

Projeto TLIBRAS digital – tradutor português x LIBRAS

Guilherme de Azambuja Lira¹

Introdução

Existem no Brasil 24 milhões de pessoas portadoras de deficiência, o que significa cerca de 14% da população. Entre elas, 5,7 milhões são pessoas com deficiência auditiva (Censo – IBGE, 2000).a Essas pessoas encontram-se excluídas de diversas formas, de várias dimensões da vida social e produtiva.

A utilização das novas tecnologias de informação e comunicação, potencializada por recursos de multimídia, tem papel relevante como fator de inclusão social dessa população.

As Nações Unidas, em especial a Comunidade Européia, têm incluído em seus respectivos documentos a igualdade de oportunidades para pessoas portadoras de deficiência e os idosos. O Brasil adequou-se a essas políticas, por meio da adoção da Lei nº 10.098 de 19/12/2000b, sobre a acessibilidade, regulamentada pelo decreto nº 5.296, de 02/12/2004.

No capítulo referente à comunicação, a referida Lei,c preconiza o uso de legendas, isto é, textos escritos em correspondência às informações sonoras produzidas, principalmente para programas de televisão e para teatros, visando permitir acesso diferenciado por deficientes sensoriais auditivos.

No caso de deficientes auditivos brasileiros, essa tecnologia se revela inacessível no momento. Isto ocorre porque a maioria dos deficientes auditivos brasileiros tem grande dificuldade de ler e entender a língua escrita. Não só em decorrência da quase inexistência de projetos ligados à capacitação do surdo, mas também de dificuldades na aquisição de uma segunda língua, pois sua primeira é a Língua Brasileira de Sinais (LIBRAS).d

Esse reduzido domínio da língua portuguesa por parte da comunidade surda é derivado de implicações lingüísticas e de cunho estrutural da escola brasileira.

Segundo PERLINE, a primeira implicação lingüística está relacionada com o fato de 95% dos alunos surdos serem filhos de pais ouvintes que não dominam a LIBRAS. Portanto, a escola para os surdos se torna mais fundamental ainda, pois precisa garantir que a língua de sinais servirá de instrumento comunicativo para o desenvolvimento educacional. Devemos salientar que a aquisição de uma língua de forma natural e espontânea é a chave para a aquisição da linguagem.

“As políticas educativas deverão levar em conta as diferenças individuais e as diversas situações. Deve ser levada em consideração, por exemplo, a importância dos sinais como meio de comunicação para os

¹ *Presidente da Acessibilidade Brasil. Mestre em Educação pela UNESA.*

surdos, e ser assegurado a todos os surdos acesso ao ensino da linguagem dos sinais de seu país. Face às necessidades específicas de comunicação de surdos e surdos-cegos, seria mais conveniente que a educação lhes fosse ministrada em escolas especiais ou em classes ou unidades especiais em escolas comuns.” (Declaração de Salamanca – 1994).f

A estrutura da própria instituição educacional ainda é uma grande barreira, pois, apesar de existirem políticas públicas educacionais avançadas, as escolas regulares, em esmagadora maioria, carecem de recursos físicos e financeiros, e principalmente humanos (professores especializados e intérpretes), para a inclusão do aluno surdo na sala de aula.

A Lei nº 10.436, g decretada e sancionada em 24 de abril de 2002, dispõe, em seu artigo 4º: “O sistema educacional federal e sistemas educacionais estaduais, municipais e do Distrito Federal devem garantir a inclusão nos cursos de formação de Educação Especial, de Fonoaudiologia e de Magistério, em seus níveis, médio e superior, do ensino da Língua Brasileira de Sinais - Libras, como parte integrante dos Parâmetros Curriculares Nacionais - PCNs, conforme legislação vigente.”

Com uma população surda cerca de 60% maior que a nossa, os EUA têm registrados cerca de 10.000 intérpretes (fonte: RID – Registry of Interpreters for the Deaf – 2002 – <http://www.rid.org>) contra 400 intérpretes registrados no Brasil.h

Pelo Censo Escolar de 2000 (fonte INEP – MEC), i 80% dos alunos surdos ingressos na escola não completam o ensino fundamental e só 3% completam o ensino médio.

