participação de professores do Instituto de Ciências Matemáticas (ICMSC) e de Física (IFSC), da USP de São Carlos, do Departamento de Informática da UFSCar e da Faculdade de Letras da Unesp de Araraquara. Esse desenvolvimento está sendo coordenado pela Profa. Maria das Graças Volpe Nunes (coordenadora), do ICMSC, e pela Profa. Lúcia Helena Machado Rino (vice-coordenadora), do Departamento de Informática da UFSCar.

A escolha do NILC para produzir as ferramentas de software no projeto UNL-BR se deveu principalmente à experiência de seus pesquisadores no desenvolvimento de trabalhos tanto acadêmicos como aplicados, que inclusive tem permitido transferência de tecnologia das universidades para o setor produtivo. A título de ilustração, um dos projetos desenvolvidos pelo NILC foi um corretor gramatical para o português que está sendo comercializado pela Itautec-Philco S.A. O projeto UNL-BR foi lançado oficialmente no dia 29 de abril de 1997, na cidade de Curitiba, com a presença de diversas autoridades brasileiras e ligadas à Universidade das Nações Unidas, inclusive com uma palestra do megaempresário japonês, Sr. Nishi, que está financiando parte do projeto UNL. Essa iniciativa da Universidade das Nações Unidas é mais um projeto de âmbito internacional que conta com participação de pesquisadores da USP e UFSCar, de São Carlos.

Texto preparado pelo Prof. Osvaldo N. Oliveira Jr.
Professor do Instituto de Física de São Carlos, USP
Pesquisador Associado do NILC - publicado no jornal Primeira Página de São Carlos

resolveu patrocinar um projeto de longa duração - 10 anos - para o desenvolvimento de ferramentas de software visando a vencer a barreira da língua. Neste projeto, ao invés de tradução de uma língua para outra, faz-se a codificação do conteúdo de um texto de uma dada língua em uma linguagem universal, chamada *Universal Networking Language* (UNL), criada especificamente para textos escritos no ambiente da Internet. O texto já codificado em UNL pode então ser decodificado para a língua destino. Na terminologia adotada, codificação e decodificação, evita-se utilizar o termo "tradução automática", por ser esta uma meta difícil de ser atingida num futuro próximo, pelos menos no que se refere a aproximar da tradução humana de qualidade. O projeto UNL foi iniciado em 1997, e é de âmbito mundial, sendo coordenado pelo *Institute for Advanced Studies*, da Universidade das Nações Unidas, dirigido atualmente por um brasileiro, Dr. Tarcísio Della Senta. Nos três primeiros anos deverão ser desenvolvidos codificadores e decodificadores para as 13 línguas mais faladas da Terra, incluindo chinês, russo, alemão, francês, italiano, japonês, inglês, espanhol e português. Já foram assinados contratos entre a UNU e instituições de pesquisa de diversos países. Nos 7 anos restantes previstos para o projeto UNL, espera-se atingir as línguas de todos os países membros da Organização das Nações Unidas.

O que distingue o projeto UNL é a tentativa de conjugar esforços de grupos de pesquisa de vários países. Ou seja, ao invés de patrocinar o desenvolvimento de software para tradução entre conjuntos de duas línguas, a UNU pretende utilizar uma linguagem universal para que o trabalho realizado para codificação e decodificação do francês, por exemplo, possa contribuir para o projeto do italiano, e vice-versa. Para que haja essa colaboração, serão realizados durante o transcorrer do projeto congressos específicos com participantes dos diversos países, ocasiões nas quais serão debatidos avanços e dificuldades no desenvolvimento das ferramentas de software. A linguagem universal, UNL, concebida por pesquisadores japoneses deverá ser estendida a partir da contribuição de especialistas do mundo todo.

O Brasil ficou encarregado de desenvolver o codificador e decodificador para a língua portuguesa. Assim, textos em português poderão ser codificados em UNL para que possam ser decodificados para outras línguas, dependendo da preferência do usuário. Da mesma forma, o decodificador para o português servirá para que tenhamos acesso a material escrito em outras línguas que foi codificado em UNL. O projeto UNL-BR é coordenado por Tadao Takahashi, que até recentemente era o presidente da Rede Nacional de Pesquisas (RNP), responsável pela implantação do sistema de rede de computadores e da Internet no Brasil. A administração do projeto é feita pelo Centro Internacional de Tecnologia de Software (CITS), de Curitiba, Paraná. No primeiro ano de projeto, o desenvolvimento do codificador e decodificador para português ficou a cargo do Núcleo Interinstitucional de Lingüística Computacional (NILC), sediado em São Carlos, que conta com a

Vencendo a barreira da língua na Internet

O grande volume de informações disponíveis na Internet continua restrito a um número muito pequeno de usuários, não só no Brasil como em muitos outros países cuja língua principal não é o inglês. Ocorre que grande parte das informações aparece em inglês, ou porque foram geradas em países de língua inglesa, como os Estados Unidos e Grã-Bretanha, ou então porque os usuários que as disponibilizam querem atingir um público mais amplo. O inglês é então escolhido como a língua internacional. A importância dessa restrição talvez ainda não seja muito sentida entre nós, pois apenas uma pequena porcentagem da população tem acesso à Internet. E é a parcela da população que, em geral, tem nível mais alto de escolaridade, com maior probabilidade de conhecer, pelo menos instrumentalmente, a língua inglesa.

Mas a tendência, como com qualquer outra tecnologia, é a de que a Internet se popularize rapidamente, atingindo várias camadas da população, como hoje acontece com televisores a cores, videocassetes e que já se prevê acontecerá com a TV a cabo. Ainda não se sabe se a tecnologia que permitirá à Internet atingir a maioria da população será a mesma empregada atualmente, ou seja, através do acesso por microcomputador. Uma alternativa bastante provável é a de que possa ser empregada a tecnologia conhecida nos Estados Unidos como *WEBTV*. Trata-se de um uso interativo da TV, em que a interação se dá através de um aparelho semelhante ao atual controle remoto. De qualquer forma, a partir do momento que grande parte da população tiver acesso à Internet, como garantir que ela possa desfrutar de todo o potencial da rede? Acredito ser consenso que é impossível ensinar inglês para toda essa massa de gente. Ou seja, um mecanismo de tradução terá que ser utilizado para que as pessoas possam receber - e transmitir - informações em sua própria língua.

Devido ao volume de informações, a tradução humana pode ser descartada. Não só pelo custo altíssimo que teria, mas também porque a velocidade com que se realiza a tradução impediria que a informação fluisse com a rapidez necessária. Ora, se a grande vantagem da Internet é o encurtamento de distâncias, já que qualquer evento ou fato importante pode ser divulgado praticamente instantaneamente, não faz sentido ter que esperar horas ou dias para que a tradução seja realizada. A alternativa seria então uma tradução automática, realizada por um software de computador. Já existem no mercado alguns softwares que realizam essa tarefa, para várias línguas. Para o português em particular, são mais comuns os softwares que traduzem textos do inglês para o português, mas a possibilidade de tradução no sentido inverso dificilmente é oferecida.

As dificuldades mencionadas são comuns à maioria esmagadora da população mundial, uma vez que nenos de 10% são falantes da língua inglesa. Sensível a essa enorme barreira para a comunicação entre os povos, a Universidade das Nações Unidas (UNU), sediada em Tóquio,