmo de esperar que o tradutor humano nem mesmo perceba a ambigüidade da forma "bois", embora em estado dicionário ela possua efetivamente mais de uma acepção. Na verdade, a sentença não seria ambígua para o falante do francês, e sua ambigüidade seria antes artificialmente induzida pela máquina, para quem tanto "A lenha estava dentro de casa" quanto "O bosque estava dentro de casa" seriam traduções possíveis para a sentença assinalada. A máquina, portanto, por causa da falta de informações adicionais, terminaria por expandir o nível e o número das ambigüidades presentes nos enunciados em língua natural.

O problema, segundo Bar-Hillel, seria como coibir a alternativa "O bosque estava dentro da casa", perfeitamente possível segundo um uso mecânico do dicionário e da gramática, sem incluir, no modelo, todo o conhecimento de mundo implicitamente mobilizado pelo tradutor humano, que sabe que, neste caso, em função da estrutura semântica da sentença, a "casa" é um continente que, mercê de suas dimensões e de seu uso social, não poderia conter, como conteúdo, uma "pequena floresta", principalmente se haveria disponível a alternativa de que pudesse conter "lenha", o que de fato é comum, tendo em vista que muitas casas, nas regiões mais frias, dispõem de sistemas de aquecimento que utilizam "lenha" como material combustível. Apenas a incorporação de todo esse conjunto de informações, e tantas outras a ele relacionadas (como o fato de que a combustão da lenha provoca calor, de que os seres humanos precisam de calor para sobreviver, etc.) - de natureza bem pouco computacionalizável, porque interminável - tornaria possível à máquina prover traduções razoáveis.

Em defesa da máquina, pode-se dizer que a) ela poderia prover todas as traduções possíveis para a sentença, de forma que o leitor poderia optar pela mais apropriada, o que, embora venha a constituir, em algum grau, certo constrangimento (e certo incômodo desnecessário), não provaria a inutilidade da tradução automática; e b) sentenças como a indicada não seriam exatamente freqüentes e não representariam, do ponto de vista estatístico, problema para o funcionamento global de uma ferramenta de tradução automática. Sobre estes dois pontos caberia dizer que não há evidência empírica ou científica a seu favor, mas antes na direção contrária; e que ambos constituem antes impressões do senso-comum (sombras na caverna, portanto) do que experiências efetivamente comprováveis de sucesso no processo de construção de ferramentas automáticas. Se não, vejamos.

O primeiro argumento está diretamente relacionado à idéia de que qualquer tradução é melhor do que nenhuma tradução. No entanto, a possibilidade de a máquina prover todas as traduções possíveis para as sentenças de entrada está associada a vários

problemas. O mais grave talvez seja o fato de que não se traduzem, normalmente, sentenças apenas, mas textos inteiros. Na medida em que a mesma sentença de entrada pode corresponder a várias sentenças de saída, instala-se o risco da explosão combinatória, diretamente proporcional ao número e à complexidade das sentenças envolvidas. Outro problema concerne à cooperatividade do usuário: embora saibamos que o usuário está geralmente disposto a aceitar, em larga medida, falhas da ferramenta, ele o faz apenas quando percebe que, feitas as contas, o processo é simplificado. Nenhuma garantia neste sentido pode ser dada por um sistema que exija a intervenção do usuário a cada caso de ambigüidade. Por fim, este primeiro argumento está também relacionado a duas visões de tradução que caberia discutir: a chamada "tradução crua" (*raw translation*) e a chamada "tradução grosseira", também conhecida por "tradução rudimentar" (*rough translation*).

Tradução rudimentar ou grosseira seria aquela utilizada tão-somente como instrumento para que o usuário possa tomar a decisão de requisitar ou não uma tradução humana para o texto. Em última instância, seria apenas a indicação, sempre superficial, e bastante imprecisa, de uma idéia ou tema geral no texto, ou de que o texto conteria alguma palavra-chave que justificaria sua tradução humana. Serviria, por exemplo, como instrumento para sistemas de busca de informações, mas jamais como objeto de publicação ou difusão. A tradução não tem aqui nenhum compromisso com a qualidade, e os resultados seriam tão problemáticos que um tradutor humano, por considerá-los de muito pouca validade, seguramente julgaria menos dispendioso (e mais rápido) retraduzir todo o texto a partir do nada do que tentar corrigir os problemas verificados. Este tipo de tradução, bastante robusta, completamente automática, rápida e de baixo custo, não requereria (para a sua produção) nenhum tradutor ou revisor profissional, e poderia ser operada por um usuário comum, na medida em que não exigiria nenhuma espécie de pré-edição do texto de entrada (talvez apenas a confirmação das propostas do sistema) ou de pós-edição do texto traduzido (limitada, quando muito, a operações de formatação). Em compensação, seus resultados são de validade bastante controversa. A maior parte dos sistemas de tradução automática atualmente franqueados ao público (como SYSTRAN, CAT, etc.) incidiria neste caso.

Tradução crua, ao contrário, seria aquela planejada para a pós-edição, ou seja, seria a produção de resultados propositadamente parciais (e imperfeitos) para que pudessem servir de ponto de partida para a correção (ou tradução) humana, feita por revisores (ou tradutores) especializados. Teria como mérito reduzir o tempo de tradução de um texto, na medida em que resolveria problemas básicos para o tradutor humano, deixando-o livre para se ocupar apenas de casos mais espinhosos (ou de difícil resolução por parte da máquina). À ferramenta caberia produzir, portanto, o primeiro rascunho (jamais o texto final), a ser trabalhado pelo especialista, para o qual seria observado um expressivo ganho de produtividade (em torno de 40 a 50% do tempo dedicado a cada lauda, segundo os registros reportados em Boitet 1995). Os sistemas desta natureza envolveriam conhecimento mais especializado e interação mais intensa com o usuário, a quem também caberia, com freqüência, a pré-edição do texto-fonte. Em função do treinamento exigido, dificilmente seriam úteis ao grande público. Quase todos os grandes sistemas corporativos (DUET, da Sharp; HICAT, da Hitachi; ATLAS-II, da Fujitsu; METAL, da Siemens; AS-TRANSAC, da Toshiba; PIVOT, da NEC; etc.) funcionam desta maneira.

A produção de traduções cruas e a de traduções grosseiras constituiriam, em última análise, duas estratégias de MAHT, ou seja, de tradução humana auxiliada por máquinas, já que o papel do tradutor humano, se considerado todo o processo, longe de subsidiário, seria fundamental. O resultado, em ambos os casos, fica normalmente muito aquém das expectativas de um usuário não-familiarizado com a complexidade da tarefa. Com efeito, ambas as abordagens estão muito distantes da idéia de tradução automática tal como concebida originalmente.

O segundo argumento a favor da máquina envolve uma impressão equivocada da linguagem. A ambigüidade não é um fenômeno periférico e marginal (e por isso rarefeito) nos enunciados em língua natural. Ela é constitutiva da própria linguagem, na medida em que todos os enunciados sofrem de vagueza e de indeterminação, se isolados os índices contextuais (relativos ao contexto extratextual) e co-textuais (relativos ao contexto intratextual) que provocam, com freqüência, a ilusão de que os enunciados seriam exatos e precisos.

