

Avaliação da Compreensão de Textos Jornalísticos em Português, em LIBROL e em LIBRAS por Estudantes Surdos

Igor Q. Silva¹, Pablo F. Matos², Crisley F. Oliveira³, Talita C. B. Leite⁴,
Karoline S. Tavares³, Bruna C. O. Silva³

¹Discente Técnico em Eletrônica Integrado, ²Docente de Informática, ³Discente Técnico em Informática Integrado, ⁴Discente Superior em Ciência da Computação

¹²³Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA)
Av. Amazonas, 3150, Zabelê – 45.075-265 – Vitória da Conquista – BA – Brasil

⁴Universidade Estadual do Sudoeste da Bahia (UESB)
Estrada do Bem Querer, km 4 – 45.083-900 – Vitória da Conquista – BA – Brasil

igorqsilva@gmail.com, pablofmatos@ifba.edu.br, c.farias.box@gmail.com,
{talitaleitecomp, karoline.lorak, brunanegalora}@hotmail.com

Abstract. *The LIBROL software automatically converts texts in Portuguese language for the LIBRAS language, in which aims to facilitate the understanding the Portuguese texts by the deaf. To evaluate this comprehension, this paper proposes to measure the understanding level of the journalistic texts in Portuguese, in LIBROL and in LIBRAS by the deaf students. With the experiment, it was found that the LIBROL is a hopeful means for deaf students learn to study without the presence of a human translator and acquire an increasingly independence in academic and professional life.*

Resumo. *O software LIBROL converte automaticamente textos no idioma Português para o idioma LIBRAS; Tem o intuito de facilitar a compreensão de textos em Português pelos surdos. Para avaliar esta compreensão, este artigo propõe mensurar o grau de compreensão de textos jornalísticos em Português, LIBROL e LIBRAS por estudantes surdos. Com o experimento, constatou-se que o LIBROL é um meio esperançoso para que os estudantes surdos aprendam a estudar sem a presença de um tradutor humano e adquiram cada vez mais independência na vida acadêmica e profissional.*

1. Introdução

A Língua Brasileira de Sinais (LIBRAS) é a forma de comunicação utilizada pelos surdos [1]. Tal como outras línguas, LIBRAS é composta por morfologia, fonologia, sintaxe e semântica. A comunidade surda se autodenomina surdos e não deficientes auditivos, já que a surdez é de origem congênita. A língua de sinais é diferenciada da portuguesa em alguns pontos, pode-se citar: a inexistência de interjeição, conjunções e conectivos além da apresentação dos verbos no infinitivo [1]. Essas diferenças são o que dificultam a compreensão de um texto por um surdo. Segundo censo de 2010 pelo IBGE, cerca de 9,7 milhões de brasileiros possui deficiência auditiva, o que representa 5,1% da população total [2]. É de extrema importância promover inclusão social escolar aos surdos. Isso requer educação de qualidade e incentivo à aprendizagem. Segundo a FENEIS [3], não basta apenas incluir os surdos no sistema escolar; Eles têm que permanecer e concluir os estudos, pois poucos o conseguem. Apesar das formas de realização de inclusão, muitos ainda sentem dificuldades em acompanhar os estudos, o que conseqüentemente os levam a desistência. Apenas 3% concluem o ensino médio [3]. A maior parte dos processos movidos por estudantes surdos contra as escolas é por falta de acessibilidade. A partir deste contexto, pretende-se avaliar o grau de compreensão de textos em Português, em LIBROL [4] e em LIBRAS pelos estudantes surdos. O restante deste artigo está estruturado como segue. Na Seção 2, são resumidos os trabalhos correlatos, enquanto na Seção 3, é apresentada a metodologia de tradução do software LIBROL. Na Seção 4, são descritos a realização do experimento com estudantes surdos, e na Seção 5, são apresentadas a Conclusão e os Trabalhos Futuros.

2. Trabalhos Correlatos

Os trabalhos encontrados na literatura podem ser classificados em três categorias principais: direcionados ao auxílio dos surdos na leitura e entendimento de textos em português, que tem o intuito de converter textos do português para o correspondente em LIBRAS [5, 6]; A segunda categoria é focada na comunicação entre surdo e ouvinte que tem a intenção de diminuir a dificuldade de comunicação, utilizando sinais [6, 7, 8]; Já a terceira é focada no ensinamento de LIBRAS aos ouvintes [9, 10].

