

Acessibilidade em Interfaces Computacionais da Web: Aspectos Lingüísticos

Ismael Mattos Andrade Ávila

Engenheiro do Centro de Pesquisas em Telecomunicações – CPqD

avila_an@cpqd.com.br

Monografia de Fatores Humanos em Interfaces Computacionais

Orientação: Profa. Dra. Maria Cecília Calani Baranauskas

Data: agosto de 2001

Local: Instituto de Computação da Universidade Estadual de Campinas, UNICAMP

Resumo: Este artigo estuda o problema da barreira lingüística como um fator de limitação no acesso da Web. As razões e características desse tipo de *handicap* serão descritas e quatro modalidades de solução serão estudadas em profundidade a fim de alcançar um bom entendimento sobre os potenciais e desvantagens de cada estratégia. Levando em consideração critérios tais como direitos humanos, granularidade, disponibilidade, e robustez das soluções estudadas temos a intenção de estabelecer seus limites e avaliar as sinergias latentes daquelas ferramentas.

Abstract: This paper examines the problem of the language barrier as a factor of limitation in the access of the Web. The reasons and characteristics of this kind of handicap will be described and four modalities of solution will be studied in depth in order to achieve a good understanding about the potentials and drawbacks of each strategy. Taking in consideration criteria such as human rights, granularity, availability, and robustness of the studied solutions we intend to establish their boundaries and evaluate the latent synergies of those tools.

Keywords: Acessibilidade, barreiras lingüísticas, deficiência lingüística, discriminação na educação, direitos humanos, tradução automática, adaptação cultural de interfaces, ensino colaborativo, semiótica, línguas internacionais auxiliares.

1. Introdução

Inicialmente, para compreender o problema da *deficiência lingüística* e seus efeitos na limitação da acessibilidade das interfaces computacionais da Web, nos referenciaremos à definição de deficiência lingüística em Piron [1]: como a “impossibilidade de exprimir-se com destreza e confiança numa situação intercultural”. Sendo a definição original genérica, ela recobre todos os campos da comunicação humana: linguagem oral e escrita, e em contextos tão diversos quanto os das assembleias internacionais, da comunicação dos refugiados, turistas e viajantes, daquela no âmbito das grandes organizações diplomáticas, e das corporações multinacionais. A observação dos diversos meios em que a comunicação intercultural é necessária demonstra a prevalência das situações de *deficiência lingüística*, em gradações que vão do simplesmente incômodo até as piores formas de discriminação e sofrimento [1,2].

Por sua vez, a transposição do conceito de *deficiência lingüística* para o âmbito da Web trará à luz outra modalidade de limitação, igualmente caracterizada pela falta de domínio de línguas por parte de um indivíduo, mas cuja consequência específica é a impossibilidade do interessado em fazer uso das informações disseminadas na Web. Em virtude dessa falta de domínio dos idiomas estrangeiros nos quais é veiculada a quase totalidade das informações contidas na rede mundial de computadores, sobretudo o inglês, o interessado se vê totalmente cerceado em seu acesso à informação, ou tem esse acesso grandemente dificultado.

1.1 Contexto

Na Rede mundial de computadores, a noção da existência das barreiras lingüísticas é muito mais nova do que a própria tecnologia de comunicação via Internet

(1969) ou o advento da *World Wide Web* (1990). Em ambos os casos, a popularização dessas novas mídias se deu em um ambiente unilíngüe (predominantemente), na medida em que quase todos os usuários iniciais da Internet se encontravam nos Estados Unidos e tinham o inglês com primeira ou segunda língua. Esse fato por si bastou para que fosse inexistente qualquer preocupação com a diversidade lingüística. Imaginava-se, em um primeiro momento, que a veiculação de informações na Rede pudesse, por definição, estar restrita à língua inglesa. Mas já em 1996, Tim Berners-Lee, que no final da década de oitenta concebera, no CERN (Centro Europeu de Pesquisas Nucleares, Genebra), a *World Wide Web*, se perguntava, em um artigo intitulado *The World Wide Web: Past, Present and Future* [3], se os efeitos do enfraquecimento do fator geográfico como protetor da diversidade cultural levariam a uma “cultura monolítica (Americana)” ou se “fomentariam grupos de interesse ainda mais distintos daqueles então existentes”. Tudo parece indicar que é a segunda hipótese que prevalece, na medida em que a concepção unidimensional da Web (i.e. monocultural) vem sendo abandonada e, com a rapidez que caracteriza a evolução da Rede, substituída por outra, mais aberta, que reconhece a importância de a expansão da Web respeitar a multiplicidade de culturas no planeta. Isso ocorre porque, embora seja notável a difusão do inglês, as cifras sobre número de falantes das diversas línguas no mundo dão conta de que aquele só é usado por cerca de 10% da população do planeta, 6% se somente considerados os falantes nativos [4]. Isso equivale a dizer que uma Web de abrangência planetária cujo conteúdo esteja totalmente em inglês tem por efeito restringir o acesso da maior parcela de seus potenciais usuários. Teríamos então, em termos de acessibilidade, a barreira

lingüística como fator limitador para 90% dos indivíduos, aqui chamados *deficientes lingüísticos*, o que configura uma forma de *handicap* muito mais prevalente do que aquelas produzidas pelas deficiências de ordem sensorial ou psicomotora.

No início da Internet, quando as condições econômicas e de infra-estrutura favoreciam um maior crescimento da rede na América do Norte, e um conseqüente predomínio de usuários e conteúdos de língua inglesa, era inexistente a percepção da deficiência lingüística e de seus efeitos sobre a acessibilidade. Entretanto, verifica-se atualmente uma mudança no ritmo de expansão da Rede, e o crescimento relativo da Web naquela parte do mundo é inferior ao das demais regiões: estima-se que o total de usuários de língua inglesa cresça 60% nos próximos 6 anos, enquanto que no mesmo período o número de usuários das demais línguas deve aumentar 150% [7]. Haverá então, com a tendência de aceleração do crescimento da Internet na Ásia, América Latina e sul e leste europeus, uma comunidade internauta cada vez mais multilíngüe. E a conseqüência social desse fenômeno já é antecipada: o surgimento de uma Web multicultural [9]. Mas com isso os efeitos perversos da limitação lingüística passarão a ser sentidos cada vez com maior intensidade. Em outras palavras, vencidas as barreiras de tempo e distância, a barreira lingüística se tornará o maior obstáculo à circulação da informação.

A compreensão desse processo vem despertando diversas iniciativas que objetivam superar a barreira lingüística, ou, em alguns casos, explorar comercialmente esse novo e promissor filão. Prover aos portadores de limitação lingüística acessibilidade às informações da Web, ou tornar esse público acessível às empresas pontocom passou a ser uma atividade economicamente interessante. Multiplicam-se empresas de consultoria tecnológica, estratégica e de mercado, desenvolvedores de ferramentas de tradução automática e prestadores de serviços lingüísticos para a Web. Contudo, essas iniciativas tendem a se limitar a aspectos comerciais, tais como o potencial de consumo do público alvo e os custos envolvidos no atendimento de seu perfil lingüístico e cultural (*localization*). Mas elas não preenchem todas as lacunas: o direito à informação pode ser considerado parte importante da concepção de cidadania, devendo assim se estender a todos aqueles que potencialmente necessitariam ou desejariam fazer uso de alguma informação na Web, independentemente de seu perfil de consumo.

As implicações da *diversidade lingüística* nas questões relacionadas aos direitos humanos têm sido tratadas por todas as declarações e convenções internacionais desde a Declaração Universal dos Direitos Humanos (1948), passando pela Convenção Internacional do Trabalho (1957, 1989), pela Convenção da UNESCO Contra a Discriminação na Educação (1960), pelo Pacto Internacional sobre Direitos Civis e Políticos (1966), a Declaração dos Direitos Humanos dos Estrangeiros (1985), a Convenção sobre os Direitos da Criança (1989), a Convenção Internacional sobre a Proteção dos Direitos dos Trabalhadores Migrantes (1990), a Declaração dos

Direitos das Pessoas Pertencentes a Minorias Étnicas, Religiosas e Lingüísticas (1992) e pela Declaração Provisória sobre os Direitos dos Povos Indígenas (1993) [5]. Os preceitos estabelecidos nessas Cartas dão um arcabouço legal à questão dos direitos lingüísticos. Em razão disso e para os objetivos do presente trabalho é fundamental tomá-los como premissa, estabelecendo assim os critérios de estudo do problema da *deficiência lingüística* na comunicação intercultural, especificamente no contexto da Web, comparando as possíveis soluções.

Na prática, já existem situações em que, por exemplo, considerações sobre discriminação na educação devem ser feitas. Se tomarmos os termos da Convenção da UNESCO acima mencionada, veremos que já em seu Artigo 1º é estabelecido o conceito de *Discriminação* como sendo “qualquer distinção, exclusão, *limitação* ou preferência que, sendo baseada em raça, cor, sexo, *língua*, religião, [...] tem o propósito de negar ou prejudicar a igualdade ou o tratamento na educação”. Se, em muitos países, mesmo do Terceiro Mundo, a Internet já está sendo introduzida na rede pública de ensino (o Brasil planeja conectar mais de sete milhões de alunos até 2002 [36]), é importante avaliar em que medida a deficiência lingüística dos alunos poderá representar um fator de limitação de seu real aproveitamento.

A percepção de que a acessibilidade da Web possa ser dificultada ou impedida pelas barreiras lingüísticas não é universal. Para alguns, o problema não seria de fato importante, uma vez que toda diferença de oportunidades é encarada com sendo inerente à condição humana. Por esse prisma, esperar que a Web se torne uma ferramenta (ou mídia) acessível a todos seria algo utópico, e a estrutura hierarquizada em que alguns têm pleno acesso enquanto outros ficam marginalizados seria então uma conseqüência natural do fato de que as sociedades mais avançadas, por conta de seu pioneirismo tecnológico e de seu dinamismo social, fizeram por merecer essa condição privilegiada. No que se refere aos indivíduos, as diferenças seriam conseqüência dos maiores ou menores esforços pessoais empreendidos na aquisição do instrumento (lingüístico) necessário à utilização da Web. Haveria portanto, segundo tal paradigma, tanto no âmbito social quanto individual, um tipo de meritocracia na qual os ganhos obtidos e a comodidade desfrutada seriam perfeitamente justificáveis pelas condições de partida.

Nossa abordagem, por sua vez, refuta o raciocínio acima, pelo simples fato de que, como nas outras modalidades de deficiência ou limitação a que um indivíduo esteja submetido, por razões naturais ou não, buscar soluções que atenuem de fato seus *handicaps* é eticamente mais defensável do que pura e simplesmente elaborar racionalizações que justifiquem sua condição. Pessoas com capacidade visual sub-ótima e aquelas com limitações motoras devem ter acesso a recursos (tais como conversores texto-fala ou fala-texto, resp.) que minimizem suas dificuldades no acesso às interfaces computacionais. Se é legítima, para essas modalidades de deficiência, a busca por meios que promovam a acessibilidade, compensando, tanto quanto possível, os

handicaps individuais, não seria menos legítimo dispensar aos indivíduos portadores de deficiência lingüística um mínimo de atenção e de consideração. Sobretudo se considerarmos que ao invés do que se imagina, o instrumento lingüístico predominante na Web, o idioma inglês, está de fato inacessível para a maior parte da população de nosso planeta, uma vez que seu aprendizado demanda investimento de esforço, tempo e dinheiro, que está fora do alcance do comum dos mortais [2]. Por conta disso, a condição de *deficiente lingüístico* não pode ser tratada como resultado de meras decisões individuais, pois o esforço exigido não é razoável.

Assim, se a solução do problema não tratar cada indivíduo de forma não discriminatória no que se refere ao direito à informação, ela terá por efeito criar *analfabetos mundiais* (ver Piron [2]) numa hierarquia em que uns poucos têm acesso à maior parte do conteúdo disponível, outros só alcançam pequenas parcelas, e a maioria fica de fato excluída.

Vemos então que a forma como encararmos o fator *limitação lingüística* no uso da Web (sobretudo na educação), fará toda a diferença. Um meio que seria potencialmente um fator de redução das defasagens, por disponibilizar acesso padronizado (do ponto de vista tecnológico) a um universo de informações, em qualquer local do mundo, pode acabar se tornando um instrumento de agravamento do desnível de desenvolvimento das nações ao privilegiar os usuários pertencentes às culturas dominantes.

É claro, a igualdade a que nos referimos é a de oportunidades, o que significa que o que se busca neste caso é uma visão mais ética e plural da Web. Essa visão levaria a um conjunto de posturas menos etnocêntricas e mais democráticas, sem que isso signifique que as diferentes comunidades culturais e lingüísticas tenham que abdicar do uso de suas respectivas línguas e da manifestação de sua identidade e de seus valores. Mas ela estimularia todos aqueles que fazem uso da Rede para fins mais abrangentes (em termos culturais e geográficos) conforme a vocação natural desta, a dar um passo em direção às demais culturas, aproximando-se tanto quanto possível de um padrão intercultural de comunicação.

1.2 Objetivo:

O objetivo do presente trabalho é tentar comparar as várias modalidades de solução para o problema acima descrito. Ao todo serão estudadas 4 soluções, contrastadas com um modelo de controle que corresponde à situação encontrada nos começo da Web (anos 90) em que todo o conteúdo estava disponível em inglês. Daremos a essa hipotética situação o nome de Solução *Cyber-english*. As outras modalidades de solução avaliadas serão:

- a) Ferramentas de tradução automática;
- b) Tradução mediada pela meta-língua UNL (*Universal Network Language*);
- c) Interfaces multilíngües, baseadas em conceitos de adequação local (*localization*) com tradução humana;
- d) Uso de língua internacional auxiliar.

Diante desse objetivo, o trabalho será dividido conforme segue:

1. Introdução
2. Referencial teórico
3. Modalidades de solução
4. Uso combinado das soluções
5. Conclusão
6. Propostas de estudo
7. Agradecimentos
8. Referências

2. Referencial teórico:

Para dar contornos mais nítidos à questão da multiplicidade lingüística e cultural no mundo, e estabelecer alguns pontos de balizamento para o presente trabalho, vamos nos referenciar às cifras reunidas em [4] e que se encontram sintetizadas na tabela 1. Nela temos a repartição percentual de todas as línguas vivas em nosso planeta e os números absolutos dos referidos percentuais. Vê-se que o total de idiomas ainda em uso no mundo ultrapassa a quantia de seis mil, sendo a Europa o continente com menor quantidade: 225.