A pesquisa americana citada por Duffy, que constatou que a média de leitura dos alunos surdos com o ensino médio completo corresponde à quinta-série do ensino fundamental, explica os resultados da pesquisa (LIRA), j pela qual 50% de pessoas surdas, com ensino médio completo, não conseguiram entender o conteúdo das informações disponibilizadas em português, via legendas ou *closed caption*.k

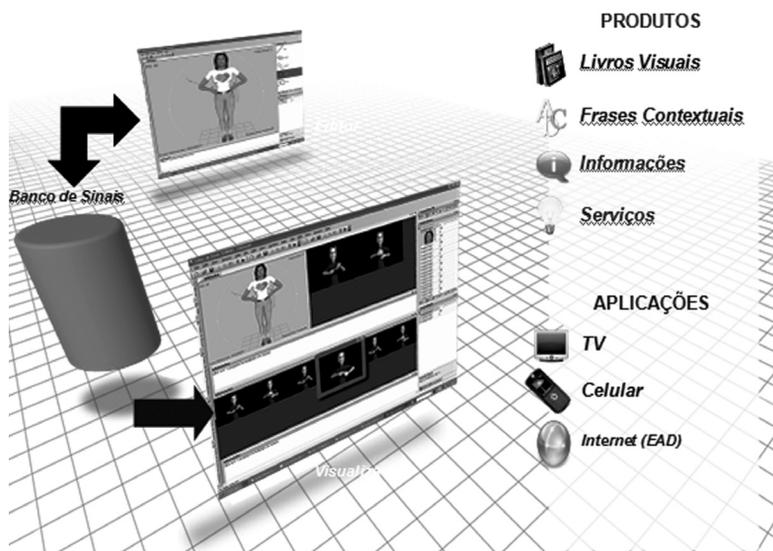
As legendas automatizadas em português, em substituição aos textos sonoros produzidos pelos meios de comunicação via *closed caption* ou mesmo legendas de melhor qualidade, produzidas para vídeos ou para filmes, não têm atendido de forma satisfatória à grande maioria da comunidade surda brasileira, que usa a LIBRAS como sua primeira língua: além do reduzido número de surdos que adquiriram a língua portuguesa, tanto as empresas de comunicação quanto os fabricantes de aparelhos de TV não têm se preocupado em tornar disponível, em larga escala, essa tecnologia.

O projeto TLIBRAS digital - o tradutor português x libras

O TLIBRAS foi iniciado em 2001 como projeto executivo, sendo desenvolvido de 2002 a 2004 com o apoio financeiro do Ministério da Educação do Governo Brasileiro, através de sua Secretaria de Educação Especial. O objetivo é a construção de um tradutor informatizado da língua portuguesa para a Língua Brasileira de Sinais - LIBRAS, com a seguinte abrangência: ser utilizado em sala de aula; em áreas públicas, pela futura televisão digital (concomitantemente ou em substituição aos textos legendados); em vídeos; pela internet; em dispositivos digitais (celulares, PDAS); na construção de livros visuais, traduzindo informações por meio de sinais animados, apresentados via computador.

Em 2005, o projeto TLIBRAS teve sua estratégia de desenvolvimento redimensionada, para as quatro fases abaixo descritas:

A primeira fase é o desenvolvimento do editor e visualizador de Sinais em LIBRAS (Língua Brasileira de Sinais).



O desenvolvimento do Editor e visualizador de Sinais em LIBRAS (Língua Brasileira de Sinais) foi dividido em dois módulos:

O primeiro módulo é um editor gráfico de sinais, apresentado no monitor do computador e apoiado por um personagem gráfico articulado (AVATAR). Ele permite que um especialista em Sinais em LIBRAS (uma pessoa surda ou um intérprete) possa simular o movimento dos sinais em LIBRAS, correspondentes a palavras em português.

Essa simulação será feita combinando a(s) configuração(ões) de mão de um sinal específico, sua trajetória, o tipo de movimento, a expressão facial, dentro de uma linha do tempo, isto é, possibilitando que com uma ou mais configuração(ões) de mão escolhida, seja na mão direita ou esquerda ou nas duas, combinadas com trajetórias de movimento, frequências e expressões faciais, cada uma em um espaço de tempo determinado, possa apresentar em LIBRAS uma palavra correspondente em português.

Sendo correta, essa combinação de movimentos que representam um sinal em LIBRAS poderá ser apresentada e armazenada em um Banco de Dados de *Scripts* Gráficos de Sinais - BDSGS, para posterior utilização e visualização.