Há, evidentemente, marcas mais ostensivas dessa ambigüidade. É o caso da ambigüidade categorial e da ambigüidade léxica, por exemplo. Existem na língua portuguesa, efetivamente, formas que podem indicar mais de uma classe gramatical, ou que, mesmo indicando a mesma classe gramatical, podem comportar acepções diferentes. O dicionário Aurélio registra, por exemplo, doze classificações gramaticais diferentes para a forma "que", que não é exatamente incomum nos textos em língua portuguesa; e raras são as entradas, no mesmo dicionário, que comportam apenas uma acepção. As formas homônimas (como "banco") representam apenas a parte mais radical dessa possibilidade de variação. Mesmo formas aparentemente não-ambíguas (como "abacaxi") comportam muitos sentidos diferentes ("a planta como um todo", "apenas a parte comestível da fruta"), principalmente se incorporadas as variedades regionais e sociais da língua (em que "abacaxi" pode figurar, por exemplo, ora como "coisa complicada", ou "pessoa desagradável", ou até mesmo "dançarino desajeitado").

A par da ambigüidade em nível lexical, a ambigüidade sintática é também facilmente perceptível. E, como no caso anterior, não é exatamente rara na língua portuguesa. A análise dos casos mais conhecidos talvez faça parecer que o fenômeno seja localizado a algumas construções específicas, como a adjunção ao verbo. No exemplo, muito explorado, de "A menina viu o menino com o telescópio", não se poderá precisar, fora de contexto, se o sintagma "com o telescópio" constitui um modificador do verbo "viu" ou do seu objeto "o menino". Mas é preciso perceber que mesmo sentenças de estruturação supostamente mais clara contêm ambigüidades estruturais. A exatidão de "A menina chegou atrasada" desaparece, por exemplo, ao considerarmos que há pelo menos duas possibilidades diferentes de enquadramento do sintagma "atrasada": como adjunto a "menina" (com o qual, aliás, concorda) ou como adjunto a "chegou" (a quem parece modificar: "chegar atrasadamente"). A escolha por uma entre essas possibilidades, embora talvez irrelevante para o português, torna-se estratégica no processo de tradução para línguas que vão optar por meios bastante diferenciados de representação dos mesmos fenômenos.

De resto, a ambigüidade em língua natural se espalha em inúmeras outras direções, como na recuperação das relações anafóricas e no preenchimento das elipses, por exemplo. Mas a ambigüidade mais insidiosa é justamente aquela que não se revela na superfície do texto, mas se reveste de suposta univocidade, a nos convencer de que a

sentença admite uma única e exclusiva possibilidade de interpretação. Essa monovalência esconde o fato de que a univocidade dos enunciados lingüísticos deriva muito mais de fatores contextuais (extralingüísticos, portanto, e em princípio irrepresentáveis para a máquina) do que de fatores propriamente lingüísticos. Estaríamos de tal forma hipnotizados pelo contexto, ou nele inseridos (como prisioneiros da caverna que não vêem senão as sombras projetadas), que seríamos incapazes de perceber o quanto há de ambigüidade nos enunciados cotidianos, mais ou menos da mesma forma que o tradutor humano muitas vezes nem mesmo percebe que a forma "bois", explorada por Bar-Hillel, poderia ter sido traduzida de outra forma.

Considere-se, a este propósito, apenas mais um exemplo: a tradução, para o inglês, de uma sentença banal (e a princípio não-ambígua) da língua portuguesa, como "João está esperando por Maria". Quanta ambigüidade não se evidenciará nesta sentença, ao se perceber que, no processo de tradução para a língua inglesa, caberá à ferramenta decidir se o "esperar" em tela corresponde aos verbos do inglês "wait", "expect" ou "hope", todos eles associados, no dicionário, à mesma forma da língua portuguesa. A ambigüidade, neste caso, como no caso de "bois", talvez não se perceba senão retrospectivamente, pela necessidade de adaptação do enunciado a uma língua (ou a uma situação) em que outros níveis de análise acabam por se revelar necessários.

Estabelece-se, portanto, o consenso de que o desenvolvimento de recursos lingüísticos (o chamado "lingware"), embora necessário, não é suficiente para o desenvolvimento de ferramentas computacionais de tradução automática. Tem sido afirmado, com freqüência, que a tradução é um problema menos lingüístico do que cognitivo, e os sistemas exclusivamente baseados em conhecimento lingüístico vêm incorporando progressivamente outros elementos externos à linguagem, como os requeridos por Bar-Hillel.

Do exposto acima percebe-se o quanto o campo da tradução automática não pode ser delimitado de forma única, constituindo antes uma dispersão, que varia conforme todo um conjunto de pressupostos sobre o que seja a linguagem humana, sua natureza, sua estrutura, sobre o papel do conhecimento lingüístico e do conhecimento de mundo na interpretação dos enunciados, e - talvez principalmente - sobre o que seja tradução.

3. Tradução Humana e Tradução Automática

Do ponto de vista teórico, a tradução humana é normalmente considerada como processo de transferência ou de substituição de conteúdo entre duas línguas naturais: a língua de origem, também chamada língua-fonte, ou língua de partida; e a língua de destino, também chamada língua-alvo, ou língua de chegada. Para Catford (1980), tradução é "substituição do material textual de uma língua pelo material textual equivalente em outra língua". Para Nida (1975) é a preservação da carga semântica da língua-fonte na língua-alvo. Dessas concepções tradicionais, derivam os princípios que deveriam nortear a boa tradução. Para Tytler, um pioneiro dos estudos da tradução no século XVIII, "a tradução deve reproduzir em sua totalidade a idéia do texto original; o estilo da tradução deve ser o mesmo do texto original; e a tradução deve ter toda a fluência e a naturalidade do texto original" (*apud* Arrojo 2000). Para Yebra (1982), seria imperativo "decir todo lo que dice el original, no decir nada que el original no

diga, y decirlo todo con la corrección y naturalidad que permite la lengua a la que se traduce".

Pelo que se pôde perceber na seção anterior, é forçoso admitir que "traduzir não pode ser meramente o transporte, ou a transferência, de significados estáveis de uma língua para outra, porque o próprio significado de uma palavra, ou de um texto, na língua de partida, somente poderá ser determinado, provisoriamente, através de uma leitura" [Arrojo 2000]. Em outras palavras: o significado de uma sentença não está totalmente encapsulado nela mesma, mas é a ela atribuído por um leitor, que mobiliza, para isso, conhecimento externo à própria sentença. Sem esse conhecimento exógeno, a sentença não passa de um amontoado de símbolos isolados, cada um dos quais apontando para múltiplos diferentes significados (como prevê o próprio dicionário), todos à espera de uma decisão (a ser tomada apenas pelo leitor) que os permitirá combinarem-se uns aos outros para que possam vir a constituir um todo. Fora de contexto, ou seja, despregada de uma situação específica de leitura, a sentença não significa; ela é apenas um conjunto de pistas que podem conduzir a muitos destinos diferentes. Se essas pistas, à primeira vista, parecem conduzir a um único lugar - o chamado "sentido literal" -, sobre o qual concordam muitos (senão todos os) leitores, isso não se deve ao fato de que a sentença contém ela mesma a chave de sua interpretação, mas ao fato de que a sua comunidade de leitores está toda ela interpelada por uma mesma convenção de leitura, que faz que todos mobilizem, para essa mesma sentença, o mesmo repertório de conhecimento. Ou seja, não existem sentenças intrinsecamente não-ambíguas, capazes elas mesmas de produzir o seu próprio significado, independentemente do leitor; a aparente não-ambigüidade de algumas sentenças é antes produto de uma convergência, entre os leitores, sobre o conhecimento a ser mobilizado para sua interpretação.