	Línguas Vivas Total	Porcentagem
América	1000	15%
África	2011	30%
Europa	225	3%
Ásia	2165	32%
Pacífico	1302	19%
TOTAL	6703	

Tabela 1: distribuição das línguas

É preciso ressaltar que na referência supracitada *todas* as línguas consideradas ainda vivas são contabilizadas, independentemente do respectivo número de usuários. Assim, dentre as 6703 línguas está o mandarim, com cerca de 885 milhões de falantes, ao lado de inúmeros idiomas cuja base populacional é inferior a uma centena de indivíduos. Essa abordagem, ainda que adequada do ponto de vista etnológico, tem pouca aplicabilidade prática para nosso objetivo de identificar e compreender as necessidades de comunicação para o acesso à rede mundial de computadores. Primeiro porque grande parte dessas línguas minoritárias só existe na modalidade oral, e ainda que se possam imaginar usos da Web na sua preservação (quase todas são ameaçadas de extinção), as medidas para evitar a eliminação desses idiomas fogem do escopo do presente trabalho. Por outro lado, é verossímil que muitos dos usuários dessas línguas regularmente se sirvam de uma outra língua de maior abrangência, ou de uma *lingua franca*, em seus contatos com as culturas vizinhas, de maneira que a possibilidade de acessar também as informações na Web através de uma língua intermediária já seja por si só um ganho em termos de acessibilidade. Isso infelizmente não torna mais simples a solução do problema de se lidar com tal multiplicidade de línguas: mesmo que nos limitássemos

Figura 1: Distribuição Populacional das Línguas (1999)

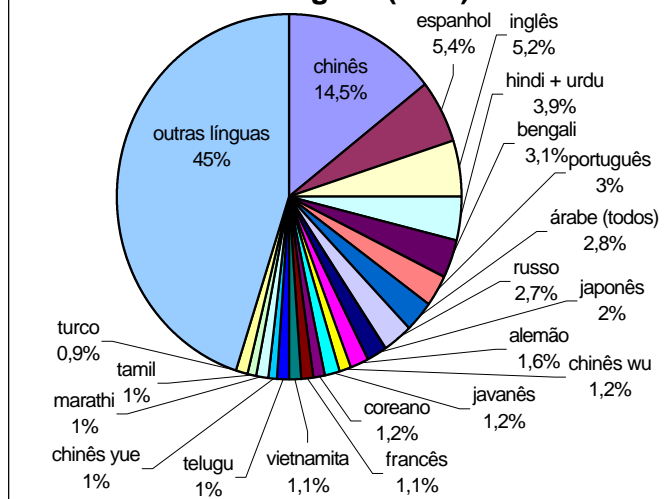


Figura 1: Distribuição populacional das línguas

às línguas que possuem pelo menos dez milhões de falantes, teríamos que contabilizar 100 idiomas [4]. Além disso, essas considerações quantitativas, ainda que possam ter uma justificativa prática, devem ser refreadas se quisermos nos pautar pelos princípios anteriormente mencionados, em particular no direito de acesso à informação, que deve ser universal e se estender a todo indivíduo, independentemente do fato de pertencer a um grupo étnico majoritário ou minoritário. Assim sendo, as soluções estudadas serão comparadas do ponto de vista de justiça: em que medida elas promovem ou não um tratamento igualitário dos interessados em relação ao acesso às informações da Web. Para isso tentaremos avaliar as soluções estudadas em termos de abrangência, para determinar em que medida cada uma se aproxima da realidade das pequenas comunidades lingüísticas e culturais.

Se direcionarmos o foco de nossa atenção para a diversidade atual de línguas dentro da *World Wide Web*, constataremos que um número muito menor de idiomas está efetivamente representado. A Figura 4, extraída de [6], e que compila dados referentes a março de 2001, indica que juntas as 11 línguas com maior número de usuários na WWW representavam 91,5% dos Internautas, sendo que mais da metade desse contingente correspondia a anglófonos.

Esse é contudo um resultado que ilustra bem a tendência de diversificação cultural mencionada acima, pois no início da década de 90 o percentual referente à língua inglesa chegava aos 100%. Trata-se então de um processo irreversível de difusão da rede Internet nos quatro cantos do mundo, e de diversificação cultural e lingüística da comunidade de usuários da Web.

Ainda mais notável é a tendência delineada na Figura 3, da mesma fonte, e que mostra a evolução do número de usuários que acessam a rede nas diversas línguas. Observamos que as projeções do estudo apontam para um total de 800 milhões de usuários de língua não inglesa no ano de 2005. O número de usuários das “pequenas línguas” somará então algo em torno de 130 milhões de indivíduos, a maior parte representada por falantes de línguas não ocidentais.

Figura 2: Populações Online por Línguas (Março de 2001)

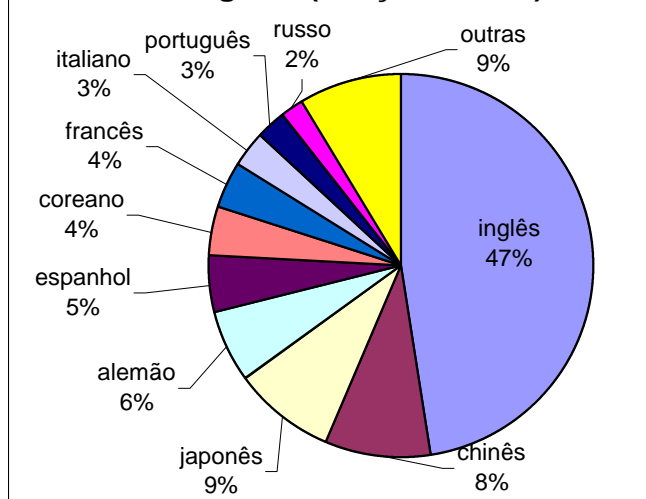


Figura 2: Populações lingüísticas online

Essa alteração da repartição da comunidade da Web de um padrão unilíngüe e monocultural para outro multilíngüe e multicultural certamente produzirá uma alteração da informação contida na Web, e também essa tenderá a refletir cada vez mais a diversidade cultural dos usuários, com a gradual redução relativa dos conteúdos em língua inglesa, e o correspondente incremento dos conteúdos nas demais línguas. A alteração das proporções dos conteúdos nas diversas línguas parece assim ser ao mesmo tempo causa e conseqüência da modificação da população *online*: quanto mais conteúdo na língua materna, maior o interesse em relação ao uso da Web, e quanto maior o uso, mais conteúdo na língua materna é gerado. Mas parece existir um certo lapso de tempo entre os dois fenômenos, como ilustra a Tabela 2, abaixo. Ela indica que mesmo com o número de usuários de língua inglesa já inferior a 50% do total, o conteúdo de informações em inglês ainda representa mais de 68% do total de páginas atualmente na Web.

Língua	Páginas da Web	Porcentagem do Total
Inglês	214.250.996	68,39
Japonês	18.335.739	5,85
Alemão	18.069.744	5,77
Chinês	12.113.803	3,87
Francês	9.262.663	2,96
Espanhol	7.573.064	2,42
Russo	5.900.956	1,88
Italiano	4.883.497	1,56
Português	4.291.237	1,37
Coreano	4.046.530	1,29
Outras	14.567.899	4,64
Total	313.296.128	100

Fonte: Vilaweb.com, 2000

Tabela 2: Número de páginas Web por língua

Em conjunto esses dois fenômenos, isto é, o aumento da base de usuários pertencentes a meios culturais e lingüísticos variados e a conseqüente pulverização dos conteúdos em um número cada vez maior de línguas não eliminarão as situações de *limitação lingüística* descritas acima para o cenário da Web dos anos 90, em que a quase totalidade do conteúdo era em língua inglesa. A diferença é que cada vez mais o problema da barreira lingüística tenderá a atingir um

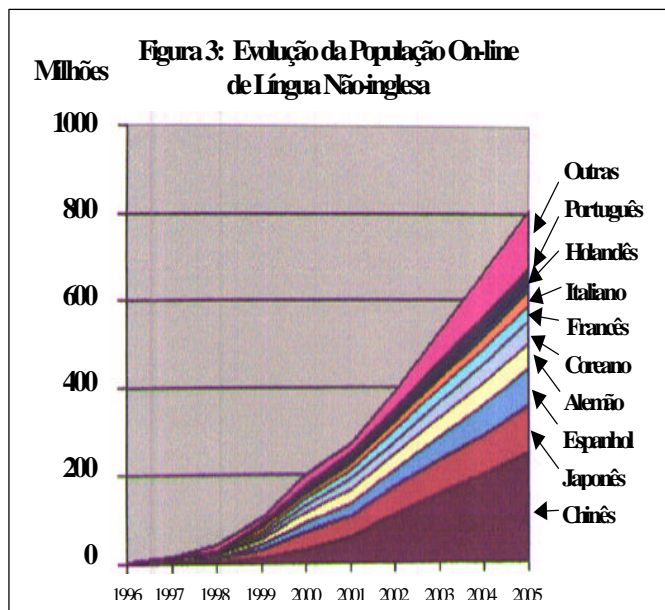


Figura 3: Evolução da população on-line de língua não-inglesa

maior número de usuários, inclusive os anglófonos, na medida em que cada vez mais páginas forem veiculadas na rede em língua não-inglesa. Com isso, cada pessoa que acessa a Web em busca de informações ver-se-á diante de uma séria limitação: o percentual de páginas da Web que ela terá em sua língua materna corresponderá a uma fração do total, maior ou menor conforme a língua do interessado. Assim, da utopia do acesso ilimitado a um mundo de informações chegaremos à dura realidade: todas as informações estão ali, ao alcance do ponto de vista tecnológico, mas distantes do ponto de vista lingüístico e cultural. Não é exagero, então, antever que num futuro não muito distante a Web refletirá, grosso modo, a espantosa diversidade de línguas existente no planeta, e seus usuários se depararão com problemas parecidos com os que afligem os *deficientes lingüísticos*, na condição de refugiados, trabalhadores imigrantes, viajantes, usuários das pequenas línguas, enfim, todos os que vivenciam, por conta de circunstâncias específicas, sofrimento, angústia, injustiça, frustração e cansaço mental em situações de limitação lingüística [1,2].

Muito embora, nos dias de hoje, a utilização da Web ainda não se faça por imposição, obrigação ou extrema necessidade, sendo em geral fruto de iniciativa individual, muitas vezes lúdica, é provável que cada vez mais ela seja utilizada para os mais diversos fins, seja como ferramenta educacional nos sistemas de ensino em todo o mundo, seja nos procedimentos relacionados à administração pública e outros serviços diretamente ligados aos direitos dos cidadãos. No primeiro caso, a utilidade da Web no contexto educacional é indiscutível se considerarmos que ela facilita pesquisas e buscas de informação, promove iniciativas de colaboração de estudantes geograficamente dispersos, viabiliza a criação de bibliotecas virtuais [37], entre outras. No âmbito da administração pública, o aspecto de redução de custos certamente fará com que muitos procedimentos de disponibilização de serviços ou de disseminação de

informações sejam totalmente substituídas pela Internet, sendo as outras interfaces abandonadas em favor da Web. Já se discute no Brasil o impacto da Internet na administração pública [33].

Na área internacional, é possível prever serviços sendo mundialmente oferecidos por organizações ligadas à ONU, tais como a OMS e a UNESCO. A OMS, por exemplo, lançou um projeto visando usar a Internet para permitir a estudantes nos 100 países mais pobres do mundo ter acesso gratuito aos principais periódicos de Medicina [32]. Esse esforço, que inclui até mesmo a doação de computadores, não leva em conta o fato de que as publicações, em sua maioria, estão em inglês. ONGs de atuação mundial também podem se servir da Web em suas atividades. Em todos esses casos será preciso lidar com a diversidade cultural e haverá sem dúvida situações de *deficiência lingüística* e discriminação cultural.

Diante disso, a preocupação com o problema da barreira das línguas nas interfaces computacionais da Web parece legítima e oportuna, e a percepção da deficiência lingüística como fator de limitação do pleno exercício da cidadania pode ser considerada pertinente.

2.1 Modelo de Controle

O modelo de controle adotado nesta análise visa reproduzir aproximadamente as características originais da Web no que se refere ao conteúdo, isto é, predominância da língua inglesa, mas associando a essa situação a distribuição populacional projetada para o ano 2005 com base nos dados de crescimento em [6] e cujo perfil está ilustrado na figura 3. A motivação de tal abordagem é tentar avaliar qualitativamente a situação hipotética em que a alteração do perfil cultural dos usuários da Web não tivesse sido acompanhada pela mudança lingüística de seu conteúdo. Isso poderia ser consequência da adoção de uma *regra de etiqueta (Netiquette)* na Rede, segundo a qual toda a informação deveria ser veiculada em inglês (e as poucas que não o fossem deveriam ser espelhadas por páginas de tradução naquela língua). Assim, todo o conteúdo seria acessível em inglês, mesmo que também houvesse dados em outras línguas. Queremos com base nesse modelo determinar em que medida um conteúdo unilíngue centrado na língua inglesa atenderia as necessidades de uma população multicultural.

Como mostrado na figura 4, as projeções para o ano 2005 estimam que de um total de um bilhão cento e vinte milhões de usuários da Web, cerca de 320 milhões (28,6%) serão de língua inglesa. A segunda maior população de usuários corresponderá aos de língua chinesa, com 250 milhões de indivíduos, enquanto que 132 milhões de pessoas acessarão a Web em línguas minoritárias, representando o terceiro maior contingente. Os japoneses devem ocupar a quarta maior fatia: 105 milhões. Os usuários do português somarão 40 milhões, representando cerca de 3,5 % do total.

Com base nessas cifras, e considerando a hipótese do modelo de controle, teríamos então a seguinte situação:

- Todas os grupos lingüísticos são minoritários: nenhum ultrapassa 30% do total;
- Os anglófonos ainda predominam, embora bem menos que nos anos 90. Eles serão doravante chamados de **Grupo 1**;
- As línguas germânicas, excetuando-se o inglês, limitam-se a 7,6% do total. Seus falantes serão referidos com sendo o **Grupo 2**;
- As principais línguas românicas (português, espanhol, francês e italiano) representam 17,7% dos usuários. A essa parcela corresponderá o **Grupo 3**;
- As principais línguas orientais somam 36,2% dos usuários. Esse contingente corresponde ao **Grupo 4**;
- As línguas minoritárias da Web, se tomadas em conjunto, formam um grupo quantitativamente muito importante: 11,8%. Nesse total estão idiomas de diferentes famílias, os principais são o árabe, o russo e todas as línguas eslavas, o hindi, o bengali, o tailandês, o malaio-indonésio, entre outros, mas nenhum estruturalmente semelhante ao inglês. Seus usuários formam o **Grupo 5**.

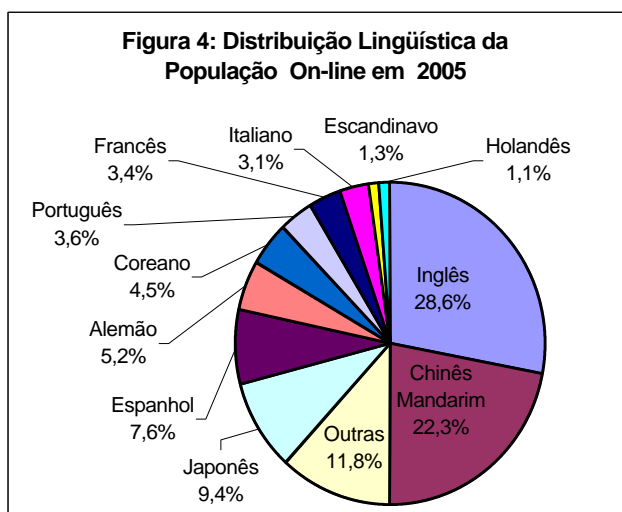


Figura 4: Distribuição lingüística da População Online em 2005

Diante desses números, podem ser feitas as seguintes considerações no que se refere à facilidade de acesso pelos integrantes dos distintos **Grupos** da Web no caso hipotético em que predomine a língua inglesa (100% do conteúdo disponível). Considerando os níveis de esforço necessários ao domínio da língua inglesa por parte dos respectivos grupos lingüísticos e os níveis de penetração daquele idioma nas respectivas culturas, teríamos a seguinte hierarquia:

- 28,6% têm pleno acesso sem limitação e sem esforço adicional → **Grupo 1**;
- 7,6% têm acesso elevado (pouca limitação) mas com um moderado esforço na aquisição do instrumento lingüístico → **Grupo 2**;
- 17,7% têm acesso parcial (bastante limitação), e mediante consideráveis esforços na aquisição do instrumento lingüístico → **Grupo 3**;

- 48% têm baixo acesso (muita limitação), e necessidade de enormes esforços na aquisição do instrumento lingüístico → **Grupos 4 e 5**;

Com relação ao aspecto ‘esforço na aquisição do instrumento lingüístico’, é importante ressaltar que os níveis de exigência variam conforme a “distância” da língua do interessado em relação ao inglês e também em função da maior ou menor penetração da língua inglesa nas respectivas culturas [2]. Assim sendo, do ponto de vista lingüístico, os falantes das línguas eslavas (Grupo 5) estão mais próximos do inglês do que os falantes das principais línguas orientais (Grupo 4), mas a influência cultural anglo-americana nos países correspondentes ao Grupo 4 historicamente é muito mais forte do que nos países eslavos, e com isso, ambos os grupos encontram-se, grosso modo, no mesmo patamar de limitação lingüística.