O editor conta, além de um visualizador de sinais produzidos, com uma biblioteca auxiliar de objetos – para reaproveitamento – composta de movimentos, trajetórias e expressões faciais já utilizadas por outros sinais armazenados.

O segundo módulo é um visualizador de frases em LIBRAS.

A partir de uma frase apresentada, o programa verifica a existência de sinais correspondentes às palavras, componentes da frase, no BDSGS. Caso existam, o programa combinará sinais em LIBRAS e os apresentará de forma contínua e combinada (*Blended*), impedindo uma solução de continuidade na apresentação da frase em LIBRAS correspondente.

Caso um sinal ou sinais, dessa frase a ser representada, não existam no BDSGS, o visualizador se encarregará de apresentar a palavra ausente utilizando a datilologia, isto é, o sinal soletrado (sinais representando as letras do alfabeto, em LIBRAS), com a continuidade de movimentos entre os Sinais em LIBRAS existentes e os apresentados via datilologia.

Observamos que a frase a ser submetida ao editor necessitará, nesse primeiro momento, de uma adequação humana, entre as gramáticas da língua portuguesa e da LIBRAS, para que as frases em LIBRAS apresentadas pelo visualizador mantenham as regras morfológicas, semânticas e sintáticas exigidas pela LIBRAS, evitando assim o português sinalizado, incompreensível para a maioria das pessoas surdas.

As metas previstas nessa fase são cadastrar e possibilitar a visualização de *scripts* de sinais gráficos povoados no BDSGS, possibilitando inicialmente a tradução de frases para a comunicação, via telefonia celular, educação a distância, via internet, e para a produção de livros didáticos visuais.

A segunda fase do projeto é o desenvolvimento de rede colaborativa para a produção de sinais em LIBRAS.

O objetivo dessa fase é estruturar uma rede colaborativa, via internet, centralizada na Acessibilidade Brasil, formada por Universidades e Escolas, Centros de Pesquisa, Especialistas, Federações de Surdos, Entidades de Pessoas Surdas, o mercado produtor de conteúdos em LIBRAS e a comunidade surda em geral, que atuarão, fundamentalmente, como alimentadores de novos sinais gráficos em LIBRAS.

Com a cessão para o uso do editor e visualizador gráfico de sinais, construído na primeira fase, os integrantes da rede colaborativa poderão contar com um novo ferramental para a criação de seus conteúdos em LIBRAS, visando atender a seus contextos específicos, como no caso do MEC, para a produção de livros didáticos visuais, principalmente aqueles indicados pelo PNLD – Programa Nacional do Livro Didático ou para a produção de textos, informações e livros para-didáticos e técnicos por entidades de pessoas surdas, e centros de pesquisa, entre outros afins.

Esse imenso número de sinais em LIBRAS, produzidos para contextos diferentes, pelos integrantes da rede colaborativa, permitirá o povoamento de um grande banco de dados, com sinais específicos para contextos determinados, possibilitando um grande e rápido avanço, na formação de uma *nominata* brasileira de sinais gráficos em LIBRAS, além da consolidação das regras morfológicas, semânticas e sintáticas da LIBRAS, ambos imprescindíveis para as traduções automatizadas do português para a LIBRAS.

Os contextos poderão ter origem, por exemplo, no atendimento de um posto de saúde, na escola, num banco, na geração de material didático/técnico visual como alternativa ao português textual, ou em sítios da internet ou mesmo na comunicação celular, como formato alternativo a informações e serviços disponibilizados em português

Esse grande volume de sinais produzidos pelos integrantes da rede será avaliado por uma comissão de especialistas, formada por integrantes da comunidade surda,

lingüistas especializados e entidades representativas de pessoas surdas, impedindo o cadastramento de sinais inadequados ou inexistentes para a comunidade surda.

As metas previstas para essa fase são:

Estruturar rede colaborativa, via internet, apoiada por rede neural, com as seguintes especificidades:

a) Integrar colaboradores através do Sítio “Tradutor LIBRAS”, tais como universidades, centros de pesquisa, escolas, especialistas, entidades de pessoas surdas e da comunidade em geral.

b) Criar e manter uma comissão permanente de avaliação dos Sinais em LIBRAS produzidos pela rede colaborativa.