Este caráter extrínseco do significado - o fato de ser derivado do processo de leitura ao invés de estar contido no objeto da leitura (e dele ser extraído, ou "decifrado") - constitui, sem sombra de dúvida, o principal problema da tradução automática, porque implica que, para traduzir sentenças, não basta representar o conhecimento contido na sentença; deve-se representar também o conhecimento mobilizado pelo leitor da sentença. Não basta combinar itens lexicais previamente associados em um dicionário bilíngüe; deve-se também escolher, a partir de conhecimento que não está representando na sentença, mas no próprio contexto de leitura, entre os vários significados disponíveis para o mesmo item lexical. Não fosse assim, meio século de investimentos na construção de ferramentas que automatizassem o processo tradutório não provocariam, no início do século XXI, o gosto amargo da decepção que tanto provocam. E é importante assinalar, mais uma vez, que esta não se trata de uma limitação válida apenas para o texto literário, objeto privilegiado de estudo dos teóricos da tradução; a interpretação de qualquer sentença somente se torna possível de fora (ou seja, do leitor) para dentro. Submetida ao crivo da ausência de referência externa, como é o caso do computador, qualquer sentença, mesmo a mais banal (como "Le bois était dans la maison"), explode pela quantidade de significados que passa a veicular. O contexto-zero do computador prova que há apenas sombras onde julgávamos antes ver luz.

Na medida em que cada leitor mobilizaria, em cada leitura, um conhecimento que lhe seria específico (no contexto bastante delimitado da leitura), a tradução - de qualquer texto - estaria inapelavelmente contaminada de subjetivismo, o que frustraria

qualquer possibilidade de automação. Para muitos teóricos, o problema delimitado - da tradução automática - é, por este motivo, uma contradição em termos, inextricável e irresolúvel. Reconhece-se, no esforço de tradução, um gesto de (re)criação que, sendo tão caracteristicamente humano (por intuitivo), não poderia ser replicado por nenhum organismo cibernético, ainda que antropomórfico.

No extremo oposto a esta posição estão aqueles que acreditam que o processo de tradução possuiria uma natureza sobretudo intersubjetiva, dado que o excesso de subjetivismo implicaria a própria intraduzibilidade dos enunciados lingüísticos, e não explicaria como, afinal de contas, a partir de experiências tão singulares, acabamos por conseguir nos entender uns aos outros. Em outras palavras: embora cada leitor possa mobilizar seu próprio conhecimento no processo de interpretação e de tradução, os leitores devem necessariamente mobilizar, também, algum conhecimento compartilhado, comum a outros leitores, ou não serão capazes de interagir eficazmente pela linguagem. O caráter comunitário desse conhecimento compartilhado asseguraria, ao processo de interpretação, alguma regularidade, e permitiria, conseqüentemente, a possibilidade de alguma automação do processo de tradução.

Os sistemas de tradução automática partem dessa possibilidade e se diferenciam pelas soluções que a ela endereçam. Tomam a tradução, não como um processo de "recriação", mas de "recomposição", que, em certa medida, deveria ou poderia preceder o processo de recriação. Embora não se possa falar em complementaridade entre as duas perspectivas (a "recomposicional" e a "recriacional"), as abordagens de tradução automática postulam que a re-representação do significado compartilhado do texto-fonte não é de todo inútil e que, ainda que não sirva à emulação do comportamento enunciativo do texto original, pode ser utilizada para outras finalidades. Principalmente, a tradução automática acredita que, para uma parcela expressiva dos textos produzidos em qualquer língua natural, "interagir" é sobretudo "comunicar", ou seja, veicular o conteúdo informacional codificado pela mensagem. Nestes casos, estaria licenciada a abstração do contexto para a análise e a reconstrução dos enunciados, o que poderia ser feito apenas pelo recurso à sua estrutura interna.

4. Estratégias de Tradução Automática

Como já anunciado na segunda seção, os sistemas de tradução automática variam ao sabor das premissas que os orientam, e há inúmeras formas de agrupá-los. Os sistemas podem ser analisados, por exemplo, em relação ao número de línguas envolvidas: há sistemas de tradução bilíngües ou multilíngües. Estes sistemas podem ser, por sua vez, unidirecionais ou bidirecionais, na medida em que permitem (ou não) que a língua-alvo possa ser também a língua-fonte, ou vice-versa. Uma outra perspectiva a partir da qual os sistemas podem ser analisados diz respeito ao paradigma utilizado. Há sistemas de tradução baseados em regras (sistemas simbólicos) e há sistemas de tradução baseados em casos (sistemas subsimbólicos, ou conexionistas). Há sistemas de tradução que utilizam apenas recursos lingüísticos (dicionários e gramáticas) e há sistemas de tradução que incorporam também outras formas de conhecimento (bases de conhecimento, ontologias, *corpora*, etc.). Os sistemas podem ser classificados também de acordo com o papel do usuário: há sistemas interativos (que requerem a ajuda ou a intervenção do usuário) e há sistemas não-interativos (que limitam a intervenção do usuário ao pré-processamento ou ao pós-processamento). Por fim, todos os sistemas

podem ser classificados segundo as estratégias utilizadas: a tradução direta ou indireta. O objetivo desta seção será analisar alguns desses paradigmas e estratégias, indicando, em cada caso, problemas e soluções.

4.1 Dois paradigmas: KBMT e EBMT

Dois grandes exemplos de alternativas aos modelos de tradução baseados unicamente em conhecimento lingüístico são a tradução automática baseada em conhecimento (ou *Knowledge-Based Machine Translation*, ou simplesmente KBMT,), e a tradução automática baseada em exemplos (*Example-Based Machine Translation*, ou EBMT). No caso da KBMT, prevê-se, além dos recursos lingüísticos tradicionais (dicionários e gramáticas), uma versão da enciclopédia (ou base de conhecimentos) reclamada por Bar-Hillel já em 1960. No caso de EBMT, procedimentos estatísticos são utilizados para comparar traduções já realizadas (e revisadas) com tarefas em curso. Ambos os paradigmas pretendem enriquecer o ponto de partida do processo de tradução, aparelhando a máquina com conhecimento adicional para que as ambigüidades e a indeterminação das línguas naturais possam ser reduzidas. Em ambos os casos, percebe-se que o processo de tradução, feito anteriormente no sentido de baixo para cima (*bottom-up*), ou seja, utilizando apenas o conhecimento lingüístico contido na própria sentença, começa progressivamente a comportar variações que procuram localizar, em primeiro plano, informações de natureza macroestrutural (como o universo do discurso, por exemplo), para, e apenas então, processar integralmente o material lingüístico. São estratégias, portanto, que pretendem representar (ou emular) o conhecimento mobilizado pelo leitor durante o processo de tradução.

Considere-se, a título de ilustração, a tradução automática, para o português, da sentença do francês proposta por Bar-Hillel, "Le bois était dans la maison". Os sistemas baseados em conhecimento (KBMT), a par do dicionário e da gramática que permitiram a associação entre os itens lexicais e as construções sintáticas do francês e do português, pressupõem o desenvolvimento de uma base de conhecimento cujos elementos estejam por exemplo, inter-relacionados por meio de uma grandeza escalar (dimensão), na qual estariam situados, entre outros, os conceitos "lenha", "casa" e "bosque", de tal forma que se possa inferir que lenhas podem estar contidas em casas e bosques, casas podem estar contidas em bosques mas não em lenhas, e bosques não podem estar contidos nem em casas nem em lenhas. Essa informação - que não está representada na sentença, mas é antes derivada da experiência de mundo do tradutor humano - permitiria bloquear a possibilidade da tradução "O bosque estava dentro da casa", na medida em que seria violada a condição prescrita na base de conhecimento. O mesmo resultado poderia ser alcançado de inúmeras outras maneiras, a depender da estrutura da base de conhecimento, que varia de sistema para sistema.