Alguns autores preferem dividir os países, no que se refere à penetração da língua inglesa, em três categorias [8]: ENL (*English as a Native Language*) que em nossa classificação corresponde ao **Grupo 1**; ESL (*English as a Second Language*) que inclui parte dos países do **Grupo 2** além de outros que por razões históricas tiveram forte influência colonial do Reino Unido (Singapura, Índia, etc.) ou dos Estados Unidos (Filipinas). Entretanto, o nível de proficiência na língua inglesa é bastante desigual entre as populações desses últimos países, e na maioria dos casos o percentual de habitantes que não domina a língua ultrapassa os 80%. Além disso, aqueles que a dominam já foram incluídos no total do **Grupo 1**. Por essa razão, preferimos utilizar o conceito de dificuldade de acesso à língua inglesa (por intermédio do sistema de ensino convencional) para o conjunto da população, o que de fato coloca a maioria dos habitantes desses países no **Grupo 5**.

A terceira categoria, chamada de EFL (*English as a Foreign Language*) pretende incluir o restante dos países, o que de certa forma falseia a realidade na medida em que ignora o caso de inúmeras regiões (basicamente na África) onde a língua ‘estrangeira’ é o francês ou o português. Mas independentemente das diferenças na classificação dos diversos países, a conclusão em [8] é de que uma hierarquização do acesso à Internet é consequência da predominância da língua inglesa em seus conteúdos, sendo o processo de internacionalização da Web limitado pela falta de domínio daquela língua em muitas regiões.

De fato, as barreiras lingüísticas são um fator determinante em relação às taxas de penetração da Internet mesmo em países ricos e classificados como ENL. No Canadá, por exemplo, as taxas de conectividade observadas no Québec, região onde predomina o idioma francês, são sensivelmente mais baixas do que no restante do país: enquanto o francês é a língua de cerca de 30% dos canadenses, somente 19,9% dos cibercidadãos daquele país pertenciam a este grupo lingüístico em 1996, sendo os falantes do inglês 76,8% do total [9]. Situação semelhante é observada nos Estados Unidos em relação aos cidadãos de língua espanhola. Em uma pesquisa feita

pelo Departamento do Comércio no ano de 2000, foi observado que entre a população hispânica o índice de penetração da Internet era mais baixo do que de qualquer um dos principais grupos étnicos do país, mesmo aqueles de baixo poder aquisitivo, como o dos negros [13]:

	Usuários da Internet	Total da População	Penetração da Internet por Etnia
Branco	93,7	186,4	50,3%
Negro	9,6	32,8	29,3%
Asiáticos	5,1	10,3	49,4%
Hispânicos	7,3	30,9	23,7%

Fonte: Departamento do Comércio - 2000

Segundo a fonte, uma explicação para o fenômeno seria a barreira da língua e o fato de que faltariam páginas Web em língua espanhola (menos de 3% do total), quando os três outros grupos, que acessam a Web em inglês, têm à disposição mais de dois terços do total de páginas.

Se nas sociedades acima citadas, ricas e do tipo ENL, as restrições na acessibilidade à Web atingem consideráveis parcelas da população, qual será a situação em sociedades mais pobres e que são do tipo EFL? Em um artigo tratando do papel atual e potencial da Internet para as comunidades indígenas da Federação Russa [14], o autor reconhece que embora as línguas daqueles grupos minoritários já disponham de recursos para veiculação na Web (escrita em alfabeto Cirílico) o fato de que o potencial da Internet exige um bom domínio do inglês é um limitador, visto que “o conhecimento do inglês não é muito comum entre os cidadãos da Federação Russa”. Além disso, “as pessoas das áreas rurais estariam ainda mais em desvantagem do que as das áreas urbanas.” Por conta disso, segundo o artigo, o fortalecimento das línguas nativas é ao mesmo tempo fomentado e dificultado pela Internet.

Os exemplos supracitados, assim como a análise de nosso modelo de controle nos levam a uma mesma conclusão: a predominância do inglês nos conteúdos disseminados na Web leva a uma hierarquização no âmbito da comunidade da Rede, com a maior parte de seus usuários potenciais sendo submetida a consideráveis limitações em seu acesso às informações disponíveis. Essa conclusão reforça a noção de que o problema da limitação lingüística na Internet não seria consequência exclusiva da pulverização dos conteúdos em diferentes línguas mas é um efeito inerente à diversificação cultural da coletividade que se utiliza da rede. O fato de que todo o conteúdo esteja disponibilizado em língua inglesa não soluciona o problema. O processo de aprendizado dessa língua, assim como de qualquer outro idioma ocidental, é extremamente penoso e exige, para um domínio considerado operacional, uma quantidade de horas de estudo muito além das possibilidades dos sistemas de ensino dos diversos países, exceção feita ao caso do ensino do inglês nos países do **Grupo 2** [2].

Além de limitar a acessibilidade da Web, a predominância da língua inglesa teria outros efeitos

negativos. Em um ensaio intitulado *Resisting Cyber-English* (1996) [11], o pesquisador de literatura norte-americana da Universidade de Berkeley, Joe Lockard, lançando um olhar crítico sobre a modalidade de comunicação predominante na Rede, baseada no inglês, aponta muitas das implicações dessa onipresença. Segundo ele, o *Cyber-English* constitui a maior barreira à participação eletrônica, sem que contudo seja percebido tal desajuste entre a *lingua franca* da Rede e o vasto mundo não-anglófono, em razão de um tipo de “cegueira coletiva que caracteriza a cultura *cyber-english*”. Ele acrescenta que as racionalizações geralmente feitas para justificar a situação baseiam-se no ‘argumento do piloto de aeronave’: “se nas comunicações aéreas o inglês foi adotado ‘por razões de segurança’ por que não tolerá-lo na Web?”¹ argumentam os defensores do *Cyber-English*. O problema, rebate o autor, é que “a grande maioria da população mundial não se encaixa nessas especificações de comunicação: eles necessitam de (e merecem) uma comunicação doméstica, em suas próprias línguas, em seus próprios termos, para seus próprios propósitos”. O autor salienta que o desenvolvimento da Internet acelerou a transformação do inglês em uma *commodity* que pode atravessar as fronteiras sob a forma de bits isentos de taxas. Esse processo de ‘comodificação’ da língua “recapitula as linhas imperialistas da produção na periferia e refinamento na metrópole”: textos em inglês são digitados a baixo custo na Índia e no Caribe enquanto que as atividades de *information design* e reformatação, de elevado valor agregado, são monopolizadas por trabalhadores bem remunerados nos Estados Unidos e na Europa. O fenômeno não envolve tão somente esses regimes estratificados de produção lingüística, mas “se baseia igualmente na capacidade dialógica, isto é, em uma habilidade semiótica adquirida relacionada ao *Cyber-English*, que determina o valor adicionado”. A periferia tem um papel limitado e não participa da produção final, nem se questiona sobre seus propósitos sociais. O valor de seu inglês se baseia em um monologismo, enquanto que o valor adicionado depende de um dialogismo que está fora de seu alcance. O *Cyber-English* limita o dialogismo aos consumidores finais ocidentalizados que podem falar a partir de um paradigma tecnológico que exige participação léxica. Nessa perspectiva, o valor adicionado no *Cyber-English* pode ser descrito como um domínio da proficiência e do controle.

Ainda segundo o texto, é bastante comum encontrar falantes não-nativos do inglês lutando para participar de listas de discussão com seus limitados conhecimentos da língua. Suas sentenças tendem a ser curtas, suas sintaxe e pontuação pobres. Por vezes, seus interlocutores os insultam por suas opiniões assim como por seu inglês.

Finalmente, o autor chama atenção para o fato de que vivemos em um ambiente tecnológico onde um

¹ De fato, problemas de comunicação devidos ao uso do inglês respondem por 11% das fatalidades na aviação mundial (David Crystal, 1998)

domínio totalmente inteligível e idiomático do inglês padrão tornou-se a autorização para uma opinião global, para uma expressão transmissível e inteligível através da rede. Falantes não-nativos do inglês seguiriam sendo os “novatos desorientados da Internet”, “uma classe global de um campesinato lingüístico incapaz de falar o latim tecnológico”. O *Cyber-English* “age como um filtro cultural que exclui as particularidades culturais e padroniza a experiência expressiva”.

O ensaio acima citado foi feito em meados da década de noventa. Os efeitos descritos podem ser extrapolados para nosso modelo de controle e, mais do que isso, podem ser exponencialmente aumentados na medida em que crescer o número de falantes não-nativos em um ambiente baseado no *Cyber-English*. Essas considerações nos levam então à constatação de que não se consegue acessibilidade através do uso do inglês como *lingua franca* da Web, e de que muitos efeitos nocivos decorrem de seu emprego, dos quais o mais grave é uma estratificação, uma hierarquização, que só acentuará as desigualdades entre países dominantes e periferia.

3. Soluções para prover acessibilidade:

Diante do exposto na seção precedente, e com base nas premissas de igualdade no direito de acesso às informações, passaremos a avaliar as possíveis soluções visando prover acessibilidade a todos os portadores de *deficiência lingüística*.

3.1. Ferramentas de tradução automática:

Assim como em todos os demais meios onde a diversidade cultural se fez notar, também na Internet surgiram, por conta da multiplicidade das línguas, situações de *deficiência lingüística*. E a exemplo do que já acontecera em outros meios, também no caso da Internet surgiram iniciativas visando resolver o problema através de ferramentas automáticas de tradução. A própria natureza computacional da Internet favoreceu esse tipo de tentativa, que conforme analisa Piron [2], funda suas raízes na tendência de nossa sociedade moderna a sempre procurar nas coisas a solução para seus problemas. Esses esforços na busca da máquina de traduzir na verdade remontam aos primórdios da computação, ainda nos anos 50. Passadas várias décadas, nas quais grandes somas de dinheiro e consideráveis esforços foram destinados a essa pesquisa, alguns dos sistemas atingiram o ponto de uso comercial, muitos dos quais no âmbito da Internet. Nesta seção tentaremos então avaliar se o estado da arte das ferramentas de tradução automática foi capaz de romper as barreiras lingüísticas na Web. Para tanto, analisaremos alguns destes *softwares*, tentando explicitar as suas características (positivas ou negativas).

Antes de iniciar a análise comparativa das ferramentas de tradução automática, é fundamental procurar entender a fundo o que representa uma atividade de tradução e quais são os requisitos de um bom tradutor. Em um trabalho apresentado no Simpósio Internacional sobre Tradução Automática (Budapeste, 1988) e intitulado *Learning From Translation Mistakes* [10], Piron, que durante quinze anos atuou como tradutor e

revisor na ONU e na OMS, e que há mais de trinta anos estuda a comunicação internacional em seus aspectos lingüísticos, enumera as qualidades que considera essenciais em um bom tradutor: discernimento e flexibilidade.

Discernimento “é a habilidade de resolver um problema através do amplo conhecimento do assunto, por meio da consciência de que o problema existe e levando em conta os vários níveis do contexto”.

i) O amplo conhecimento do assunto implica a necessidade de que o tradutor conheça a fundo a natureza da matéria na língua original e possa interpretar no texto, em termos semióticos [20], informações que fazem parte dos níveis pragmático (jargão, gíria, etc.) ou social (costumes, tradições, etc.), além, é claro, daquelas dos níveis sintático e semântico, que são essencialmente lingüísticos. Para tomarmos um dos exemplos da referência acima, se é preciso produzir em inglês a tradução da sentença francesa “*une messe de neuf heures*”, o tradutor deve estar ciente de que, no âmbito semântico, a palavra francesa *heure* pode significar em inglês *hour* ou *o'clock*. Para poder transpor corretamente a sentença ele precisará optar por uma entre as duas soluções. Mas a superfície do texto não fornece essa informação visto que sintaticamente é a mesma estrutura da sentença *un voyage de neuf heures* (uma viagem de nove horas). Ele tem então que apelar aos níveis pragmático e social para perceber que a tradução *a nine hour mass* é inadequada pois seria extremamente improvável uma missa católica durar nove horas. Ele deverá então optar pela tradução *a nine o'clock mass*.

Essa capacidade de fazer uso de diferentes níveis semióticos no processo de tradução é um dos ingredientes que formam um bom tradutor humano, mas é algo extremamente difícil de se incorporar em uma ferramenta de tradução automática.

ii) A consciência de que o problema existe é uma capacidade que o tradutor adquire após alguns anos de atividade, e que consiste em tomar consciência da existência de problemas dos quais anteriormente ele não tinha qualquer idéia. O processo leva algum tempo mas se o tradutor muda sua área de atuação (ex.: da área tecnológica para a área médica) o processo deve ser reiniciado até que ele readquira uma consciência dos problemas naquele terreno novo. Por exemplo, um tradutor que tenha que transpor a expressão *Japanese encephalitis vaccine* para outras línguas precisa estar consciente de que ela dá margem a duas interpretações (sintática e semanticamente), e que é então preciso recorrer a considerações pragmáticas, o que nesse caso significa empreender uma pesquisa para se informar de que a vacina em questão não é um medicamento produzido no Japão e que trata as encefalites em geral, mas sim uma vacina contra a encefalite japonesa. Sem estar consciente do problema (da ambigüidade) um tradutor corre o risco de introduzir contra-sensos no texto que produz. Assim como o item anterior, também este é de difícil implementação em sistemas de tradução automática;

iii) A consideração dos vários níveis do contexto é o terceiro componente do discernimento de um bom tradutor. Ele é importante quando na interpretação de um texto a análise semântica dependa de níveis contextuais diferentes, que devem ser identificados a fim de evitar erros. Tomando uma vez mais um exemplo de [10], a palavra inglesa *repression* possui duas traduções em francês: em política ela equivale a *répression* enquanto que em psicologia ela se traduz por *refoulement*. Assim sendo, a tradução correta depende do contexto em que o termo aparece. Entretanto, na prática, as coisas não são assim tão simples, pois o autor pode utilizar o sentido psicológico dentro de um contexto político mais amplo, como em um artigo tratando da era Stalinista em que o autor escreve *Repression by the population of its spontaneous critical reactions led to...* Nesse caso, embora o texto trate de política, a sentença trata de psicologia. O contexto estrito foge do contexto amplo. Embora o software de tradução possa inferir o contexto amplo através de heurísticas de termos encontrados, isso não ajuda a tradução em tais situações, e não existe uma forma mecânica para determinar que o autor passou, no contexto estrito, para outra significação do termo.

Por vezes, nem mesmo o conhecimento do contexto permite evitar as ambigüidades do inglês: em um texto de higiene veterinária a expressão *stable odours* permite duas traduções igualmente plausíveis: *odores estáveis* e *odores de estrebaria*. O mesmo se dá com a expressão *light particles* em um texto sobre física: pode significar *partículas leves* ou *partículas de luz*².

Flexibilidade é, ao lado do discernimento, uma das duas qualidades indispensáveis a um tradutor. Ela se refere ao aspecto malabarístico de seu trabalho. Dominar uma especialidade e as duas línguas relevantes não é o bastante. É preciso dominar a arte de constantemente pular de um idioma para outro, pois línguas são mais do que estruturas intelectuais, são universos [10]. Cada língua tem uma índole própria e tem soluções muito particulares para exprimir idéias. A dificuldade está no fato de que essas diferentes abordagens devem ser levadas em conta não somente no âmbito das palavras, mas também no das sentenças. É por isso que a sentença inglesa *Private education is in no way under the jurisdiction of the government*, embora composta basicamente por palavras de origem francesa, expressa o pensamento de uma forma bem diferente da feita nessa língua, a despeito da proximidade etimológica. Uma boa versão em francês seria: *L'enseignement libre ne relève en rien de l'Etat*. As diferenças no nível sentencial ficam mais evidentes quando fazemos uma tradução literal do francês para o inglês: *Free teaching does not depend in any way from the State*, que significa algo bem diferente da sentença inglesa original. Por essa razão, para traduzir de maneira adequada é preciso **sentir** quando e como passar de uma atmosfera para outra. Nenhum tradutor iniciante sabe como fazê-lo. Essa habilidade de ajustar-se às várias abordagens da realidade ou da fantasia que as

diversas línguas incorporam, e que está ligada à capacidade de constantemente passar de uma para outra é o que Piron chama de flexibilidade. Essa é a capacidade mais difícil de se encontrar quando se selecionam profissionais tradutores, e no que concerne sistemas de tradução automática, é inverossímil que algum chegue a esse grau de refinamento.