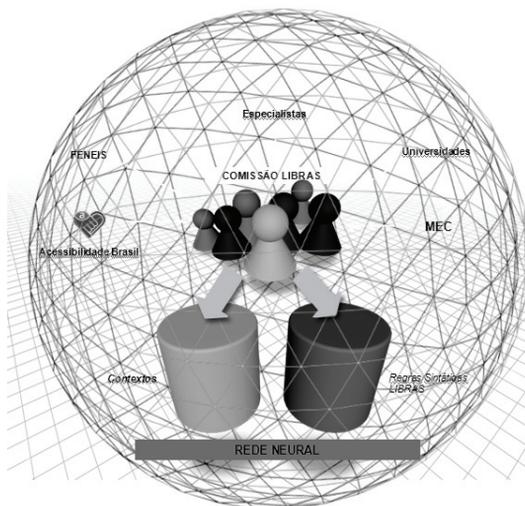
c) Produzir tecnologias que permitam:

- A comparação automática entre sinais existentes e os sinais candidatos e o aprendizado de frases e estruturas contextuais.
- A geração de tabelas de dados para os sinais, *scripts* de sinais, partes de sinais, estruturas de sinais em contextos específicos e isolados.
- O cadastramento de regras semânticas e sintáticas para contextos diversos.

A terceira fase do projeto é a estruturação da ferramenta para a geração de notação em LIBRAS utilizando a *Universal Networking Language* - UNL. Para a estruturação da ferramenta para a geração de notação em LIBRAS optou-se pela UNL tendo em vista:

- a) a experiência existente no mercado com este modelo de formalização;
- b) o seu caráter efetivamente plurilingüístico, dado que a iniciativa envolve grupos lingüísticos muito mais variados do que as outras abordagens disponíveis;
- c) o seu caráter público, na medida em que as patentes associadas pertenceriam à ONU;
- d) a sua abrangência, já que a flexibilidade da representação (notação LIBRAS) permitiria à UNL contemplar as informações de natureza visual, frequentemente expurgadas dos outros modelos, de base estritamente oral-auditiva.

O objetivo da notação LIBRAS é representar, de forma não ambígua, todas as informações necessárias (e apenas elas) para a geração de sinais de LIBRAS, a partir de



uma sentença em língua portuguesa, considerando, particularmente:

- a) que Português e LIBRAS são línguas completamente distintas, não apenas em relação à estrutura (gramática e léxico), mas ao próprio suporte;
- b) que a sentença em língua portuguesa possui redundâncias (como a concordância de número, gênero e pessoa gramatical) que seriam suprimidas no processo de representação em LIBRAS;
- c) que a sentença em língua portuguesa possui informações – como a ordem dos constituintes – que nem sempre são relevantes para a representação em LIBRAS;
- d) que a sentença em língua portuguesa possui lacunas (informação sobre classificadores, elipses, anáforas e outras pró-formas, nominais e verbais) que deverão ser preenchidas, ainda que de forma arbitrária, para que se possa assegurar a gramaticalidade e a semânticidade dos enunciados de libras.

Para a análise semântica do português, utilizou-se uma linguagem de representação do conhecimento (a UNL), que operou no protótipo como interlíngua, para a qual era convertida a sentença em língua portuguesa, e da qual era gerada a representação linearizada da LIBRAS (notação-LIBRAS).

A representação UNL ofereceu a perspectiva de desenvolvimento de um sistema de tradução baseado, principalmente, em informações de natureza semântica, em detrimento das estruturas sintáticas, que são diferentes entre o português e a LIBRAS.

Optamos pela estratégia de acumulação através da incorporação de informações no *corpus* através dos contextos iniciais, de modo que o modelo proposto terá sua abrangência gradativamente ampliada, para o tratamento de fenômenos, à medida que o trabalho da rede colaborativa se amplie.

As metas previstas nessa fase são:

- a) Desenvolvimento de banco de contextos para a alimentação acumulada visando ao desenvolvimento de corpus necessário para traduções;
- b) Implantação da rotina de tradução indireta, via UNL, visando à geração da notação LIBRAS.

A quarta fase – fase de integração – é a utilização automática do tradutor do português textual para sinais gráficos animados em LIBRAS e sua integração com a tecnologia de reconhecimento de voz.