No caso dos sistemas baseados em exemplos (EBMT), a desambigüização seria feita por consulta a um repertório de traduções prévias, em que seria observado que a ocorrência, por exemplo, do par "lenha" e "casa" seria estatisticamente mais relevante do que a ocorrência do par "bosque" e "casa". Essa maior probabilidade orientaria o processo de seleção, e muitas outras variáveis poderiam ser aí intervenientes (como a presença, por exemplo, da preposição "dentro", ou fato de que "lenha" ocupa a posição de sujeito gramatical).

É forçoso reconhecer, contudo, que as estratégias, tanto no caso de KBMT quanto no caso de EBMT, encontram-se ainda em estágio experimental. Oferecem-se vários problemas para ambas as abordagens, muitos dos quais aparentemente insolúveis. Os sistemas de EBMT, por exemplo, estão amparados em juízos de similaridade (entre sentenças já traduzidas e sentenças a serem traduzidas) de natureza bastante controvertida. A similaridade estatística entre enunciados lingüísticos muitas vezes não corresponde à identidade de significado. Não se pode afirmar, categoricamente, que uma mesma estrutura sintático-semântica, utilizada em contextos diferentes, seja portadora do mesmo significado. Provam-no todas as formas de uso figurativo da linguagem (caso da metáfora, por exemplo). Além disso, a versatilidade dos sistemas de EBMT está geralmente reduzida à amplitude do conjunto de traduções prévias, cuja variabilidade sacrifica, por sua vez, a exatidão do sistema, na medida em que passa a incluir, para uma mesma estrutura, possibilidades diversas. Ou seja, os sistemas de tradução baseados em exemplos têm sido prisioneiros de uma lógica perversa: quanto mais limitado o *corpus* de sentenças previamente traduzidas, piores os resultados; quanto mais rico, mais inexatas as respostas.

Algumas dessas mesmas ciladas valem para os sistemas de KBMT. É o caso, por exemplo, da construção de bases de conhecimento, enciclopédias e ontologias que venham a representar o conhecimento que o homem tem do mundo. O conhecimento humano tem parecido, em muitos instantes, não-formalizável. Nem sempre é discreto, preciso e, principalmente, nunca é estático. Sua organização é variável, e o repertório de conceitos e de relações entre conceitos é dependente não apenas da cultura, mas da experiência muitas vezes pessoal e intransferível dos interlocutores. A comunicação parece basear-se antes em um jogo de inferências de regras pouco conhecidas.

Também a categorização da realidade, fundamental para a organização de ontologias, é problema complexo, com desdobramentos na Filosofia, nas Ciências Sociais e na Psicologia, e não é de todo incontroverso que a realidade possa ser de fato encapsulada nas unidades descritivas que vêm sendo propostas. Os modelos de categorização sugeridos pela Psicologia, por exemplo, são bastante variados: ora admite-se que categorias são integradas por todos os elementos que respeitam um conjunto finito e discreto de condições necessárias e suficientes de pertencimento categorial (o chamado modelo clássico, de inspiração aristotélica); ora propõe-se que seriam formadas por todos os elementos para os quais se pudesse estabelecer uma relação de parentesco (semelhança) com uma instância considerada prototípica (o modelo dos protótipos); ora pulveriza-se esta mesma instância prototípica, admitindo-se a existência de várias instâncias exemplares (o modelo dos exemplares); ora afirma-se que os processos de filiação a uma determinada categoria são externos à própria categoria e governados pelo contexto (o modelo, por exemplo, da teoria da dependência conceitual, que prevê uma realidade compartimentalizada em planos, scripts e frames). Na medida em que as bases de conhecimento fazem, obrigatoriamente, escolhas entre as inúmeras teorias disponíveis, sacrificam inevitavelmente o alcance de suas propostas, e passam a estar confinadas a uma representação antes parcial do conhecimento humano.

Transfere-se, portanto, o problema da ambigüidade lingüística, criando-se o problema da ambigüidade contextual. Se se revela efetivamente plausível que a tradução não seja uma atividade estritamente lingüística, mas lingüístico-cognitiva, não se revela igualmente óbvio em que medida a introdução de categorias cognitivas (ou de

exemplos prévios) poderia resolver a ambigüidade e a indeterminação das formas lingüísticas. Os modelos propostos têm ainda um sabor experimental, e se vêm efetivamente aprimorando os resultados dos sistemas anteriores, fazem-no em uma escala ainda relativamente modesta.

4.2 Duas estratégias: Tradução Direta e Tradução Indireta

A par dos recursos utilizados (se apenas dicionários ou gramáticas, ou se também bases de conhecimento ou *corpora*, como indicado na seção anterior), os sistemas de tradução automática diferem entre si também em relação às estratégias utilizadas. São duas, neste caso, as possibilidades: a tradução direta (os chamados sistemas de primeira geração) ou indireta (sistemas de segunda geração). A tradução indireta admite ainda duas variações: as abordagens de transferência (sintática ou semântica), e aquelas realizadas por meio de uma língua pivô, intermediária, também chamada "interlíngua".

A tradução direta prevê, em linhas gerais, que a língua-alvo seja considerada o próprio instrumento de análise da língua-fonte. Ou seja, não haveria, em princípio, nenhum estágio intermediário entre língua-fonte e língua-alvo. O vocabulário da sentença de entrada seria automaticamente vertido para a língua-alvo por meio de um dicionário bilíngüe, com a ajuda, talvez, de algum processamento morfológico. Uma vez geradas as equivalências lexicais na língua-alvo, haveria algum reordenamento (bastante superficial e localizado) dos itens lexicais, para produzir resultados mais aceitáveis (como a posposição do adjetivo, por exemplo, no caso das traduções do inglês para o português). Não haveria propriamente processamento sintático das sentenças originais da língua-fonte, ou qualquer outro tipo de processo semântico.

Em última instância, os sistemas de tradução direta constituiriam sistemas de tradução palavra-por-palavra, com a possibilidade de alguma pós-edição, automática, dos resultados. A Figura 1 abaixo ilustra a arquitetura geral de um sistema desta natureza.

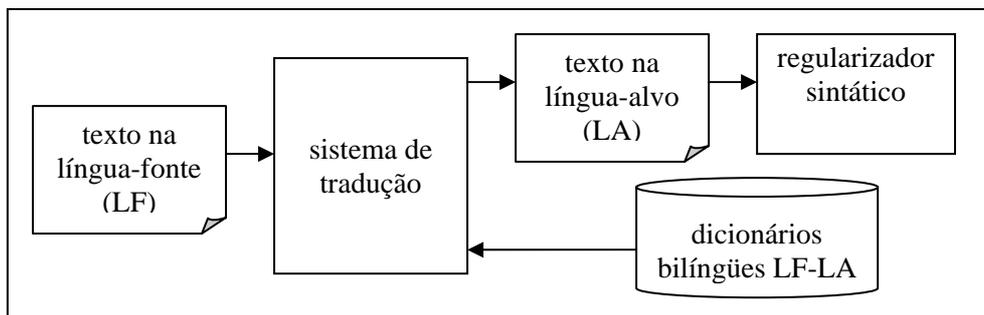


Figura 1. Arquitetura geral de um sistema de tradução direta

Este modelo de tradução já provou há muito não ser adequado, não apenas porque há uma relação de muitos para muitos entre os conjuntos de palavras que integram as línguas naturais, mas também porque são freqüentes as expressões idiomáticas e de sentido formulaico (isto é, não composicional) que não podem ser traduzidas a partir de suas unidades constituintes. Provérbios, por exemplo, perdem completamente o sentido quando traduzidos a partir das palavras que os compõem. É forçoso reconhecer, no entanto, que sistemas desta natureza podem passar a incorporar dicionários específicos de expressões idiomáticas (e mesmo de sentenças inteiras recorrentes), ou podem ser enriquecidos para lidar com a falta de correspondência estrita entre o vocabulário das duas línguas, criando-se, por exemplo, regras de desambigüização categorial amparadas em alguma sensibilidade ao co-texto mínimo (à esquerda e à direita).