Os atributos acima descritos são necessários por conta dos problemas inerentes às línguas e que se dividem em dois tipos: gramatical e semântico. No primeiro caso, a língua ideal para os tradutores seria aquela que expressasse as relações entre palavras e seus elementos constituintes, através de recursos gramaticais precisos e bem delimitados. Neste aspecto, entre as línguas mais ingratas, segundo o autor [2,3], estariam o inglês e o chinês, na medida em que ambas são, em suas estruturas gramaticais, extremamente parcimoniosas, com poucos recursos que permitiriam eliminar ambigüidades. Esse é por exemplo o caso quando é preciso traduzir um texto que trata de uma epidemia desencadeada em um *Japanese prisoner of war camp* durante a Segunda Guerra Mundial. Como traduzir? A análise da superfície do texto admite duas possibilidades: uma epidemia em um campo japonês de prisioneiros americanos ou em um campo americano de prisioneiros japoneses. Os recursos gramaticais da língua inglesa não dão elementos para operar a escolha. É preciso então apelar para pesquisas ou tentar, quando possível, contactar o autor.

Em razão disso, traduzir a partir de textos em inglês (ou chinês) é sempre uma atividade de detetive, em que freqüentemente o tradutor tem que empreender buscas em bibliotecas, ligações telefônicas, apelar para correspondência, etc. Fica claro então que os sistemas de tradução automática não teriam como superar de forma confiável esses impasses, tendo que, quando muito, apelar para decisões estatísticas, nem sempre corretas.

No campo semântico, por outro lado, os problemas são especialmente difíceis para tradutores humanos por duas razões: ou o problema não é aparente, ou ele é visto prontamente, mas a solução requer um bom discernimento, como nos casos vistos acima, ou não existe, como naquelas situações em que uma palavra não tem tradução exata na língua alvo, e deve ser substituída por uma circunlocução que pode não ser totalmente fiel.

Diante dessas considerações, e levando em conta os pressupostos deste trabalho, que são buscar soluções que promovam a acessibilidade lingüística na Web, com um padrão de qualidade suficiente para aplicações em ensino à distância, atividades de intercâmbio científico, comercial ou cultural, se fazem necessárias ferramentas de tradução capazes de transmitir, senão integralmente, pelo menos eficazmente os textos em todos os seus níveis semióticos de interpretação. Para tanto, vamos testar alguns dos tradutores automáticos disponíveis na Web, avaliando em que medida eles conseguem verter um dado texto sem deformá-lo. Será também examinado o aspecto de abrangência da ferramenta: quantos pares de idioma ela consegue atender, tipos de texto, etc. A fim de estabelecer uma referência válida, escolheremos um texto que contenha informações cuja interpretação requiera

² Ambos os exemplos levantados a partir de erros produzidos por tradutores humanos profissionais.

algum tipo de análise pragmática e que tenha informação associada ao contexto. Optaremos por um texto original em inglês em virtude da predominância dessa língua nos conteúdos da Web.

Como nossa amostra destacamos um trecho de texto em um portal dedicado à popularização científica, sendo portanto uma fonte típica de informações para pesquisas escolares, mas cujo conteúdo está inteiramente em inglês. A amostra extraída foi a seguinte³:

“Bones are all well and good when it comes to figuring out body size and structure and even what dinosaurs ate. Most animals, however, growl, cry and sing with vocal cords and other soft tissues that just do not survive the ages like bones do when fossilized.”

Após submeter o texto acima ao sistema de tradução automática Systran, reputado como um dos melhores na Web, obtivemos a seguinte versão em português:

“Os ossos são tudo bons e bons quando vem a figurar para fora do tamanho e da estrutura de corpo e nivelam o que os dinosauros comeram. A maioria de animais, entretanto, rosnam, gritam e cantam com cabos vocais e outros tecidos macios que apenas não sobrevivem as idades como os ossos quando fossilized.”

Uma rápida leitura do texto produzido basta para nos darmos conta de que algo estranho deve ter ocorrido durante o processo de tradução. Se não somos capazes de ler o texto original, não conseguiremos extrair todas suas idéias com base na tradução. De fato, algumas palavras sequer foram traduzidas pela ferramenta, o que denota uma relativa pobreza lexical desta, ao menos no que se refere ao português. Mas o mais relevante na análise da tradução é perceber que o sistema não foi capaz de transpor as expressões idiomáticas *to be well and good*, e *to figure out*. Por conta disso, a tradução literal dos elementos dessas expressões produziu um péssimo efeito *non sense*. De resto, a ferramenta mostrou não ser capaz de inferir o contexto a partir do vocabulário disponível, o que se pode constatar pela tradução do termo *ages* como *idades*, ao passo que em se tratando de um contexto de paleontologia a melhor tradução seria *eras*. Fica evidente então que, do ponto de vista semiótico, a ferramenta mostrou-se incompleta, na medida em que tenta fazer uma transposição dos níveis sintático e semântico sem considerar o nível pragmático. A qualidade da tradução parece depender pouco das línguas envolvidas, pois o mesmo texto original deu origem a textos em espanhol e francês com os mesmos defeitos, exceto pela menor falta de vocabulário:

“Los huesos son todos bien y buenos cuando viene a calcular fuera de talla y de la estructura de cuerpo e igualan lo que comieron los dinosaurs. La mayoría de los animales, sin embargo, gruñeron, gritan y cantan con las cuerdas vocales y otros tejidos finos suaves que apenas no sobreviven las edades como lo hacen los huesos cuando están fosilizados.”

³ <http://www.discovery.com/exp/fossilzone/sounds/dinosounds.html>

“Les os sont tous bons et bons quand il vient à figurer hors de la taille et de la structure de corps et égalisent ce que les dinosauros ont mangé. La plupart des animaux, cependant, grognent, pleurent et chantent avec les cordes vocales et d'autres tissus mous qui juste ne survivent pas les âges”

Outras ferramentas de tradução automática foram testadas, a partir de um portal de tradução por máquina localizado em <http://www.foreignword.com/>. A interface permite a opção entre 28 línguas-fonte (todas européias, com exceção do japonês, do swahili, do africâns e do tswana, mas o número de línguas-alvo depende da língua fonte escolhida: um original em inglês permite 28 línguas-alvo, dentre as quais cinco não européias: africâns, zulu, persa, japonês e chinês. Já um original em português permite 23 traduções enquanto que um em africâns só pode ser traduzido para o inglês. A interface escolhe o software de tradução em função do par de línguas escolhidas, mas poucos pares lingüísticos permitem mais de uma opção. No teste, mantivemos o texto anteriormente usado e solicitamos uma tradução para o português. A única máquina disponível, chamada de InterTran™, produziu as seguintes linhas:

“Ossos são tudo tanto melhor quando é necessário calculando fora corpo tamanho e estrutura e já [what] dinosauros comido. Muita animais, agora, rosnar, choro e cantar com cordas vocais e outro fraca tecidos que justamente fazer não sobreviver as idades como ossos fazer quando [fossilized]”.

Um detalhe interessante é que para cada termo utilizado são oferecidas soluções alternativas: a palavra ‘tamanho’ vem acompanhada de 9 outros substantivos considerados equivalentes. Detalhes à parte, uma análise do texto produzido não deixa dúvida de que este software está ainda mais longe dos nossos objetivos do que o Systran acima descrito. Do texto não se conseguem extrair nem as idéias gerais.

Como alguns pares lingüísticos dispõem de mais de uma ferramenta de MT, resolvemos testar se uma alteração desta resolveria o problema da qualidade. Para, por exemplo, uma tradução em espanhol, dispomos de um total de 7 ferramentas. Destas, algumas se mostraram inacessíveis (por razões de servidor fora de serviço). Mas a ferramenta IBM Alphaworks estava disponível e gerou a tradução abaixo:

“Huesos son todo bien y bueno cuando venga figurando fuera de tamaño de cuerpo y estructura y aún qué dinosaurios comieron. La mayoría de los animales sin embargo gruñido llora y canta com cordones vocales y otros tissu suaves que ahora mismo no sobreviven las edades huesos similares dan cuando fosilizado.”

A maioria das demais ferramentas teve desempenhos comparáveis a este. A ferramenta e-lingo, por sua vez, levou ao seguinte texto:

“Los huesos son todo bien y bueno cuando viene a deducir tamaño del cuerpo y estructura e incluso qué dinosaurios comió. La mayoría de los animales, sin embargo, gruñe, lamento y canta con los cordones vocales y otros tejidos suaves que simplemente no sobreviven las edades como los huesos haga cuando se

fossilizó.” Percebe-se que uma qualidade um pouco melhor foi alcançada, e já é possível ter uma idéia geral do que trata o texto, mas uma das expressões idiomáticas continuou sem transposição adequada, enquanto que casos gramaticais da língua-alvo, tais como concordância e modos verbais, permanecem incorretos, o que é no mínimo inconveniente para um contexto de ensino.

Se ferramentas como essas, disponibilizadas abertamente na Web, não são capazes de transpor corretamente as idéias de um texto em inglês para línguas como o espanhol, o francês e o português, é pouco provável que possamos contar com elas para promover de fato o acesso de todo e qualquer internauta aos conteúdos da Web. Devemos, além disso, ressaltar que só foram avaliados softwares de uso gratuito, para não ferir os pressupostos do presente trabalho, e que muitos deles não se mostravam acessíveis quando necessário, o que caracteriza uma limitação da solução MT: ela depende do bom funcionamento do servidor de tradução, e se este fica indisponível, o usuário monoglota, por conta de sua deficiência lingüística, fica à mercê da sorte, à deriva em um oceano de informações em línguas estrangeiras. No mais, as ferramentas só tratam textos circunscritos, não sendo de muita ajuda quando se trata de recursos de bate-papo ou quaisquer outros em tempo real.

Diante desses resultados, registramos que a *tradução automática* não atende aos requisitos neste trabalho estabelecidos: prover acessibilidade lingüística a todos os portadores de limitação, com qualidade suficiente e sem discriminar as culturas minoritárias.

3.2. Tradução mediada pela meta-língua UNL (*Universal Network Language*):

O ano de 1996, já citado várias vezes neste trabalho, parece ter sido o ponto de inflexão que marcou o início da mudança do perfil dos usuários da Web, de uma comunidade estritamente unilíngüe para uma

multicultural e multilíngüe. Foi também naquele ano que a Universidade das Nações Unidas lançou um programa de pesquisa com o objetivo de criar uma “língua universal das redes” capaz de fomentar e facilitar a comunicação universal na Web por meio de Processamento de Linguagem Natural, apta a codificar o “sentido” e o “significado” (e não só a “forma” e o aspecto) dos documentos da Web, respeitando a diversidade das tradições lingüísticas. A coordenação do projeto, que foi denominado *Universal Network Language (UNL)*, coube ao *Institute of Advanced Studies* de Tóquio.

A UNL é uma língua formal projetada para prover um intercâmbio automático de informação multilíngüe e visa ser uma representação semântica interlingüística do significado de sentenças em língua natural. O cerne da UNL é um sistema KBMT (tradução por máquina baseada em conhecimento) desenvolvido por H. Uchida (1996).

Os *softwares plug-in* de codificação e de decodificação das línguas naturais estão sendo desenvolvidos por grupos de pesquisa em suas próprias línguas. Uma vez prontos, esses pacotes de software serão disponibilizados em servidores da Web e serão acessíveis via navegadores da Internet, “superando a necessidade de que as pessoas ao redor do mundo saibam a língua de seus interlocutores”.

Em sua primeira etapa, o projeto UNL conta com equipes de desenvolvimento ocupando-se de 14 idiomas: do grupo indo-europeu, português, francês, espanhol, italiano, inglês, alemão, russo, letão, e hindi; do grupo semítico o árabe; do grupo sino-tibetano o chinês mandarim; do grupo ural-altaico o mongol; do grupo malaio-polinésio o indonésio, além do japonês.

A arquitetura do sistema UNL está ilustrada na figura 5. Ela consiste de dois processadores principais, o codificador e o decodificador. Cada processador faz por

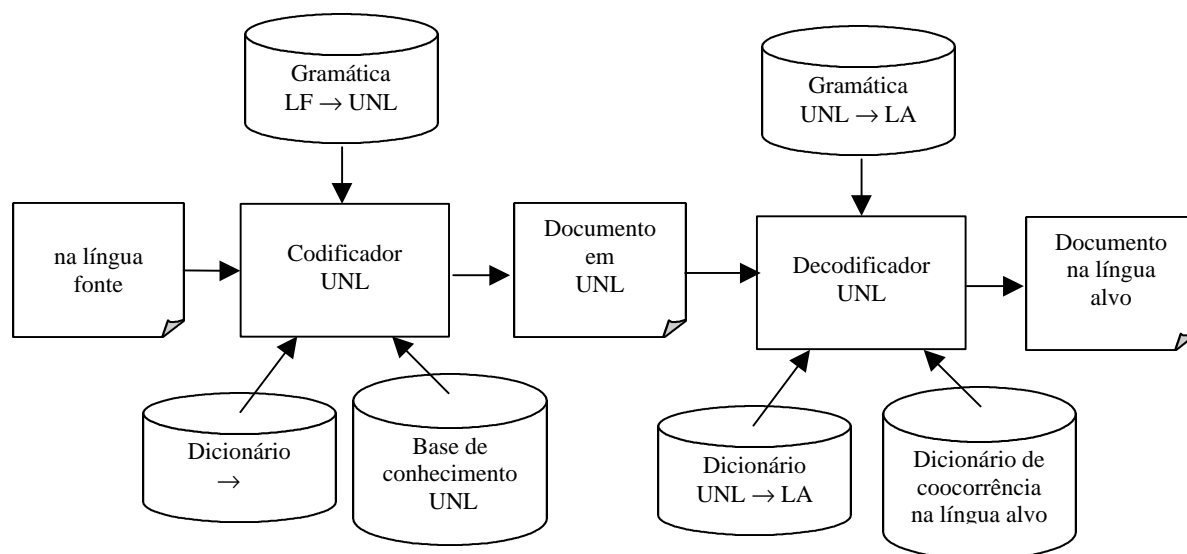


Figura 5: Arquitetura do Sistema UNL

sua vez uso de vários recursos lingüísticos: bases de conhecimento, dicionários e gramáticas mapeando a língua natural na UNL. Para cada língua envolvida no projeto devem ser desenvolvidos todos esses elementos. Um documento na língua-fonte é então convertido em um documento UNL. Esse documento UNL será eventualmente decodificado, produzindo um outro na língua-alvo [16].

A UNL aplica uma teoria tridimensional de significação de sentenças, cujos componentes se dividem em três grupos de entidades distintas: Palavras Universais (UWs), Rótulos de Relação (RLs), e Rótulos de Atributos (ALs), que recobrem os três níveis de representação: lexical, interlexical e sentencial, respectivamente. Pela sintaxe da UNL, a informação trazida por cada sentença pode ser representada por um hipergrafo cujos nós contêm UWs e cujos arcos significam RLs.