A integração das diversas ferramentas e metodologias desenvolvidas nas fases anteriores do projeto (1, 2 e 3) possibilitará que textos escritos em português possam ser automaticamente traduzidos para animações gráficas em LIBRAS, bastando que, para isso, o contexto já seja conhecido e que as palavras utilizadas pelo texto já existam no banco de dados de sinais gráficos.

Caso contrário, será necessário o auxílio humano, visando resolver alguns fenômenos ainda não tratados por traduções anteriores.

A eficiência e eficácia do processo dependerá, efetivamente, da quantidade e qualidade das traduções realizadas pela rede colaborativa e na consolidação do conjunto de regras semânticas e sintáticas da LIBRAS, necessárias para a qualidade das traduções.

Nessa fase, ainda, será integrada uma tecnologia de reconhecimento de voz que permita a conversão de textos sonoros em português para textos escritos em português, possibilitando que o processo de tradução se inicie a partir de textos sonoros em português, limitado também a contextos e regras ainda não formalizados.

As metas previstas são:

- a) Integração das ferramentas e metodologias desenvolvidas nas fases 1, 2 e 3;
- b) Ampliação dos bancos de contexto e de sinais gráficos em LIBRAS;
- c) Integração com ferramenta(s) de reconhecimento de voz.

Conclusão

Apesar de ser de grande utilidade para a comunidade de 1 milhão de surdos brasileiros, o projeto TLIBRAS teve seu desenvolvimento descontinuado em 2004, por falta de recursos financeiros.

Em 2005, através de uma parceria com o instituto VIAS, de Florianópolis, com recursos oriundos do Ministério das Comunicações, a sua primeira fase pôde ser reiniciada, já com o desenho atual. Entretanto, nesse mesmo ano, por contingenciamento de recursos, o projeto novamente foi interrompido.

Espera-se que, no ano de 2007, a sociedade como um todo (governo, empresas e terceiro setor) entenda a importância desse projeto e o apóie, pois ele, implantado, será de fundamental para a inclusão da comunidade surda, nas novas tecnologias de comunicação e informação, imprescindíveis para a melhoria da qualidade de vida dessa grande parcela da população.

a IBGE. **Censo demográfico 2000 : resultados preliminares.** Rio de Janeiro, 2000. 156 p. Tab. Graf. Mapas. Inclui 1 CD-ROM.

b BRASIL. *Leis, Decretos. Lei nº 10.098 de 19/12/2000, Diário Oficial da União, Brasília, 20, dez., 2000. p. 2. Estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida, e dá outras providências.*

c BRASIL. *Leis, Decretos. Lei nº 10.098 de 19/12/2000, Diário Oficial da União, Brasília, 20, dez., 2000. p. 2. Estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida, e dá outras providências.*

d Libras is the Brazilian Portuguese sign language system of communication.

e PERLIN, Gladis. *Identidades surdas.* In: SKLIAR, Carlos. *A Surdez: Um Olhar sobre as Diferenças.* Porto Alegre: Mediação, 1998.

f **DECLARAÇÃO de Salamanca. Sobre Princípios, Políticas e Práticas na Área das Necessidades Educativas Especiais.** [online] 10 jun. 1994. [capturado em 10 jan. 200] Disponível em: <http://www.iie.min-edu.pt/proj/inclusivas/declaracao-salamanca.doc>. Essa Conferência Mundial contou com a participação de noventa e dois países e vinte cinco organizações internacionais, reunidos em Salamanca, Espanha, de 7 a 10 de Julho de 1994

g BRASIL. *Leis, Decretos. Lei nº 10.436, de 24/04/2002. Diário Oficial da União, Brasília, 25, abr., 2002. p. 23. Dispõe sobre a língua brasileira de sinais - Libras e dá outras providências.*

h REVISTA DA FEDERAÇÃO NACIONAL DE EDUCAÇÃO E INTEGRAÇÃO DE SURDOS nº 18.

i INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS (Brasil). **Sinopse estatística da educação básica: censo escolar 2000.** Brasília, 2000.

j Pesquisa realizada no INES – Instituto Nacional de Educação de Surdos, em março de 2002, com pessoas portadoras de surdez profunda, com ensino médio completo.

k *O closed caption (legenda oculta) é um tipo de legenda concebido para permitir aos surdos e pessoas com dificuldades auditivas o acesso a programas, comerciais e filmes veiculados na televisão e em vídeo, entre outros.*