O grande obstáculo ao sucesso dos sistemas de tradução direta consiste no grau de distanciamento das estruturas entre as línguas a serem traduzidas. O sistema está amparado na hipótese, hoje considerada ingênua, do isomorfismo sintático entre as línguas naturais, seja na sua versão forte (o mapeamento sintático é completamente dispensável), seja na sua versão mais fraca (o mapeamento sintático pode ser localizado em algumas construções bastante específicas, envolvendo quase sempre itens lexicais contíguos).

Contra essa possibilidade, acumulam-se hoje evidências de toda sorte. Observa-se hoje que mesmo línguas historicamente muito próximas - como o português e o espanhol, por exemplo - envolvem processos de gramaticalização muito diferentes, e que essas diferenças de estruturação sintática não são acidentais ou excepcionais, mas extremamente freqüentes, e não podem ser negligenciadas no processo de tradução, sob o risco de serem produzidos resultados inúteis. Traduzir uma construção bastante corriqueira do português, como "Gosto de Pedro", para a correspondente em espanhol, "Me gusta Pedro", a par dos problemas relacionados à ambigüidade de "gosto" (que pode ser substantivo ou verbo, sem que o co-texto possa resolvê-lo), envolveria inverter completamente as relações sintáticas estabelecidas pelo verbo, fazendo do sujeito o objeto da oração, e do objeto, o sujeito (ainda que posposto). Sem esta inversão, dramática para a máquina, os resultados seriam agramaticais, e quiçá mesmo ininteligíveis para o usuário final monolíngüe, desavisado das diferenças entre as duas línguas.

No entanto, em defesa dos sistemas de tradução direta é preciso dizer que surgiram no início da história da tradução automática, quando eram ainda utilizadas calculadoras numéricas para o processamento das informações. A ingenuidade teórica foi, em muitos casos, derivada de restrições de natureza prática, em um momento em que não havia ainda tecnologia e recursos disponíveis para um processamento mais refinado das línguas naturais.

A tradução indireta prevê o desenvolvimento de uma forma de representação intermediária entre a língua-fonte e a língua-alvo. Esta forma de representação pode ser dependente das línguas envolvidas, no sentido de constituir uma interface específica (unidirecional ou bidirecional), ou pode ser independente tanto da língua-fonte quanto da língua-alvo, procurando organizar-se como uma outra língua, artificial, autônoma, neutra, porém mais adequada ao processamento automático (porque livre de ambigüidade, por exemplo). No primeiro caso, fala-se em tradução indireta baseada em transferência; no segundo, em tradução indireta baseada em interlíngua.

A tradução por transferência envolve o desenvolvimento de uma metalíngua entre a língua-fonte e a língua-alvo. Esta metalíngua pode possuir caráter ora sintático, ora semântico, ora sintático-semântico, e consiste, quase sempre, em uma tabela de correspondências entre as duas línguas envolvidas. Na medida em que esta metalíngua seria diretamente decalcada da língua-fonte e da língua-alvo, prevendo apenas suas possibilidades de combinação, não consistiria, ainda, uma interlíngua, mas tão-somente um filtro necessário para o equacionamento das diferenças (principalmente estruturais) entre as duas línguas. A Figura 2 abaixo apresenta a arquitetura geral de um sistema de tradução por transferência.

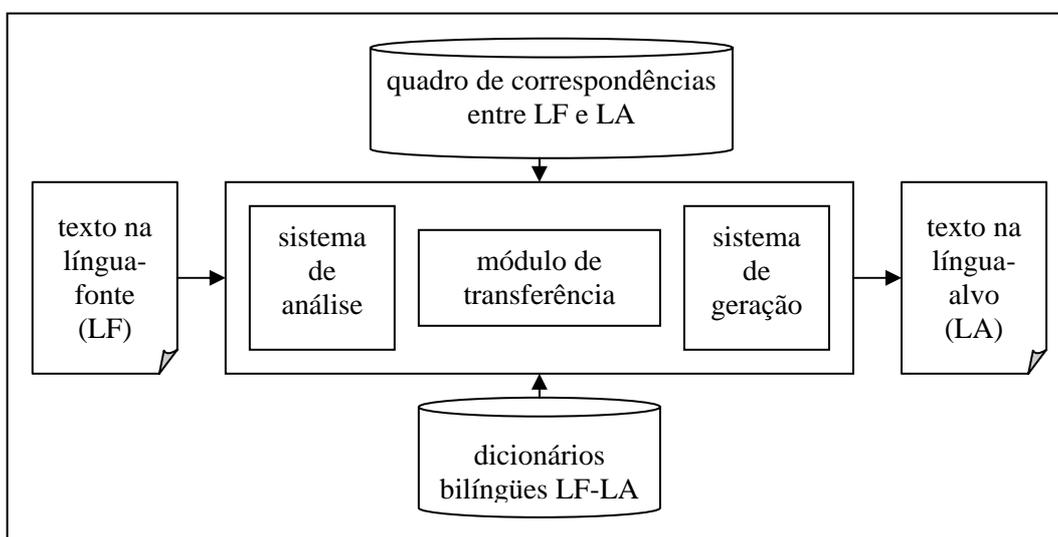


Figura 2. Arquitetura geral de um sistema de tradução por transferência

A abordagem baseada em interlíngua consiste, em última análise, no aprofundamento da abordagem baseada em transferência. A metalíngua de análise ganharia, neste caso, o estatuto de componente autônoma, a princípio desligada, seja da língua-fonte, seja da língua-alvo. Ofereceria um sistema de representação de natureza não-ambígua, para o qual seriam vertidas as informações disponíveis na sentença de entrada, e da qual seriam geradas as informações a serem incluídas nas sentenças de

saída. Diferentemente do que ocorre em um sistema de transferência, esta representação seria (completamente) abstrata, no sentido de modularizar o processo de tradução, desmembrando-o em duas etapas marcadamente independentes: a projeção ou representação do texto-fonte na língua intermediária seria feita independentemente da língua-alvo; e a geração do texto de saída, na língua-alvo, a partir desta representação intermediária, seria executada à revelia das informações sobre o processo de análise. A Figura 3 apresenta a arquitetura geral de um sistema baseado em interlíngua.