As UWs são rótulos para informação similar a conceito, correspondendo aproximadamente ao nível léxico da estrutura da sentença. Elas compreendem um amplo cabedal de palavras, supostamente capaz de denotar qualquer significado não composto que possa ser produzido por um falante de toda e qualquer língua. Por simplicidade de representação, optou-se por associar esses itens semânticos ao vocabulário da língua inglesa, muito embora não se trate sempre de um mapeamento puro e simples do léxico inglês, pois a UNL pretende ser um código de representação multilíngüe, e assim ser maior que o vocabulário daquela língua. As UWs estão sujeitas a relações ontológicas (sinonímia, antonímia, etc.) enquanto que os valores semânticos (atributos de agente ou objeto) são representados em uma Base de Conhecimentos. O conjunto de atributos e relações visa evitar ambigüidades. Mas resta o problema das diferenças de correspondência entre os vocábulos nas diversas línguas, visto que cada idioma categoriza o mundo e seus conceitos de uma forma totalmente particular. Assim, para cobrir os *gaps* semânticos entre as unidades léxicas a UNL propõe um léxico analógico, na medida em que as UWs, embora discretas, geram entidades contínuas, com as diferenças de campo semântico sendo preenchidas pela Base de Conhecimentos.

No que se refere aos Rótulos de Relação, eles correspondem a relações de valor semântico ligando informação similar a conceito. Existe atualmente um total de 44 RLs, mas esse valor tem sido freqüentemente alterado por razões empíricas de falta ou redundância. Embora se acredite que os RLs atuais sejam suficientes para codificar qualquer sentença em qualquer língua natural, isso ainda não pôde ser comprovado pois têm sido observados casos de superposição ou competição entre diferentes RLs.

Os ALs, por sua vez, indicam informações sobre as intenções do falante a respeito de cada parte específica da sentença (foco, tópico, ênfase, tema) ou sobre a estrutura como um todo (exclamação, interrogação, etc.), além de incorporar informações espaciais e temporais.

3.2.1. Limitações do Sistema UNL

Com base em uma publicação da equipe encarregada do desenvolvimento dos codificadores e decodificadores para a língua portuguesa [16], foram destacadas as seguintes considerações a respeito do projeto UNL:

- Não há restrições de domínio ou de estrutura do documento na língua-fonte, mas necessariamente há uma perda de expressividade semântica durante a codificação LN-UNL. O objetivo da UNL não é, em princípio, preservar totalmente o significado do texto, mas somente seus principais conteúdos, i.e., aqueles considerados essenciais. Entretanto, não existe, no Projeto UNL, um padrão mensurável do que seria 'essencial'. Por convenção, assume-se como sendo o sentido literal, aquele que está diretamente ligado à estrutura superficial das sentenças. Por conta disso, não há meios de representar conteúdos que não estejam mapeados diretamente nas estruturas sintático-semânticas da língua natural (níveis pragmático ou social);

- O fato de que os pesquisadores envolvidos tratam de questões relacionadas a suas línguas maternas seria um dos pontos fortes do projeto, mas, por outro lado, essa segregação dos esforços representa um problema crucial na medida em que os distintos grupos tendem a interpretar de forma diferente as especificações da interlíngua. Em razão disso é preciso um consenso acerca do formalismo da UNL;

- Para decodificar qualquer texto em UNL, a codificação LN-UNL é feita manualmente, uma vez que a customização do codificador para cada LN ainda não foi empreendida no projeto;

- Apesar de os resultados serem considerados promissores a qualidade atingida varia enormemente com a codificação da sentenças em UNL, a qual pode ser diferente para as distintas equipes de desenvolvimento. De resto, aspectos comunicativos da troca de informações na Web não são explorados em profundidade;

- O uso do inglês como *lingua franca* para comunicação entre as equipes de desenvolvimento e como língua de representação do conhecimento trouxe limitações ao projeto, visto que ele implica um nível indesejável de dependência lingüística. Línguas cujas estruturas sintáticas diferem profundamente daquelas do inglês podem apresentar um nível de complexidade adicional que torna impossível (ou não realista) o mapeamento da/para UNL. Somente um conhecimento do mundo, associado a índices contextuais, ambos ausentes no hipergrafo da UNL, evitaria codificações inadequadas;

- Outro problema diz respeito à ambigüidade inerente aos vocábulos da língua inglesa, a qual dá margem a análises conflitantes, mas que do ponto de vista da UNL são consideradas igualmente válidas e assim podem levar a corretas expressões de superfície, embora digam respeito a fatos semânticos distintos. A distinção de tais mapeamentos sintático-semânticos problemáticos é uma tarefa extremamente sensível ao contexto, muitas vezes ultrapassando os limites da sentença. As descrições da UNL não tratam desses níveis refinados de representação

de significado, estando limitadas aos significados derivados das sentenças fonte independentes do contexto, mesmo quanto tal independência implica informação insuficiente. A UNL realiza uma análise *default* para sentenças semanticamente ambíguas, e por isso pode-se dizer que a representação na UNL é probabilística e não determinística.

Ainda a respeito do Projeto UNL-Brasil, Martins [17] em um artigo sobre os problemas da ambigüidade semântica, relata que, “com freqüência desconcertante, os diversos grupos envolvidos no desenvolvimento dos codificadores/decodificadores para as várias línguas não têm sido capazes de decodificar, para a sua língua de trabalho, todas as sentenças codificadas em UNL, de forma automática ou manual, pelos outros grupos. [...] Embora cada um dos grupos venha tendo êxito na codificação e decodificação entre sua língua e o código UNL, repetindo, nesse sentido, o comportamento humano para a própria língua, o grupo não consegue decodificar automaticamente todas as sentenças codificadas por outros grupos, o que cancela o fim último da proposta, que seria permitir geração multilíngüe de representações de origem distinta”. Esses problemas teriam neutralizado as ações do projeto no último ano. O autor acrescenta que as facilidades e dificuldades encontradas independiam da competência dos codificadores ou da proximidade lingüística, visto que todos os códigos produzidos eram consistentes no âmbito de seu grupo, embora quando decodificadas por outros grupos algumas das sentenças geradas fossem problemáticas. Esses casos, porém, variavam de forma caótica, sem que se destacasse uma ou outra ‘sentença difícil’.

Quando a partir de uma sentença inicial em inglês⁴ os grupos tiveram que produzir uma tradução humana em suas respectivas línguas e com base nesta gerar (manualmente) uma codificação em UNL, os códigos produzidos, embora privilegiando o conteúdo semântico, foram bastante diferentes entre si. Como não se verificaram erros nas codificações, constatou-se que “era mistificadora a universalidade do significado da sentença”. Em outras palavras, cada grupo havia interpretado a especificação de forma correta mas totalmente idiossincrática, uma vez que uma leitura atenta da sentença autorizava três diferentes interpretações das relações entre alguns dos componentes.

Quando, a fim de evitar as ambigüidades de classificação, tentaram-se eliminar as imperfeições da especificação propondo uma nova redação do texto, explicitando, de forma mais precisa, as diferentes representações aplicáveis, a especificação tornou-se impraticável para alguns grupos, mas sem que houvesse, em termos de codificações geradas, uma redução significativa das divergências.

Segundo o autor, esse fenômeno “constitui um impedimento insuperável para os postulados originais do Projeto: trata-se da validade da dicotomia língua-objeto e metalíngua, ou entre línguas naturais e UNL”[17].

⁴ “Long time ago, people began to build a tower, wich seemed about to reach the heaven”

O autor salienta que o problema não estaria na língua utilizada (UNL), mas sim no fato de que “a univocidade das palavras e expressões da linguagem é sempre aparente, ou culturalmente (contextualmente) determinada”. Isso explicaria o atual processo de dialeção da representação UNL, e o fato de que não se pode afirmar a superioridade de um dialeto em relação ao demais. Diante disso, ele conclui que por conta da heterogeneidade dos usuários não é possível uma metalíngua totalmente unívoca, e que seria então válido investigar como inserir a ambigüidade na definição da própria especificação, sem impedir a heterogeneidade de seu uso, controlando contudo suas implicações.

Para os propósitos do presente trabalho, é possível concluir então que a ‘Solução UNL’, embora impulsionada por um legítimo desejo de confrontar e contornar o problema da barreira das línguas na Web, ainda apresenta muitas limitações graves, das quais as principais são, ao que parece, intransponíveis. Por conta disso o Projeto parece incapaz de eliminar os efeitos de restrição da acessibilidade lingüística na Internet.

Destaquemos então as características principais da UNL para posterior confrontação com as demais propostas:

i) No estágio atual, o projeto só contempla 14 idiomas. Muito embora neste conjunto estejam incluídas algumas das línguas mais faladas do mundo (e mais presentes na Rede), o critério da difusão certamente não foi o determinante na formação dessa lista, uma vez que línguas com grandes contingentes de usuários como o polonês (44M), o bengali (189M), o turco (59M), o coreano (75M), entre outras, não participam do projeto, enquanto que o letão, falado por menos de 2 milhões de pessoas, já faz parte.

ii) Levando-se em conta o enorme esforço feito nos últimos 5 anos no desenvolvimento de codificadores e decodificadores para as 14 línguas citadas, é inverossímil que semelhante energia venha ser despendida em favor de línguas minoritárias cuja base de usuários esteja em países pobres, sem tradição tecnológica e sem recursos financeiros para empreender por conta própria um projeto dessa envergadura. É também pouco provável que as Nações Unidas consigam patrocinar tais gastos para um número significativo de línguas, lembrando que há centenas delas com número de usuários maior ou igual ao do idioma letão. Em outras palavras, o projeto UNL não possui a granularidade desejável para os pressupostos de igualdade de oportunidade e de direito no acesso às informações da Web.

iii) A solução UNL limita-se à transposição lingüística de documentos, não propondo solução para as demais modalidades de comunicação oferecidas pela Web, tais como as listas de discussão, os diversos fóruns e ambientes de bate-papo, os quais constituem elementos fundamentais das aplicações de ensino à distância. Essa lacuna é grave, uma vez que, como discutido em seção precedente, é nesses ambientes de comunicação em tempo real (ou síncrona) que se manifestam as mais insidiosas formas de discriminação e de limitação lingüística;

iv) A solução UNL não dá aos autores de documentos da Web a certeza sobre a fidelidade das transposições lingüísticas produzidas na passagem pela meta-língua. Efeitos dos erros de decodificação sobre a integridade do texto original estão fora do controle das partes interessadas.

v) A solução depende da disponibilidade e do bom funcionamento de recursos computacionais, o que significa que se os servidores UNL não estiverem acessíveis, o interessado não será capaz de obter a tradução do documento.

Como no caso das traduções automáticas tratadas na seção anterior, também a solução UNL parece incapaz de solucionar por completo e em definitivo os problemas decorrentes da diversidade de línguas sobre a acessibilidade na Web. Ao que parece, as ferramentas computacionais vem sendo sistematicamente derrotadas pela complexidade da comunicação interlingüística.

3.3. Interfaces multilíngües, baseadas em conceitos de adequação local (*localization*) e em tradução humana;

Uma diferente abordagem visando superar as barreiras culturais e lingüísticas na Internet consiste em providenciar que o conteúdo e o formato das páginas Web estejam totalmente adaptados ao perfil cultural de um determinado local. O termo inglês *locale* designa exatamente essa noção: uma região com especificidades culturais e lingüísticas próprias. Por esse princípio, o número de *locales* é ainda maior que o número de línguas, visto que duas regiões onde predomina o mesmo idioma podem ser *locales* diferentes por conta de pequenas diferenças de ortografia ou de vocabulário. Esse é o caso do português do Brasil e de Portugal; do francês da França, do Canadá, etc.; do Inglês (britânico, australiano, americano, etc.). Além disso, devem ser considerados vários outros aspectos tais como moedas, datas, horários, formatos de números, etc.

As iniciativas de adaptação cultural se tornaram tão necessárias e comuns que muitos conceitos antes desconhecidos passaram a merecer atenção e interesse cada vez maiores. Dentre esses novos conceitos destacaremos os mais importantes para os propósitos deste trabalho [15]:

Internacionalização: é o processo de generalização de um produto a fim de que ele possa tratar múltiplas línguas e convenções culturais sem a necessidade de *redesign*. Ela tem lugar no âmbito do *design* do programa e no desenvolvimento da documentação.

Adequação local (*localization*): envolve tornar um produto lingüística e culturalmente apropriado para o local (país/região e língua) onde ele será usado e vendido.

Tradução: é só uma das atividades na adaptação cultural, juntamente com outras tais como gerência de projeto, engenharia de software, teste, etc.

Engenharia lingüística: aplicação de conhecimento da linguagem oral e escrita ao desenvolvimento de sistemas de informação, transação e comunicação, a fim de que eles possam reconhecer, compreender, interpretar e gerar linguagem humana. A tecnologia lingüística inclui

tradução automática ou auxiliada por computador, verificação, reconhecimento e síntese de voz, buscas semânticas e recuperação de informação, etc.

UNICODE: conjunto de caracteres de comprimento fixo (16 bits) que é capaz de representar mais de 65000 caracteres, o que é suficiente para incluir praticamente todos os alfabetos existentes, além de símbolos de uso geral. Seu objetivo é substituir o ASCII e outros conjuntos de caracteres. Ele já é suportado pela maioria dos sistemas operacionais atuais, incluindo MAC OS e Windows. Sua maior vantagem é a capacidade de representar qualquer texto com uma codificação não ambígua.

Linguagem controlada: é um subconjunto de uma língua natural, usada principalmente para escrever material técnico, e que é mais estruturada e fácil de entender do que linguagem não controlada. Também chamada de *easy (simplified) english* é um método de escrita que ajuda autores a reduzir a ambigüidade e aumentar a legibilidade dos textos. O uso de linguagem controlada é um fator importante na redução dos custos e do tempo necessário para a produção de documentação e textos de ajuda. Não existe um padrão único, vários sistemas são usados com variados graus de sucesso [15]. O sistema conta com uma lista de termos e expressões proibidos assim como uma lista de sinônimos permitidos, além de um conjunto de regras gramaticais. Uma das vantagens da linguagem controlada está no fato de que ela aumenta a tradutibilidade do texto, especialmente quando a tradução auxiliada por computador é usada no processo de adequação local (o ganho de tempo chegaria a 25% [15]).

Considerações sobre adequação local de páginas Web:

Segundo Esselink[15], são atualmente muito comuns sites Web que fazem uso de repositórios de bases de dados e que exibem dinamicamente texto e outras informações para as páginas HTML ou ASP. Com a introdução do XML, o uso de bases de dados para criar conteúdos de sites Web se tornará ainda mais difundido. Para fins de adaptação cultural, isso significa que em vez de se ter que adaptar um grande número de páginas HTML, será necessário traduzir tabelas de bases de dados com conteúdo estruturado, o que fará com que a atividade de adaptação cultural de páginas Web se torne cada vez mais comparável àquela de adaptação de softwares. Um outro ponto complexo na adaptação cultural de páginas Web diz respeito à velocidade em que o conteúdo é atualizado. Idealmente, toda atualização de portais multilíngües deveria ser publicada simultaneamente em todas as línguas visadas. Isso significa que o processo de manutenção de um portal multilíngüe tenderá a exigir soluções de *workflow* além de uma impressionante equipe de profissionais diversos: engenheiros de adequação local e de qualidade, tradutores, revisores, gerentes de engenharia e de lingüística, gerentes de projeto, revisores nos países-alvo, etc. [15].

Cuidados como o uso de caracteres UNICODE e de linguagem controlada estão longe de ser suficientes para garantir um processo de adaptação fácil. Em muitos casos a adaptação cultural de uma página leva a

problemas de redimensionamento por conta de diferença de comprimento de textos nas diversas línguas: em geral um texto original em inglês resulta 30% mais longo após traduzido, mas o incremento pode chegar a 100% [15]. Esse problema é mais grave em línguas ideográficas (chinês, japonês) que não admitem abreviações e podem exigir expansões verticais.

Quanto à tradução dos textos, o problema tende a crescer com o número de línguas-alvo, visto que a tradução auxiliada por computador não está disponível para todas as línguas, e como já visto, nem mesmo o uso de linguagem controlada garante a fidelidade dos textos produzidos, pois muitas das decisões dependem de considerações de ordem pragmática ou social, que estão fora do alcance dos sistemas informáticos.

Uma alternativa seria a tradução humana, que, se competente, costuma dar resultados bem melhores que os da tradução automática. Nesse caso a maior dificuldade está nos custos. Os preços dos serviços de tradução giram em torno de US\$0,11 por palavra de um texto comum, mas em se tratando de páginas Web, mais exigentes em termos de formatação hipertextual, o valor pode chegar a US\$0,35 (o que significa que a tradução de uma página de tamanho médio pode custar de US\$45 a US\$75) [18].

Porém a tradução humana tem outras limitações: a disponibilidade de tradutores para todas as línguas minoritárias é algo impraticável, e portanto a solução tende a ter pouca granularidade. A maioria dos portais internacionais se contenta com um leque de 6 idiomas, enquanto os mais ambiciosos chegam a 30. A escolha das línguas depende então do potencial de consumo das comunidades e dos custos de adequação ao seu *locale*. Por conseguinte, as iniciativas de adaptação cultural de páginas Web tendem a ser norteadas só por fatores de mercado, o que as leva a ser incompletas e propensas a promover discriminação das comunidades lingüísticas mais minoritárias. Essa é uma lacuna da solução. Mas todos os conceitos por ela desenvolvidos podem ser úteis para nossos propósitos. A disponibilidade de um sistema como o UNICODE é um passo fundamental para a democratização da informática e da Web. Com ele deixam de existir as limitações de ordem técnica para a disponibilização de conteúdo nas diversas línguas, e isso sem dúvida estimulará as diversas comunidades lingüísticas (minoritárias e não-ocidentais) a aumentar sua participação na rede, produzindo mais conteúdo em suas próprias línguas. A pulverização cultural da Web tenderá então a se acelerar nos próximos anos, o que reforça os motivadores do presente trabalho.

As modalidades de solução descritas nas seções precedentes se mostraram incompletas na medida em que não atendem os pressupostos de qualidade, robustez, disponibilidade e abrangência desejáveis quando se busca solucionar o problema da limitação lingüística nas interfaces da Web. Os motivos para esse desempenho não satisfatório podem ser resumidos da seguinte forma:

i) Um conteúdo centrado na língua inglesa (tal como em nosso modelo de controle) deixa uma grande parcela dos usuários potenciais da Web em situação de muita limitação, e os que conseguem, com considerável esforço,

ter algum acesso, se sujeitam a uma estrutura hierárquica na qual eles certamente não terão mais do que um papel marginal: não terão capacidade dialógica, estarão sempre na periferia, sendo influenciados sem jamais serem capazes de influenciar os que estão no centro;

ii) As ferramentas de tradução automática tentam eliminar as barreiras lingüísticas na Rede, mas, como visto anteriormente, os requisitos de um bom tradutor estão fora do alcance das implementações atuais, principalmente devido ao fato de que a língua-fonte predominante na Web, o inglês, é particularmente ambígua em suas estruturas gramaticais e semânticas, o que dificulta muito a tarefa à qual os *softwares* de tradução se propõem. A tendência de que os usuários do chinês mandarim se tornem a curto prazo o segundo maior contingente da Rede não é alentadora na medida em que também sua língua é de difícil transposição (por razões similares àquelas do inglês);

iii) O Projeto UNL, ainda que motivado por uma legítima preocupação em promover a acessibilidade lingüística da Web, vem se deparando com grandes contratemplos em seu desenvolvimento, talvez por conta de suas limitações intrínsecas. No mais, no estágio atual está longe de ter o alcance desejável em termos de comunidades lingüísticas atendidas;

iv) As iniciativas de adaptação cultural de interfaces da Web baseadas em tradução humana (auxiliada ou não por tradução automática) têm como principal inconveniente os custos envolvidos, os quais necessariamente sujeitarão os esforços às leis de mercado, o que equivale a dizer que as sociedades mais pobres (sobretudo na Ásia e na África) formadas por grupos lingüísticos minoritários, que potencialmente seriam as grandes beneficiárias das informações que a Internet pode difundir a baixo custo, não terão acesso a esses conhecimentos pois eles certamente não serão disponibilizados em suas línguas locais. A veiculação dessas informações nas *linguas francas* tradicionais daquelas regiões (inglês, francês, etc.) além de não atingir o grosso da população, que só domina as línguas locais, poderá ter efeitos nocivos: ampliar a desigualdade social naquelas populações, na medida em que a elite ocidentalizada estiver ainda mais privilegiada por gozar de acesso a uma mídia inacessível às demais camadas da população; haverá um impulso a mais no processo de desaculturação daquelas sociedades na medida em que se firmar a percepção de que o acesso ao mundo só é possível através das línguas ocidentais.

Alguma solução alternativa?

3.4. Uso de língua internacional auxiliar

Diante da natureza dos problemas acima descritos, que é essencialmente lingüística e não tecnológica, somos levados a considerar o fato de que também a solução para as questões que ora discutimos possa ser lingüística. Assim, a título de hipótese de trabalho, vamos avaliar agora em que medida o uso de uma língua internacional auxiliar, uma das soluções propostas para a comunicação internacional no sentido amplo, poderia ser de alguma utilidade para o problema específico da comunicação intercultural via Web. Baseando-nos nas conclusões de um estudo empreendido

pelo pesquisador suíço Claude Piron em *Le Défi des Langues* [2], constatamos que, em diversos aspectos e nas mais variadas situações interculturais, a utilização de uma língua internacional auxiliar apresentou muitas vantagens quando comparada às soluções usuais: interpretação simultânea, tradução automática, ou o recurso a uma língua estrangeira. As razões, conforme demonstrado na referida obra, se devem a um conjunto de fatores:

1º) O aprendizado de qualquer língua ocidental, para que se atinja um domínio considerado operacional, exige uma quantidade extremamente grande de horas de estudo e prática e, com raras exceções, nunca é alcançado pelos sistemas de ensino da maioria dos países. Isso se deve não a falhas dos sistemas de ensino mas ao fato de que as línguas em questão são estruturadas de forma totalmente arbitrária (por razões inerentes à sua condição de línguas étnicas), com um número incontável de regras e exceções que tornam a sua assimilação extremamente demorada e penosa: a carga horária mínima gira em torno de 2000h para o domínio operacional, e 12000h para o pleno [2]. Dessa forma, o domínio dessas línguas (em particular o inglês) fica restrito a uma minoria que dispõe de recursos para longos estudos e para estadas no exterior.

2º) Os sistemas de interpretação simultânea têm uma relação custo/eficácia extremamente baixa, devido à complexa natureza da atividade de intérprete do ponto de vista semiótico. Esses sistemas, de resto, estão fora de nosso escopo atual;

3º) Por fim, as limitações da tradução automática já foram suficientemente tratadas em seções anteriores.

Logo, somos levados a reconhecer que as soluções correntes para o problema da comunicação internacional não atendem nossos requisitos. A solução ‘língua internacional auxiliar’ será então considerada.

Inicialmente, é oportuno lembrar que para os nossos propósitos uma língua internacional auxiliar deve ser bem adaptada para sua finalidade básica de facilitar o intercâmbio entre pessoas de diferentes culturas e línguas, a fim de evitar os inconvenientes das línguas naturais usadas para a comunicação internacional. Ela deve ser capaz de atingir um ponto ótimo em termos de riqueza e expressividade por um lado, e por outro lado ter simplicidade e clareza. Ela deve ter atributos que a tornem acessível ao maior número possível de indivíduos (a rigor a todos) independentemente da cultura a que estes pertençam, e ser além disso capaz de veicular todas as nuances do pensamento humano. Esse ponto ótimo de equilíbrio foi descrito por Umberto Eco em sua obra *La ricerca della lingua perfetta nella cultura europea* [19] ao descrever a língua internacional auxiliar Esperanto: “Enquanto as línguas *a priori* e as gramáticas ‘lacônicas’ se esforçavam tanto quanto possível a buscar o *princípio da economia*, o Esperanto visa ao *da otimização*.” Assim, segundo Eco, essa otimização dos recursos gramaticais é conseguida através da utilização de um mínimo de recursos que seja suficiente para dar ao idioma clareza e flexibilidade. Além disso, o esperanto possuiria outros atributos interessantes: “O alfabeto, de 28 letras, baseia-se no princípio ‘para cada letra um som, para cada som uma letra’. O acento tônico é sempre na penúltima sílaba.

O artigo tem uma única forma [...], não existe artigo indefinido”, o léxico se baseia em um mínimo de vocábulos que é expandido por meio da lógica de sufixos e prefixos. De resto, o léxico é otimizado ao tirar proveito, sempre que possível, de raízes comuns a várias línguas (as chamadas palavras internacionais).

Visto que uma língua internacional deve privilegiar ao máximo a clareza, no Esperanto as ambigüidades lexicais foram evitadas pela adoção de diferentes raízes para designar diferentes conceitos (baixa homonímia) e um mínimo de raízes para uma dada noção semântica (pouca sinonímia). Para exemplificar tais critérios, comparemos o léxico do inglês com o do esperanto. Vemos que enquanto o inglês usa um mesmo vocábulo para designar os distintos conceitos *letra* e *carta* (*letter*), o esperanto usa dois: *litero* e *letero*. Com freqüência, em inglês um mesmo vocábulo representa uma gama de diferentes conceitos, ao passo que em esperanto cada um desses conceitos é representado por um vocábulo específico. A tabela seguinte dá uma idéia da ambigüidade inerente à língua inglesa: um mesmo item lexical pode pertencer a três distintas categorias gramaticais (substantivo, adjetivo ou verbo) e conter uma enorme gama de significados. Esse tipo de fenômeno, que torna o inglês particularmente ingrato como língua fonte em sistemas de tradução automática, parece ser uma das limitações inerentes ao sistema UNL, o qual, como vimos, extrai do léxico inglês os rótulos de suas UWs (palavras universais). Esse é também um dos motivos do surgimento do conceito de *linguagem controlada*, já tratado. O esperanto, por sua vez, torna a análise sintática mais simples uma vez que cada elemento da sentença tem sua função gramatical explicitada pela desinência utilizada: *-o* (substantivos), *-a* (adjetivos), *-i* (verbos no infinitivo), etc. Além disso, do ponto de vista semântico, os itens léxicos são mais precisos que os das línguas étnicas em geral, o que torna a interpretação das sentenças bem menos problemática, tanto para leitores humanos quanto para sistemas de tradução. Nesse sentido, produzir um texto em esperanto é verbalizar os pensamentos de forma clara, usando os elementos necessários e suficientes, expressando as nuances através dos moduladores (afixos) (R. Saussure [20]).

Inglês	categoria	português	esperanto
mean	substantivo	meio	perilo
		recurso	rimedo
		média	averaĝo
	adjetivo	mau	malbona
		mesquinho	malnobla
			senhava
			fianima
		pobre	malriĉa
		miserável	mizera
		mediocre	mezbona
		meio	meza
		sovina	avara
		malvado	malica
	verbo	pensar	pensí
		significar	signifi
ter em vista		celi	
tencionar		intenci	
pretender	pretendi		

Além disso, como a língua conta com mais de um século de uso em que serviu de veículo para interações entre pessoas de variadas culturas e nos mais diferentes campos de conhecimento, ela herdou um substrato interétnico no qual se firmaram aquelas expressões cujo sentido era o mais adequado do ponto de vista de um público multicultural, diferentemente do que se verifica com as línguas étnicas, inclusive o inglês, onde os costumes lingüísticos e as expressões idiomáticas surgem e se estabilizam em meio a uma única comunidade cultural, sendo, por isso, totalmente idiossincráticas, sem correspondente direto nas demais línguas. A título de exemplo, em inglês, como em muitas línguas européias, as designações geográficas trazem consigo um referencial eurocêntrico: *far east* (extremo oriente) contém em inglês a idéia de que os países do leste da Ásia seriam ‘distantes’. Em esperanto, tal estrutura seria inaceitável pois a língua pretende ser internacional e não européia. A forma utilizada é então *orienta Azio*. De resto, alguns idiomatismos do inglês são culturalmente tendenciosos, como por exemplo a expressão *who wears the trousers* (quem usa as calças), que significa *quem toma as decisões, quem está no comando*, e que não é uma convenção válida em muitas outras culturas.

Mas o aspecto mais relevante do esperanto diz respeito a sua estruturação, como demonstrado por Piron [2]. Sua gramática, em vez de estabelecer um conjunto de diretrizes arbitrárias e incoerentes, como na maioria das línguas étnicas (sobretudo as ocidentais), baseia-se no nível mais fundamental e universal da verbalização humana, aquele descrito por Piaget como sendo o mecanismo da assimilação generalizadora. Segundo esse princípio, a fase mais fundamental do aprendizado lingüístico, em toda e qualquer criança, de qualquer parte do mundo, é caracterizada pela adoção de uma gramática regida por esquemas com validade absoluta, que não trazem casos de exceção. Em Norman [21] esse mecanismo é denominado *overgeneralization*. Na medida em que as línguas étnicas, em sua maioria, carregam estruturas gramaticais com uma enorme quantidade de exceções, a linguagem das crianças nesse estágio, que predomina entre as idades de 2 e 7 anos, se mostra freqüentemente em desacordo com a norma culta. No entanto, do ponto de vista psicológico, a criança é mais lógica do que as normas da língua, pois ela generaliza (extrapola) os esquemas assimilados, seguindo o fluxo natural de sua verbalização, inconscientemente buscando seguir a lei do mínimo esforço (mínimo gasto de energia nervosa) [2]. Ao longo desse período, as incontáveis exceções (desvios) da gramática da língua materna deverão ser incorporadas, isto é, transformados em esquemas mentais, mas o processo é lento: a formação desses esquemas e sua transformação em reflexos demandam anos de imersão total naquele universo lingüístico.

Por conta desse mecanismo espontâneo de verbalização, se a criança eventualmente se depara com a tarefa de aprender uma língua estrangeira, ela terá um esforço ainda maior que aquele feito na assimilação de sua língua materna: ela deverá incorporar todos os

reflexos correspondentes às especificidades gramaticais da língua estrangeira. Mas para isso ela terá tanto que inibir a sua tendência natural de generalizar os esquemas aprendidos (assimilação generalizadora), quanto inibir os reflexos condicionados que tratam das exceções da língua materna, os quais em geral não têm correspondente direto na língua estrangeira. O pequeno exemplo abaixo dá idéia desse tipo de esforço: uma criança de língua materna portuguesa incorpora em seus esquemas de verbalização a estrutura consistindo em formar o tempo futuro com o verbo auxiliar *ir*. Porém, devido ao mecanismo da assimilação generalizadora, ela dá ao esquema uma validade absoluta, o que a conduz ao erro quando o verbo a pôr no futuro é *ir*:

Português	Inglês
Eu vou comer	I will eat
Eu vou brincar	I will play
Eu vou poder	I will can (incorreta)
Eu vou ir (incorreta)	I will go

Após algum tempo, por força das muitas correções, ela se conforma à norma e incorpora essa exceção como esquema mental. Se, posteriormente, ela tiver que aprender a língua inglesa, ela encontrará um mecanismo muito semelhante na formação do tempo futuro por meio de um verbo auxiliar. Mas assim como em português, também no inglês o mecanismo não é absoluto e comporta exceções, mas estas não correspondem àquelas do português, e é por isso que a aprendizagem neste caso deverá se fazer tanto pelo bloqueio da verbalização espontânea quanto pela inibição dos esquemas da língua materna, que a conduziram diretamente ao erro.

Casos como o do exemplo acima existem aos milhares todas as vezes que nos defrontamos com o aprendizado de uma língua estrangeira. Isso explica por que tanto tempo é necessário para alcançar-se um nível satisfatório, e por que tão poucas pessoas conseguem atingir a meta sem apelar para recursos de imersão total por longos períodos.