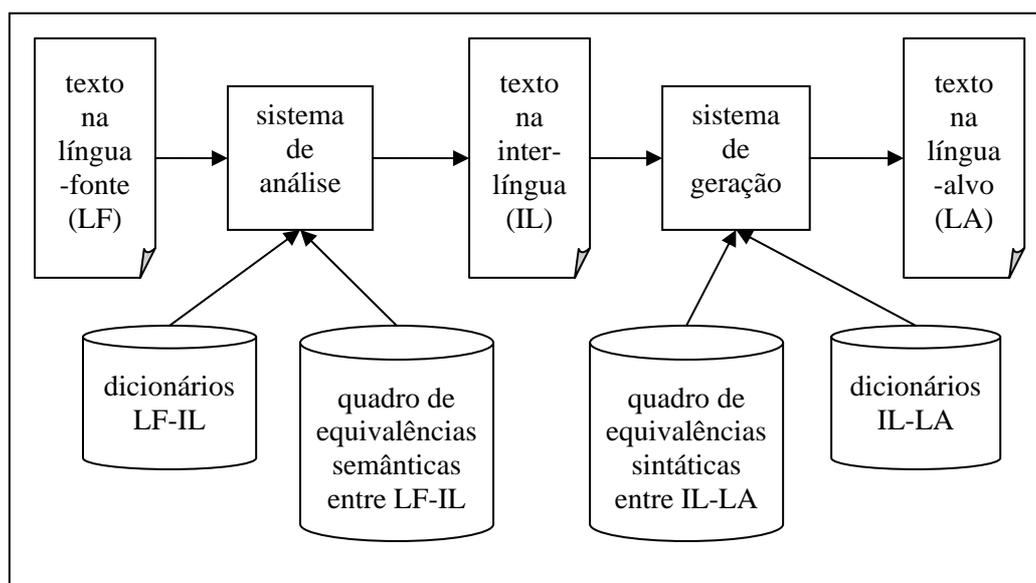


Figura 3. Arquitetura geral de um sistema de tradução por interlíngua

Para que pudesse integrar sistemas multilíngües, essa interlíngua - também chamada língua-pivô - deveria ser genérica (e plástica) o suficiente para acomodar diferenças, não apenas de conteúdo, mas de forma de representação, que seriam privativas de cada uma das línguas envolvidas. Reside neste ponto a grande crítica à abordagem interlingual: em última análise, ela proporia a formalização de um sistema adâmico, pré-babélico, que pudesse conter todas as línguas existentes, das quais constituiria uma parte universal (tal como uma gramática subjacente a todas e a cada uma das línguas naturais). Seria, portanto, a língua perfeita, que compreenderia todas as demais (ou que estaria compreendida em todas elas). Os críticos da abordagem interlingual querem crer que esta é uma perspectiva exageradamente ingênua, na medida em que a existência de princípios universais (de uma gramática universal, enfim) é controversa, e envolve um estágio de conhecimento da estrutura lingüística a que não se teve ainda acesso.

Do ponto de vista teórico, a abordagem interlingual seria mais adequada para sistemas multilíngües, na medida em que a efetividade da modularização permitiria o desenvolvimento de sistemas independentes de análise e geração e, principalmente, reduziria o custo de incorporação de novas línguas ao sistema. A abordagem interlingual prevê a necessidade de um número de sistemas equivalente apenas ao dobro do número de línguas envolvidas (2n), dado que seriam necessários apenas os módulos da tradução da língua-fonte para a interlíngua, e da interlíngua para a língua-alvo. Nas

abordagens de transferência, para a produção dos mesmos resultados seriam necessários $n(n-1)$ sistemas, em que "n" responde pelo número de línguas envolvidas.

A par da vantagem operacional, a abordagem interlingual admite que os desenvolvedores tenham apenas o conhecimento de sua própria língua materna e da interlíngua proposta. Como a interlíngua, por seu caráter universal, conservaria várias das propriedades da língua do desenvolvedor, e dado que seria necessariamente não-ambígua, é de esperar que as relações entre língua natural e interlíngua sejam mais simples do que as existentes entre duas línguas naturais, cada uma das quais organizadas a partir de seus próprios princípios de imprecisão e vagueza. Por fim, e como vantagem adicional, a abordagem interlingual permitiria ainda a geração, de volta para a língua-fonte, do texto já projetado para a língua intermediária. Seria esta uma excelente estratégia de validação do processo de representação e verificação dos resultados. Apenas os sistemas de transferência de natureza bidirecional, e ainda assim de forma bastante imprecisa (porque os resultados estariam inevitavelmente contaminados pela imprecisão da língua-alvo), poderiam realizar o mesmo movimento.

No entanto, e apesar de todas essas vantagens, a tradução baseada em interlíngua vem sendo preterida pela abordagem por transferência, mesmo em sistemas multilíngües. Não apenas porque não se pôde ainda chegar a uma interlíngua que contivesse, efetivamente, princípios mais gerais (universais), participantes de todas as outras línguas, ou de pelo menos um subconjunto expressivo de línguas (como as línguas neolatinas, por exemplo), mas pela complexidade do processo de projeção da língua-fonte para a interlíngua. Apesar de teoricamente menos vantajosa, a abordagem por transferência tem provado que o desenvolvimento de interfaces específicas entre a língua-fonte e a língua-alvo, embora exija formação bilíngüe por parte do desenvolvedor, é menos complexo (e conseqüente menos oneroso, mais rápido e mais factível) do que os módulos de projeção para a interlíngua, envolvidos com sistemas de representação de natureza muito abstrata. Da mesma forma, tem sido observado que esses mesmos módulos de transferência podem ser otimizados, de forma a serem reaproveitados, em alguma medida, por novas línguas a serem incorporadas ao sistema.

4.3 Um estudo de caso

Como sempre, talvez a virtude esteja no meio-termo. O mais celebrado entre todos os sistemas em operação, o Météo, cuja precisão alcançaria 90% dos resultados, mistura estratégias da abordagem direta com abordagem indireta (por transferência sintática). O hibridismo do sistema denuncia que o isolamento das diferentes abordagens de tradução automática presta-se muito mais a finalidades pedagógicas (como a deste capítulo) do que ao desenvolvimento efetivo de sistemas funcionais. A coabitação entre essas diferentes metodologias é não apenas possível, mas talvez imprescindível para que os resultados operacionais possam ser aprimorados.

O sistema Météo, cuja primeira versão foi desenvolvida, em 1976, pelo grupo TAUM (Traduction Automatique de l'Université de Montréal), traduz boletins meteorológicos do inglês para o francês, e está ainda hoje em operação. É um sistema de tradução bilíngüe unidirecional (do inglês para o francês) que consiste de três dicionários bilíngües característicos da abordagem direta: um para expressões idiomáticas, um para topônimos (nomes de lugares) e um terceiro para o vocabulário

meteorológico de maneira geral. Além dos dicionários, o sistema possui três módulos de processamento (característicos, por sua vez, da abordagem indireta por transferência): um para a análise sintática do inglês, um para a geração sintática do francês, e um para a geração morfológica do francês. Não há processamento morfológico do inglês: todas as variantes são representadas no dicionário.

Embora preveja a análise sintática do inglês, o sistema a reserva para algumas construções específicas, e opera, na maior parte dos casos, apenas com os dicionários, como uma abordagem direta, em função da identidade estrutural entre as sentenças de entrada e as sentenças de saída. As sentenças de entrada, sempre muito curtas, caracterizam uma espécie de sublíngua do inglês, bastante rígida e controlada, de estilo telegráfico, onde não aparecem pronomes, sentenças relativas e a voz passiva. Essas restrições de redação permitem que a substituição lexical preceda a análise sintática e dela seja independente. A estratégia de tradução indireta é ativada, pois, se e apenas se a estratégia de tradução direta não se revela suficiente, o que é verificado por critérios de validação internos ao próprio sistema.

Do ponto de vista operacional, o sistema Météo integra um sistema maior (neste sentido, é considerado um sistema "embutido"), situado junto ao Centro Meteorológico Canadense. Este macro-sistema recebe, por meio de uma rede de comunicação, os boletins meteorológicos redigidos em inglês (por um humano); pré-processa os dados, formatando-os para que possam servir de entrada ao sistema Météo; executa a tradução automática; encaminha, para tradutores humanos, o material para o qual o sistema não pôde prover tradução (por causa de problemas de digitação no texto original, por exemplo); reformata os dados de saída; e, por fim, transmite a versão final do texto, em francês, pela rede.