Diante do exposto, é então possível compreender em que medida um idioma cuja estruturação se baseie no mecanismo da assimilação generalizadora permite o recurso aos níveis mais profundos e espontâneos da verbalização. Esse é particularmente o caso do Esperanto, pois todos os seus aspectos gramaticais são plenamente generalizáveis (não existem exceções às regras), o que equivale a dizer que o processo de aprendizado ocorre com muito mais eficácia e em muito menor tempo quando comparado ao das demais línguas. O processo de verbalização ocorre sem desvios e com o menor gasto de energia nervosa possível. Esse sistema tem além disso a vantagem, como já mencionado, de ser universalmente válido, visto que o mecanismo da assimilação generalizadora faz parte do desenvolvimento lingüístico de todos os indivíduos, nas mais diferentes culturas. Isso confere ao esperanto uma real capacidade de preencher sua finalidade: servir de veículo para os intercâmbios entre pessoas de culturas distintas.

Para os propósitos do presente trabalho é então razoável avaliar em que medida esta língua internacional

auxiliar poderia resolver o problema do *handicap* lingüístico nas interfaces da Web.

Como visto anteriormente, o fato de que o conteúdo da Web estivesse em inglês não resolveria satisfatoriamente o problema da acessibilidade lingüística pois o domínio operacional daquela língua exige uma quantidade de horas de estudo que está fora do alcance da maioria dos potenciais usuários da Rede. Além disso, seu uso estabelece uma hierarquização dos níveis de acessibilidade que tem muitos efeitos nocivos do ponto de vista cultural. Diante disso, podemos avaliar como seria a situação hipotética em que o conteúdo da Web fosse acessível em uma língua internacional auxiliar, neutra e de fácil aprendizado, mais especificamente o esperanto. As primeiras questões a se levantar são de ordem prática:

- i) Quanto tempo é necessário para seu aprendizado?
- ii) Quantas pessoas já usam a língua?
- iii) A língua possui recurso para veicular todos os assuntos e conteúdos disponíveis na Web?
- iv) A Web dispõe de recursos para suportar o esperanto?
- v) Existem outras alternativas entre as línguas internacionais auxiliares?

Vamos tentar encontrar respostas para essas questões:

i) Todas as características da língua esperanto tornam seu aprendizado mais fácil e rápido quando comparado àquele dos idiomas estrangeiros. Vamos resumir esses atributos:

- Estruturação baseada na assimilação generalizadora: verbalização espontânea;
- Alfabeto fonético: eliminação dos problemas de ortografia e de pronúncia;
- Vocabulário básico reduzido (exponencialmente ampliável por meio de afixos);
- Uso de raízes internacionais: um certo número de cognatos da língua materna;
- Ausência de expressões idiomáticas;
- Grande flexibilidade sentencial: adapta-se aos diversos costumes lingüísticos);

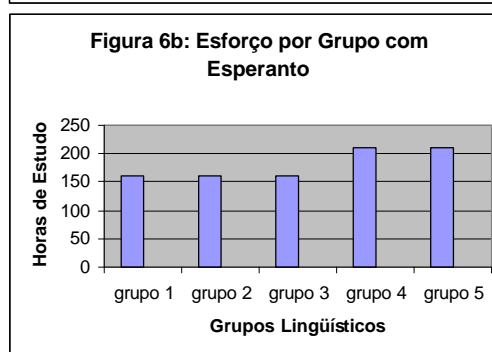
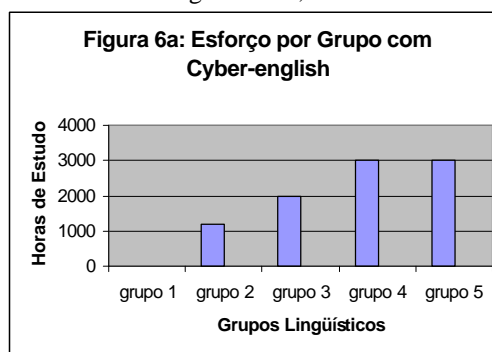
Em consequência desses atributos, os resultados empíricos ao longo de várias décadas comprovaram a maior facilidade do aprendizado do esperanto em relação aos demais idiomas. Em termos quantitativos, a maior ou menor facilidade de aprendizado de um idioma pode ser determinada com base no número de horas de estudo necessárias para o seu domínio operacional. Como mencionado anteriormente, para o inglês a facilidade depende muito da língua materna do aluno: aqueles de língua germânica têm menos dificuldade pois as estruturas da língua inglesa já lhes são familiares. Os de língua românica têm bem menos facilidade, mas ainda assim tiram proveito do grande número de palavras de origem latina no inglês. Os demais povos têm dificuldades progressivamente maiores. Mas, em termos gerais, temos os seguintes limiares [2]:

de 1500 a 2000 h	domínio parcial
de 2000 a 12000 h	domínio operacional
acima de 12000 h	domínio pleno

As cargas horárias mencionadas incluem atividades extra-classe, além de todas as situações de imersão total.

No que se refere ao aprendizado do esperanto, os resultados das pesquisas feitas pelo Instituto de Pedagogia Cibernética da Universidade de Paderborn (Alemanha) determinaram que seriam necessárias em média 160 horas de ensino para atingir em esperanto um nível que, no caso do inglês, demanda 1200 horas [2]. É uma relação de 7,5 vezes. Se levar-mos em consideração o fato de que para sujeitos cujas línguas maternas são muito distantes do inglês (línguas orientais, por exemplo) a proporção é ainda maior (da ordem de 14 vezes [2]), somos levados a concluir que a utilização de uma língua como o esperanto tem o efeito de nivelar (tanto quanto possível) a relação custo/ benefício de todas as culturas do planeta no que diz respeito ao acesso à informação: se na hipótese *Cyber-english* o esforço exigido das diversas comunidades da Web tinha aproximadamente os patamares da figura 6a, no caso da hipótese esperanto, os patamares de esforço teriam o perfil indicado na figura 6b. A observação dos gráficos dá uma idéia dos diferentes graus de exigência nas duas estratégias: no primeiro cenário a distribuição dos esforços é bastante desigual, e os valores absolutos são consideráveis, mas no segundo cenário há um equilíbrio na distribuição do esforço exigido, o qual, em todo caso, é sempre bastante razoável para qualquer um dos grupos lingüísticos. Vale ressaltar que no primeiro cenário, mesmo com todo investimento em horas de estudo, os Grupos 2, 3, 4 e 5 estão longe de ter um domínio do instrumento lingüístico equivalente ao do Grupo 1.

ii) A Segunda questão de ordem prática diz respeito ao número de pessoas capazes de fazer uso do esperanto. É uma pergunta relevante pois conforme visto, o inglês já dispõe de uma considerável base de usuários (322 milhões de falantes nativos [4]) além de um contingente de usuários como segunda língua, cujas estimativas variam de algumas dezenas a algumas centenas de milhões de indivíduos. O caso do esperanto é peculiar pois este não é uma língua étnica, e sua base de usuários



não pode ser estimada a partir das populações dos países (como no caso do inglês). É em todo caso provável que este número seja da ordem de 2 milhões de pessoas.

Embora esse número seja ínfimo comparado ao do inglês, vale lembrar que em termos relativos a diferença não é assim tão grande, uma vez que os percentuais de pessoas que não falam inglês e esperanto em nosso planeta são de 90% e 99,97%, respectivamente. Por esse ângulo, parece mais promissor esperar que uma língua cujo esforço de aprendizado seja em média dez vezes menor possa alcançar os 99,97% de usuários potenciais do que outra, 10 vezes mais exigente, possa atingir seu público latente (90%). Independentemente desses aspectos quantitativos, o fato é que a solução aqui chamada Esperanto exigiria, assim como o caso do uso do inglês, uma considerável ampliação de sua base de usuários, a fim de torná-lo uma solução efetiva para o acesso aos conteúdos da Web, os quais por sua vez teriam que ser espelhados por páginas em esperanto. Essa é uma lacuna da proposta, mas diferentemente das lacunas das soluções anteriormente tratadas (inglês, tradução automática, UNL e tradução humana) esta não parece ser intransponível a médio prazo. Na seção sobre propostas de estudo serão apresentadas algumas idéias sobre as maneiras de alcançar este objetivo, fazendo uso da própria Web.

iii) O Esperanto surgiu como uma língua planejada, mas desde sua publicação, há mais de um século, ele se tornou uma língua viva, animada por uma comunidade internacional que dele se serve para os mais diversos fins (ver Piron [2]). Mesmo no âmbito da Web) já existem milhares de páginas, listas de discussão e comunidades virtuais em e sobre o esperanto. Algumas páginas interessantes podem ser vistas em [31]. Por essa razão, a língua tem sido capaz de evoluir, e seu léxico acompanha as revoluções da ciência e da tecnologia (com um certo atraso, mas isto é também verdadeiro para todos os idiomas exceto o inglês, de onde os novos termos geralmente provêm). Assim sendo, para os propósitos aqui estabelecidos de encontrar um instrumento capaz de prover acessibilidade na Web para uma comunidade multicultural, não parece haver limitações intrínsecas na língua internacional auxiliar denominada Esperanto.

iv) Conforme visto na seção que tratou do padrão UNICODE, a exemplo de outras línguas, também o esperanto passou a ter seu caracteres totalmente suportados pelos recentes navegadores da Web. Com isso, as iniciativas de disponibilizar versões das páginas da Web em esperanto não teriam restrições de ordem técnica. Além disso, já existem máquinas de busca capazes de procurar documentos e páginas da Web a partir de palavras-chave em esperanto [23, 24].

v) Certamente existem outras línguas que se encaixam na definição de língua internacional auxiliar. Ao longo dos últimos séculos vários projetos foram iniciados, e alguns chegaram a ser publicados. Porém, só o esperanto atingiu a massa crítica de usuários capaz de perenizar o idioma. De resto, todas as tentativas de reformar o esperanto para torná-lo mais palatável aos gostos ocidentais resultaram em fracasso (ver [25]), o que reforça a tese de que como instrumento lingüístico voltado para a comunicação

internacional o esperanto representa um ponto ótimo (e estável) em uma curva que começa nas línguas filosóficas da Renascença e continua até os projetos de línguas naturais adaptadas do Século XX (e.g. Basic English)[25].

4. Uso Combinado das Soluções

Conforme discutido acima, uma das estratégias para contornar as deficiências da língua inglesa como língua-fonte em sistemas de tradução automática, é o uso de uma linguagem controlada em que o autor adota uma série de precauções para evitar que o texto produzido seja passível de interpretações conflitantes. Essa solução, contudo, exige do autor conhecimentos ainda mais refinados de seu instrumento lingüístico (inglês), na medida em que ela pressupõe uma capacidade de produzir um texto em inglês normal e além disso ser capaz de reformatá-lo para atender os preceitos da linguagem controlada. Trata-se então de um grau de domínio da língua que ultrapassa em muito o que o usuário não-anglófono da Web normalmente tem. Em outras palavras, a solução é ainda mais inacessível que o domínio do inglês padrão. Diante disso, a solução esperanto apresentaria uma característica positiva: em um certo sentido, escrever em esperanto é utilizar uma linguagem controlada (inconscientemente), visto que as estruturas da língua já são inerentemente claras (não ambíguas) e seus recursos gramaticais favorecem uma correta interpretação do texto. Esse fato, aliado à relativa facilidade do domínio ativo do idioma, sugere usos potenciais: o simples espelhamento em esperanto dos conteúdos das páginas Web nas diversas línguas possibilitaria aos tradutores automáticos fazer suas versões a partir de textos com elevado grau de traductibilidade. Esse fato, por si só, teria um efeito organizador no caos lingüístico da Web: em lugar de se desenvolverem ferramentas de tradução para incontáveis pares lingüísticos (lembrando que se existem no mundo 6000 línguas, existem dezenas de milhões de pares lingüísticos, e alguns milhares de pares se considerarmos as 50 principais línguas), os desenvolvedores de sistemas de tradução poderiam concentrar seus esforços no atendimento da demanda por traduções do esperanto para as diversas línguas étnicas (tantas quanto possível).

Seria então muito oportuno e bem-vindo o desenvolvimento dessas ferramentas juntamente com o surgimento de uma ética lingüística na Internet, com base na qual todo conteúdo de relevância intercultural fosse sempre acompanhado de uma versão em esperanto. Assim, mesmo os usuários não proficientes na língua internacional auxiliar poderiam se beneficiar de traduções de nível aceitável (embora não perfeitas) em suas línguas maternas. Elas poderiam ter assim uma participação, ainda que passiva, no fluxo de idéias e informações da rede mundial. Lembrando que uma participação ativa estaria ao alcance das mãos: bastaria um (aceitável) esforço no aprendizado da língua internacional auxiliar para poder produzir textos que teriam livre trânsito na Rede ou para ser capaz de participar de discussões e bate-papos (que já existem mas que certamente veriam seu público se ampliar exponencialmente).

Outro benefício do espelhamento em esperanto de conteúdos da Web seria a possibilidade de utilização de ferramentas de busca muito mais eficientes que aquelas feitas nos textos em línguas étnicas [26], visto que a estruturação do esperanto possibilita uma inferência mais confiável dos assuntos e dos contextos de cada documento.

5. Conclusão

As vantagens inerentes ao uso do esperanto como solução para aumentar a acessibilidade dos conteúdos da Web, conforme visto, decorrem da facilidade de aprendizado deste idioma por pessoas de todas as culturas e de diferentes classes sociais. No entanto, o fator preponderante no sucesso de qualquer aprendizado de línguas é a motivação. É provável que muitos dos potenciais beneficiários do uso de uma língua internacional auxiliar na Web não se interessem o suficiente pelo assunto para se prontificarem a fazer a sua cota de esforço (dedicar uma certa quantidade de horas para aprendê-la), e prefiram esperar que alguma solução milagrosa, como uma máquina de tradução perfeita, caia do céu. Entretanto, como visto ao longo do texto, tal ferramenta provavelmente nunca existirá e os inconvenientes da solução MT persistirão, sobretudo na medida em que os conteúdos a serem traduzidos estiverem em inglês.

Para outros, a solução *Cyber-english* pode continuar sendo a mais atraente, na medida em que ela reflete uma visão “meritocrática” de mundo na qual se considera natural o privilégio de uns em detrimento da maioria. Assim, essas pessoas não se incomodarão com as eventuais situações de discriminação ou injustiça, desde que pelo menos consigam ter uma participação passiva no fluxo de idéias e informações da Web.

Ainda assim, a utilização do esperanto como instrumento para aumentar a acessibilidade dos conteúdos da Web parece viável, e poderia beneficiar mesmo aquelas pessoas que não se dispusessem a aprendê-lo, pois, como sugerido na seção precedente, sua simples adoção permitiria organizar os intercâmbios culturais na Internet.

Assim, é inegável que com base em seu desempenho em situações interculturais [2], o esperanto tem uma contribuição a dar à solução do problema da limitação lingüística na Web.

6. Propostas de estudo

Como visto acima, a solução esperanto apresenta uma importante lacuna no que se refere à efetiva promoção de acessibilidade lingüística aos conteúdos da Web: sua base de usuários, embora grande para uma língua planejada, é pequena se considerarmos que nos próximos anos o número de usuários da Web deve ultrapassar a cifra de 1 bilhão. Esse fato, em lugar de ser visto como um empecilho, pode ser encarado como um interessante e desafiador campo de estudo: como fazer uso da própria Web para difundir e massificar a língua internacional auxiliar, a fim de esta possa se tornar um

instrumento lingüístico eficaz e beneficiar os usuários da Web? Poderia a Web se tornar um fator decisivo para a formação da massa crítica capaz de fazer do esperanto uma realidade nos intercâmbios intelectuais e culturais da sociedade na era da Internet (visto que o idioma já possui atributos que o tornam particularmente adaptado para sua finalidade)?

Segundo Piron [2], o desbloqueio da situação poderia ser conseguido com pequenas alterações nos sistemas de ensino dos diversos países. Essas alterações, embora mínimas do ponto de vista curricular, exigiriam uma ação coordenada dos diversos governos (como já conseguido na área de saúde com a eliminação de doenças por meio de esforços mundiais de vacinação). Entretanto, esse esforço coordenado não parece muito fácil de se conseguir a curto prazo, visto que as nações dominantes desfrutam de vantagens por conta da atual ordem lingüística.