Para muitos estudiosos, o Météo, apesar de sua excessiva especialização, e das já referidas tentativas malsucedidas de extensão de seu escopo de atuação para além do domínio dos boletins meteorológicos, constitui um exemplo concreto da viabilidade da tradução automática. Antes da implantação do sistema no Centro Meteorológico Canadense, a tarefa de tradução dos boletins meteorológicos era considerada tão repetitiva e desmotivante que poucos eram os tradutores humanos que permaneciam por muito tempo na função. O sistema veio, portanto, a automatizar o que seria, em princípio, automatizável, derivando desta decisão todo o seu sucesso. Principalmente: na consecução de seu objetivo maior, os desenvolvedores do sistema não hesitaram em mesclar estratégias que, embora contraditórias, revelaram-se complementares no tratamento do problema.

4.4 A tradução automática para língua portuguesa

A tradução automática para a língua portuguesa está, de maneira geral, ainda na pré-história, principalmente porque os sistemas comercializados existentes - de resultados bastante precários e de validade bastante controvertida - não produziram ainda documentação pública que permitisse a consideração dos detalhes de implementação e a análise das estratégias utilizadas. O quadro é agravado pelo fato de que não existe, na comunidade universitária brasileira, nenhum sistema em estágio operacional, embora

existam algumas iniciativas neste sentido - o Projeto UNL-Brazil², por exemplo, desenvolvido junto ao NILC, de onde retiramos a experiência para a produção deste texto.

Os muitos sistemas comerciais, produzidos principalmente fora do Brasil, muitos deles à venda nas prateleiras de supermercados e revendas de *software*, pecam, de maneira geral, por tomarem apenas o inglês como língua-fonte (ou como língua-alvo), pela dimensão (geralmente diminuta) do léxico, pela utilização de apenas recursos lingüísticos (em detrimento de bases de conhecimento e *corpora*, por exemplo) e pelo privilegiamento da abordagem direta (tradução palavra-a-palavra). Os poucos sistemas baseados em transferência (sempre sintática) parecem fazer uso de gramáticas excessivamente simplificadas, incapazes de lidar com construções mais complexas (como coordenações e subordinações).

É o que se depreende, por exemplo, da avaliação dos resultados dos sistemas Translator Pro, Alta Vista (na verdade, SYSTRAN), Intertran, GO Translator, Tradunet e Enterprise Translator Server, reportados por Oliveira et al. (2000). Todas estas ferramentas foram testadas em relação à tradução, tanto do inglês para o português quanto do português para o inglês, de 20 passagens de textos jornalísticos, envolvendo uma ou mais sentenças, extraídas do primeiro caderno de dois jornais diários, a Folha de S. Paulo e o *The New York Times*. Os resultados podem ser observados nos quadros 1 e 2 abaixo, que apresentam o desempenho das ferramentas para as sentenças do inglês que envolveram, respectivamente, o menor e o maior número de problemas.

Tabela 1. Resultados de tradução, para o português, da sentença: “The child is at the center of an international fight between his father in Cuba, Juan Miguel González, and relatives in Miami.” (Extraído de Oliveira et al. 2000)

Ferramentas	Resultados
TRANSLATOR PRO	A criança está no centro de uma briga internacional entre o pai dele em Cuba, Juan Miguel González, e parentes em Miami.
Alta Vista	A criança está no centro de uma luta internacional entre seu pai em Cuba, em Juan Miguel González, e em parentes em Miami.
Intertran	A criança é no centro de um internacional batalhar entre dele pai em Cuba, Juan [Miguel] [González], e familiares em Miami.
TRADUNET	A criança está no centro de uma briga internacional entre seu pai em Cuba, Juan Miguel González, e parentes em Miami.
Enterprise Translator Server	A criança está no centro de uma luta de internacional entre seu pai em Cuba, González de Miguel de Juan, e parentes em Miami.

² O Projeto UNL tem por objetivo o desenvolvimento de um sistema multilíngüe de tradução automática baseada numa interlíngua de natureza semântica - a UNL (ou *Universal Networking Language*) - desenvolvida por pesquisadores vinculados à Universidade das Nações Unidas, órgão da ONU, em Tóquio. A UNL pretende instrumentalizar o processo de representação dos conteúdos veiculados pelas sentenças das línguas naturais, concebendo-os como uma estrutura do tipo rede, em que conceitos exercem a posição de nós, e as relações (semânticas) entre conceitos, a posição de arcos entre nós. A fim de representar as circunstâncias em que se realiza a enunciação, esses nós podem ser anotados por atributos que denotam informação de natural contextual. Maiores informações sobre a linguagem UNL podem ser obtidas em Uchida et al (1999). Detalhes sobre o Projeto UNL-Brazil podem ser encontrados em Martins et al (1999) e Nunes et al (2001).

Ferramentas	Resultados
TRANSLATOR PRO	O Irlanda paz processo Do norte foi mergulhado mais cedo em crise este mês quando o IRA arrancou de conversas com o corpo de desarmamento independente da província britânica e retirou uma proposta para pôr seus braços `` além de uso.’’
Alta Vista	O processo da paz de Irlanda do Norte foi mergulhado na crise mais cedo este mês em que o IRA puxou fora das conversas com o corpo independente do disarmament da província britânica e retirou uma proposta para pôr seu uso do ``beyond dos braços.’’
Intertran	O Irlanda do Norte processo de paz era mergulho para a crise anteriormente este mês quando o IRA puxado ausente de conversas com o Britânico competência independente desarmamento corpo e retirado um proposta colocar o seu armas [``beyond] use.’’
TRADUNET	A paz do Northern Ireland processa Esteve mergulhado Em crise Mais cedo Este mês quando o IRA puxado sem conversas com o desarmamento independente da província Britânica corporifica E withdrew uma proposta pôr seus braços `` além de uso.’’
Enterprise Translation Server	O processo do norte de paz de Irlanda foi mergulhado em crise mais cedo este mês quando o IRA arrancaram de conversas com a província Britânica corpo independente de desarmamento e retirou uma proposta a por seu uso de ççbeyond de braços.’’

Da análise dos dados reportados acima, pode-se perceber o quanto falta ainda às ferramentas brasileiras para que resultados de qualidade possam ser obtidos. A maior parte delas parece operar de forma razoável para sentenças (curtas) de ordem direta, com vocabulário comum e sem grandes acidentes sintáticos; mas esta aparente eficiência é antes uma sombra que nos desvia da verdadeira razão do sucesso: a proximidade entre as construções do inglês e do português, de que a comparação da sentença de entrada e as sentenças de saída da Tabela 1 pode funcionar como prova. Para sentenças (ou mesmo fragmentos de sentença) em que esta semelhança é menor - como indicado na Tabela 2 - o resultado é visivelmente desalentador. E os problemas se espalham nas mais variadas direções: há falta de correspondências nos dicionários bilíngües que servem às ferramentas; há o clássico problema do tratamento das formas homônimas e das expressões idiomáticas; há a generalização, para a língua-alvo, de construções figuradas de validade restrita à língua-fonte; há o problema da ordem dos itens lexicais na sentença (principalmente no interior dos sintagmas nominais); há os problemas sintáticos relativos à concordância e à regência; a recuperação de relações anafóricas e o preenchimento de elipses; etc. O grande desafio da tradução automática para a língua portuguesa continua sendo, portanto, operar, com razoável eficácia, não apenas quando o português replica a estrutura do inglês, mas também nas situações - muito mais comuns, é preciso dizer - em que as duas línguas se distanciam.