Por outro lado, a Internet fornece um campo aberto para iniciativas de ensino a distância, não somente as vinculadas aos sistemas educacionais instituídos. Nesse contexto, iniciativas de ensino do esperanto através da rede são muito promissoras e algumas já existem desde há alguns anos. No entanto, até o presente, as estratégias adotadas têm se baseado no paradigma instrucionista. Por essa abordagem, a informação é tida como a unidade fundamental do ensino, sendo os esforços voltados para processos de aquisição, armazenamento, representação e transmissão da informação ao aprendiz [27]. Embora esse paradigma tenha uma longa tradição nas atividades de ensino, as atuais iniciativas em geral não tiram proveito de todo o potencial dos ambientes de ensino baseados no computador. Assim, é oportuno avaliar o que outras abordagens trariam de contribuição para o problema específico aqui proposto: fazer uso da Web para difundir o uso de uma língua internacional como o esperanto.

Considerando a natureza da estruturação do esperanto a qual, como anteriormente visto, se baseia no princípio da assimilação generalizadora, descrito por Piaget, é razoável supor que o paradigma construtivista, por ele proposto, seja de fato adequado a essa matéria. Nessa abordagem, e conforme descrito em [27], o aprendizado é entendido como sendo a construção individual do conhecimento a partir de atividades de exploração, investigação e descoberta, em ambientes interativos. O fato de a língua se basear em esquemas de validade geral sugere que uma interface interativa bem concebida poderia oferecer um campo satisfatório e rico aos estudantes, os quais se valeriam das informações disponíveis para construir novos conhecimentos sem se sujeitarem à freqüente experiência do erro [2] como em geral ocorre no aprendizado das línguas étnicas.

Como a língua esperanto já conta com mais de cem anos de vida cultural, em que os mais diversos conteúdos foram produzidos, de poesia a música, enigmas, quadrinhos, etc., esse material poderia ser selecionado para compor essa interface interativa. Restariam, entretanto, algumas questões em relação a tal

abordagem: como delinear o tipo de conhecimento que deve ser construído e como avaliar tais construções [27] ?

Um paradigma alternativo já proposto é o Aprendizado Socialmente Distribuído, o qual se vale das novas possibilidades surgidas com a Internet e com a globalização da informação dentro de uma perspectiva colaborativa [28] em que recursos como áudio, vídeo e texto, em interações síncronas (bate papo, conferência, etc.) ou assíncrona (correio eletrônico, listas de discussão, etc.) seriam usados de forma compartilhada com outras pessoas. Esse conceito é bastante apropriado ao ensino do esperanto pois esse deve ter um forte componente colaborativo na medida em que o aprendizado da língua tem como fim último a comunicação, a troca de idéias e de conhecimentos, e pode com isso se beneficiar das facilidades oferecidas pela Web. A comunicação por correio eletrônico pode levar o aluno a exercitar seus conhecimentos através da troca de mensagens com outros estudantes (distantes geográfica e culturalmente), o que teria, além do mais, um efeito motivador. Assim, esse paradigma parece bastante adequado para a interface aqui proposta. A respeito de iniciativas de colaboração educacional ver também [30].

A concepção de um ambiente de ensino baseado na Web que faça um uso inteligente dos paradigmas de ensino é então um desafio. Recentemente, tem-se buscado na teoria Semiótica um referencial teórico para a análise e *design* de tais ambientes, explorando as características da rede mundial: um ambiente dinâmico, interativo, repleto de informações, sem restrição de espaço e tempo para a comunicação [27].

No caso específico aqui tratado, ensino do esperanto por meio da Web, o uso de um referencial semiótico é promissor. Segundo os três modos de inferência definidos por Peirce: abdução, dedução e indução, a semiosis pode ser entendida como um processo de se aplicar signos para entender algum fenômeno (indução), raciocinar de signo para signo (dedução), e/ou usar signos para fazer sentido de alguma nova experiência (abdução) [29], ou, segundo Gudwin [38], representam, respectivamente, seleção, extração e geração de conhecimentos. No ensino do esperanto, as situações em que a abdução se manifesta são muito freqüentes por conta da regularidade dos esquemas: ao ser confrontado com um texto, o estudante, por exemplo, se dá conta de que a cada vez que encontra uma palavra terminada em *-o*, ela representa um substantivo, e todas aquelas terminadas em *-a* representam adjetivos. Por força dessas repetidas experiências ele abduz a idéia (verdadeira) de que os substantivos se formam com a terminação *-o* e os adjetivos com *-a*. Como no mecanismo da abdução a hipótese produzida pode não ser verdadeira, o fato de a língua ter uma estrutura regular aumenta a taxa de hipóteses corretas. No aprendizado de uma língua que não é regida por esquemas regulares as abduções levam freqüentemente ao erro.

O mecanismo da indução, por sua vez, se manifesta quando com base em idéias o estudante induz experiências válidas. Isso se dá, por exemplo, quando ele

aprende que os substantivos coletivos são sempre formados com auxílio do sufixo

-ar, e assim ele extrapola a idéia para todos os casos, sempre chegando a um resultado correto: se *stelaro* é o coletivo de *stelo* (estrela), e *bovaro* é o coletivo de *bovo* (boi), ele induz que o coletivo de *abelo* (abelha) é *abelaro*.

Finalmente, o processo de dedução ocorre quando a partir de uma idéia o estudante é capaz de extrair outra idéia que estava “empacotada” naquela primeira: esse é o caso quando o estudante aprende que *kio*, *nenio* e *tio* significam *o que*, *nada* e *isso*, e que *kie*, *nenie* e *tie* significam *onde*, *em nenhum lugar* e *lá*, ao se deparar com o correlativo *kiam* significando *quando*, ele deduzirá que *nunca* e *então* são, respectivamente, *neniam* e *tiam*.

Em outras palavras, o aprendizado do esperanto pode tirar proveito de todas as modalidades de inferência, e assim o *design* da interface de ensino seria muito facilitado com o referencial semiótico.

Mas há outros aspectos que tornam o ensino do esperanto na Web particularmente interessante. De início está o fato de que o público alvo (latente) de tal curso seria totalmente multicultural, pois todo ser humano é em princípio usuário potencial de um idioma internacional. Isso equivale a dizer que uma interface de ensino para uso mundial não poderia se valer de explicações ou orientações em línguas étnicas, a fim de que o mesmo curso fosse acessível a qualquer pessoa, não importando a sua cultura de origem. Isso nos levaria a explorar em profundidade conceitos de metáforas em interfaces, índices, ícones e desenhos, a fim de que o aprendizado não fosse limitado pela língua do estudante, pois isso representaria uma contradição em relação à proposta de uso do esperanto para democratização do acesso à Web. Já existem métodos (impressos) de ensino de esperanto baseados neste princípio: através do uso de figuras o estudante é capaz de descobrir o significado das palavras sem necessidade de tradução em sua língua materna. Sua transposição para a Web seria relativamente simples.

Outra consideração a respeito da interface diz respeito à modalidade de tutoria a ser utilizada. Para os propósitos de rápida ampliação da base de usuários do esperanto, seria fundamental superar o principal gargalo dos cursos desse língua na Web: o número de alunos atendidos é limitado pela disponibilidade de formadores. Isso significa que uma demanda exponencialmente crescente não poderia ser atendida pelas abordagens tradicionais. Uma alternativa a ser estudada seria o uso de dois níveis de tutoria: o primeiro baseado em agentes inteligentes, capazes de realizar tarefas de correção de exercícios e avaliação da evolução dos estudantes com base em taxas de erros, riqueza do vocabulário utilizado, etc. Os agentes inteligentes poderiam, além disso, organizar os aprendizes, encontrando afinidades a serem exploradas em termos colaborativos (correio eletrônico, etc.). Em um nível mais alto haveria tutores humanos, cuja função seria atender todas as demandas que os agentes inteligentes não fossem capazes de tratar.

O aprendizado do esperanto como meio

Numa perspectiva mais ampla, o aprendizado do esperanto não seria então um fim em si mesmo, mas um pré-requisito para a organização de uma rede de ensino colaborativo na qual a disponibilidade de um meio eficaz de comunicação intercultural possibilitaria o intercâmbio de experiências e informações entre estudantes de todo o mundo. Isso valeria igualmente para cientistas. Segundo E. Forte, coordenador do projeto europeu ARIADNE, cujo objetivo é desenvolver ferramentas que contribuam para um padrão europeu de estratégias de treinamento transculturais e colaborativas baseadas em IT, a escolha de uma língua comum para a comunicação intercultural seria, em alguns aspectos, um problema de seleção de padrões [34], e a solução esperanto “pode ter se tornado mais viável hoje graças à rápida difusão das ICTs (tecnologias de informação e comunicação) e ao inevitável advento da *cooperação* como valor e método para a transmissão otimizada e em larga escala de conhecimentos, i.e. educação” [35].

Assim, o desenvolvimento de uma interface de ensino a distância tal como discutida acima seria um desafio viável para as recentes teorias desse campo da computação, e os benefícios dela advindos, no que diz respeito à promoção da acessibilidade lingüística da Web, seriam inestimáveis.

7. Agradecimentos

Este trabalho não teria sido possível sem as seguintes pessoas:

Maria Cecília C. Baranauskas pelo estímulo e orientação; Moisés Braga Ribeiro e Rodrigo Túlio Costa Rocha pela revisão do texto; Amanda Meinke Melo e Moisés Braga Ribeiro por seus trabalhos usados como referência; e todos os colegas pelas esclarecedoras discussões em sala.

8. Referências

- [1] Piron, Claude, “Language Constraints and Human Rights”, Anniversary Symposium on Language and Human Rights, Nações Unidas, Genebra, 28 de abril de 1998, http://www.geocities.com/c_piron
- [2] Piron, Claude, “Le Défi des Langues”, L’Harmattan, Paris, 1994
“O Desafio das Línguas”, Editora Pontes, Campinas, 2001.
- [3] Berners-Lee, Tim, “World Wide Web, The World Wide Web: Past, Present and Future”, MIT, 1996
<http://www.w3.org/People/Berners-Lee/1996/ppf.htm>
- [4] Ethnologue: Languages of the World, <http://www.sil.org/ethnologue>
- [5] Direitos Lingüísticos
<http://www.unesco.org/most/>
- [6] Global Reach, “Global Internet Statistics (by Language)”, abril de 2001, <http://www.greach.com/globstats/>

- [7] NUA Internet Surveys, “Majority of Users Will Be Non-English Speakers”, 10 de junho de 1999, <http://www.nua.ie/surveys/>
- [8] Gupta, Anthea F., “The Internet & the English Language”, First Conference on Postcolonial Theory, University of Leeds, 1997, <http://www.faz.nus.edu.sg/staff/conf/poco>
- [9] Lebert, Marie, “Le Multilinguisme sur le Web”, CEVEIL, 1999, <http://www.ceveil.qc.ca/multifr2.htm>
- [10] Piron, Claude, “Learning from the translation mistakes”, Budapeste, 1988, http://www.geocities.com/c_piron
- [11] Lockard, Joe, “Resisting Cyber-English”, Universidade de Berkeley, 1996, <http://eserver.org/bs/24/lockard>
- [12] Shade, Leslie R., “www.non-digerati.com: who are you?”, Canadian Cultural Research Network Colloquium, University of Ottawa, 1998, http://www.arts.uwaterloo.ca/ccrn/documentes/colloq98_shade.htm
- [13] Brink, Collin, “US Hispanic Internet Users: Thinking in Spanish and Surfing in English”, eMarketer, 27 de fevereiro de 2001, http://www.emarketer.com/analysis/edemographics/2001_0227_edemo.htm
- [14] Habeck, Joachim O., “The Existing and Potential Role of the Internet for Indigenous Communities in the Russian Federation”, Symposium “Bicultural Education in the North: Ways of Preserving and Enhancing Indigenous Peoples’ Languages and Traditional Knowledge”, Franckesche Stifungen in Halle (Alemanha), 13 a 15 de dezembro de 1997, <http://www.spri.cam.ac.uk/people/jeoh2/2halle.htm>
- [15] Esselink, Bert, “A Practical Guide to Localization”, John Benjamins, Amsterdam / Philadelphia, 2000
- [16] Martins, Ronaldo T. et al., “An inter-lingua aiming at communication on the Web: How language-independent can it be?”, NILC São Carlos, 1998, <http://nilc.icmsc.sc.usp.br>
- [17] Martins, Ronaldo T., “Dos problemas da ambigüidade semântica em um modelo de tradução automática baseado em interlíngua: Apontamentos do Projeto UNL-Brasil”, NILC / Universidade São Francisco, 1999
- [18] Hopkins, Robert, “Website Translations”, 1997, <http://www.wdvl.com/Internet/Web/Site/Translation>
- [19] Eco, Umberto, “La ricerca della lingua perfetta nella cultura europea”, trad. Esperanto, Edistudio, 1994
- [20] Saussure, René, “La Vort-strukturo en Esperanto”, Laborgrupo “Antauen”, 1985, original de 1916.
- [21] Norman, D.A. “Human Information Processing”, P.H. Lindsay ; Academic Press, 1972
- [22] Ávila, Ismael M. A., “Um framework semiótico para SI”, Unicamp, Seminário de 21 de maio de 2001, <http://www.dcc.unicamp.br/~cpg/material-didatico/mo622/200101/Agenda.htm>
- [23] Meta Monster, <http://www.metamonster.com/esperanto/>
- [24] Euroseek,

<http://www.euroseek.com/page?ilang=eo>

[25] Harlow, Donald J., “How to build a Language”, 2000,

<http://www.best.com/~donh/Esperanto/EBook/chap03.html>

[26] Ribeiro, Moisés B., “Aplicações Computacionais Relacionadas à Gramática do Esperanto” Unicamp, Julho de 2001, <http://www.aleph.com.br/kce/oquee.htm>

[27] Melo, Amanda M., “Semiótica e Ambientes Educacionais Baseados na Web”, Unicamp, Julho de 2001,

<http://www.dcc.unicamp.br/~ra007250/mo622/artigo/index.html>

[28] Baranauskas, M. C. et al., “Uma Taxonomia para Ambientes de Aprendizado Baseados no Computador. O Computador na Sociedade do Conhecimento”, NIED-Unicamp, 1999

[29] Cunningham, D. J., “Cognition as Semiosis: The Role of Inference”, 1998,

<http://www.indiana.edu/~educp550/steps.html>

[30] Escola Virtual Tibor Sekelj,

<http://lps.uniroma3.it/kler/>

[31] Páginas Amarelas do Esperanto,

<http://www.esperanto.net/veb/flavaj-pagxoj.html>

[32] “Países pobres terão acesso a periódicos de Medicina pela Internet”, Agência Estado, 09 de julho de 2001,

<http://www.estadao.com.br/tecnologia/internet/2001/jul/09/213.htm>

[33] CONIP VII Congresso de Informática Pública, “Governo Eletrônico e Inclusão Digital: o futuro da cidadania na era da informática”, São Paulo, 5 a 7 de julho de 2001,

<http://www.conip.com.br>

[34] Forte, Eddy N., “Standards for ICT-Based Education: Facts, Promises and Pitfalls”, in Proc. GMW’2000 Conference, Insbrück, setembro de 2000.

[35] Forte, Eddy N., “Towards a Humanistic Solution to the Language Problem in IT-Based Cooperative Education and Beyond”, Ecole Polytechnique Fédérale de Lausanne, setembro de 2000,

<http://prometeus.deis.unibo.it/Prometeus/Journal00/Index.htm>

[36] Veiga Filho, João Pimenta da, “A Universalização da Informação”, Revista Ciência da Informação, Brasília, v. 30, n.1, p. 7-12, jan./abr. 2001,

<http://www.ibict.br>

[37] Marchiori, Patrícia Z., “Ciberteca ou biblioteca virtual: uma perspectiva de gerenciamento de recursos de informação”, Ciência da Informação v.26, n.2, p.115-124, mai./ago. 1997,

<http://www.ibict.br>

[38] Gudwin, Ricardo R.; Notas de aula, Unicamp, 2001.