5. Conclusões

Por tudo que se disse neste capítulo, pode-se perceber o quanto a tradução automática constitui hoje, no início do século XXI, um domínio de encruzilhada. Nenhuma outra aplicação talvez represente de forma tão cabal os desafios da interação entre modelos matemáticos, lingüísticos e cognitivos. Nenhuma outra aplicação talvez evidencie de forma tão plena a caverna em que estamos metidos.

Na medida em que mobiliza competências e habilidades lingüísticas de natureza antes geral - cujos recortes metodológicos e reducionismo têm significado, não a viabilização do processo, mas a má qualidade dos resultados - a tradução parece requerer, como condição *sine qua non*, a formalização de um modelo global da linguagem e do conhecimento, de que a Lingüística e a Inteligência Artificial estão ainda muito distantes.

A combinação de estratégias de formalização e representação do conhecimento lingüístico e do conhecimento de mundo representa, sem sombra de dúvida, o grande desafio atual do domínio da tradução automática. Existe mesmo quem proponha que a tradução automática deva aguardar o desenvolvimento desses recursos lingüísticos e conceituais antes de enveredar pela proposição de sistemas, que significariam, no atual estágio, apenas desperdício de tempo, energia e dinheiro. Ou seja, em função da precariedade de informações a respeito da estrutura e do funcionamento da linguagem e do cérebro humano, não restaria alternativa senão investir, primeiramente, no desenvolvimento das ciências cognitivas (entre elas a Lingüística, a Psicologia e a Neurologia) para, e apenas então, chegar-se a um modelo minimamente capaz de emular a tradução humana.

No entanto, e na corrente oposta, existem aqueles que, como bem salienta Martin Kay, sabem que "aviões não batem asas". Acreditam que, assim como a emulação de um aspecto específico do comportamento das aves dispensou a imitação da fisiologia do vôo e pôde ser alcançada por meio de estratégias alternativas, estaríamos também dispensados da obrigatoriedade de modelos diretamente baseados no comportamento humano para a simulação da habilidade de tradução. Afirma-se mesmo o contrário: o desenvolvimento de estratégias automáticas de tradução (embora imperfeitas e restritas) pode permitir que se amplie o conhecimento da habilidade (ou dos requisitos) do homem para a tradução. Longe de conduzir ao imobilismo, pois, o reconhecimento das limitações da tradução automática tem alimentado os desenvolvedores, certos de que apenas a experimentação poderá produzir, neste caso, o conhecimento necessário para a implementação (futura) de sistemas bem-sucedidos. O desenvolvimento de sistemas de tradução automática revela-se, pois, lugar extraordinário para a experimentação de teorias lingüísticas e cognitivas, e para o acúmulo processual e contínuo de conhecimento sobre a linguagem, o pensamento e seu funcionamento.

6. Leituras Complementares

A bibliografia sobre tradução automática em português é praticamente inexistente. Salientem-se os textos de Maria Helena Mira Mateus (1995), sobre a história da tradução automática, e de Pedro Santos (1995), sobre alguns problemas do campo. Ambos integram as atas de um curso de Engenharia da Linguagem, oferecido, em 1993, pelo Instituto de Lingüística Teórica e Computacional (ILTEC), de Portugal. As atas foram reunidas em um livro, organizado por Maria Helena Mateus e António Horta Branco, chamado "Engenharia da Linguagem", que contém também outros textos sobre o processamento automático das línguas naturais.

Em inglês, destacam-se, entre os textos introdutórios, principalmente Hutchins e Somers (1992, segunda edição: 1997). O texto apresenta, de forma bastante detalhada, não apenas os princípios e as estratégias do domínio, mas a análise de alguns entre os

sistemas mais famosos no início dos anos 1990. Os detalhes do sistema Météo, aqui apresentado, foram retirados dali. Discutem-se também estratégias de avaliação de sistemas de MT.

Dois outros textos interessantes foram editados por Serguei Nirenburg: *Machine Translation* (de 1987) e *Progress in Machine Translation* (de 1993). Trazem discussão de métodos, descrição de projetos, de sistemas, de ferramentas, e analisam também políticas públicas de investimento na área.

Discussões mais recentes podem ser encontradas na principal revista científica do campo, o *Journal of Machine Translation*, com quatro volumes por ano, editada pela Kluwer Academic Publishers.

Referências

- Arrojo, R. (2000). *Oficina de tradução; a teoria na prática*, Editora Ática, São Paulo.
- Boitet, C. (1995). "(Human-Aided) Machine Translation: A Better Future?" In: *Survey of the State of the Art in Human Language Technology*, edited by Cole, R.A.; Mariani, J.; Uszkoreit, H.; Zaenen, A.; Zue, V., NSF/CEC/CSLU. Oregon Graduate Institute. November. (<http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>)
- Boitet, C. (1995). "Machine-aided Human Translation", In: *Survey of the State of the Art in Human Language Technology*, Edited by Cole, R.A.; Mariani, J.; Uszkoreit, H.; Zaenen, A.; Zue, V., <http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>, November.
- Catford, J. C. (1980) *Uma teoria lingüística da tradução*, Cultrix, São Paulo.
- Hutchins, W. J. and Somers, H. L. (1992) *An introduction to Machine Translation*, Academic Press, San Diego (CA).
- Kay, M. (1995). "Machine Translation: The disappointing past and present", In: *Survey of the State of the Art in Human Language Technology*, Edited by Cole, R.A.; Mariani, J.; Uszkoreit, H.; Zaenen, A.; Zue, V., <http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>, November.
- Martins, R. T.; Rino, L. H. M.; Nunes, M. G. V.; Montilha, G.; Oliveira Jr., O. N. (2000). "An interlingua aiming at communication on the web: how language-independent can it be?", In: *Proceedings of the Workshop on Applied Interlinguas: Practical Applications of Interlingual Approaches to NLP*, pp. 24-33.
- Nida, E. (1975). *Language structure and translation*. Stanford University Press, California.
- Nirenburg, S. (Ed). (1987). *Machine Translation - Theoretical and Methodological Issues*, Cambridge University Press, Cambridge.
- Nirenburg, S. (Ed). (1993). *Progress in Machine Translation*, IOS Press, Amsterdam.
- Nunes, M. G. V.; Martins, R. T.; RINO, L. H. M.; Oliveira Jr. (2001). "The use of the Universal Networking Language for devising an automatic sentence generator for Brazilian Portuguese", In: *Cadernos de Computação*, Volume 2, Number 2, October.
- Oliveira Jr., O. N.; Marchi, A. R., Martins, M. S.; Martins, R. T. (2000). "A Critical Analysis of the Performance of English-Portuguese-English MT Systems", In *Anais*

do V Encontro para o processamento computacional da língua portuguesa escrita e falada.

Platão. (s.d.) A República, tradução de Leonal Vallandro, Edições de Ouro, Rio de Janeiro.

Santos, P. (1995). "Tradução Automática", In: Engenharia da Linguagem, Organizado por Maria Helena Mateus e António Horta Branco, Edições Colibri, Lisboa.

Slocum, J. (1985) "A Survey of Machine Translation: Its History, Current Status and Future Prospects", In: Machine Translation Systems, organized by J. Slocum, Cambridge University Press, Cambridge.

Uchida, H.; Zhu, M.; Della Senta, T. (1999). A gift for a millenium, IAS/UNU, Tokyo.-

Yebra, V. G. (1982). Teoría Y práctica de la traducción, Gredos, Madrid.