3. Metodologia de Tradução

O software LIBROL converte automaticamente textos no idioma Português para o idioma LIBRAS; Tem o intuito de facilitar a compreensão de textos em Português pelos surdos, proporcionando ao surdo uma independência nos estudos do tradutor humano. O software conta com uma interface desenvolvida para tornar a experiência do usuário única. Basicamente há uma caixa de texto de entrada, na qual o usuário informa o texto em Português, e uma caixa de texto de saída, na qual o programa mostra o texto traduzido para LIBRAS. Não foi possível apresentar a interface do LIBROL devido à limitação de páginas. Contudo, a interface segue a mesma simplicidade do projeto original [4].

A metodologia de tradução do LIBROL é composta por quatro etapas: (1) *Part-Of-Speech*, (2) *Stopwords*, (3) *Stemmer* e (4) Banco de Dados. Na **Etapa 1**, *Part-Of-Speech* (POS), o software classifica as palavras de acordo com sua classe gramatical utilizando o etiquetador morfossintático MXPOST [11]. A título de exemplo, considera-se a sentença em Português a seguir: “O LIBROL converte textos de português para LIBRAS.”. “O” é classificado com a etiqueta `_ART` por ser um artigo. “LIBROL” é classificado com a etiqueta `_NP` por ser um nome próprio. Contudo, nem todas as classificações são 100% corretas. O MXPOST possui 96,86% de acerto durante a fase de treinamento. Na fase de teste, o percentual de acerto baixa para 86,91%. Na **Etapa 2**, aplica-se a lista de *Stopwords* [12], cujo objetivo principal é excluir as palavras desnecessárias do texto que o surdo não compreende, tais como, interjeição, preposição e artigo. Utiliza-se a técnica Expressão Regular (ER) para excluir todas as *Stopwords*. O artigo “O” e as preposições “de” e “para” foram excluídas. Na **Etapa 3**, utiliza-se da técnica de Processamento de Língua Natural *Stemmer* [12] para converter todos os verbos para seus respectivos radicais (e.g., converte → convert). Como o texto já está etiquetado pelo POS, utiliza-se da ER para casar com um padrão preestabelecido. O padrão a ser casado é com palavras classificadas com a etiqueta `_VERB`. Por fim, na **Etapa 4**, utiliza-se de um Banco de Dados, que tem o objetivo de recuperar o verbo em LIBRAS que está no infinitivo, cujo correspondente seja o verbo em Português que foi armazenado previamente como radical do verbo (e.g., convert → converter). Ao todo foram armazenados aproximadamente 11.500 verbos [13]. Ao final do processo de tradução, o texto convertido pelo LIBROL é mais fácil de ser compreendido pelo surdo do que o texto originalmente em Português, a saber: “LIBROL converter textos Português LIBRAS.”.

4. Experimento

A partir da metodologia apresentada, pretende-se avaliar o quão viável é a compreensão de textos convertidos pelo LIBROL. Assim, o objetivo do experimento é mensurar a compreensão de textos jornalísticos aplicados a estudantes surdos. O experimento foi realizado com a aplicação de três textos com dois estudantes brasileiros surdos: um do ensino técnico e um do ensino superior. Para avaliar a compreensão dos textos pelo surdo, o mesmo teve que ler o texto, elaborar uma síntese e responder questões objetivas sobre os textos em três momentos e em dias diferentes: **Momento 1**, textos no idioma Português; **Momento 2**, textos convertidos pelo software LIBROL; e **Momento 3**, textos no idioma LIBRAS. O *Baseline* é o texto em Português; O *Gold Standard* é o texto em LIBRAS. Deseja-se responder a seguinte questão a partir da realização deste experimento: A leitura de um texto em Português, convertido em LIBRAS pelo LIBROL, é **fácil** o suficiente para viabilizar a independência do surdo nos estudos? Com isto, deseja-se justificar e validar a utilização do software LIBROL.

4.1. Etapas do Experimento

Os três momentos foram realizados seguindo as quatro etapas, a saber: **Etapa 1**, Explicação do Experimento em LIBRAS por um Intérprete. O intérprete foi responsável por explicar o experimento para o surdo avaliado. **Etapa 2**: Compreensão do Texto. O surdo teve o tempo necessário para ler e compreender o texto. **Etapa 3**: Elaboração de Síntese. Após a leitura do texto, o surdo elaborou uma síntese do que entendeu sobre o texto. **Etapa 4**: Resposta do Questionário. Após a síntese, o surdo teve que responder questões objetivas sobre os textos. As perguntas foram feitas pelo intérprete, sem o conhecimento prévio do surdo e do intérprete sobre o assunto abordado nos textos. A resposta da pergunta foi escrita pelo próprio surdo.

4.2. Características dos Textos e das Perguntas Objetivas

Características dos Textos: Foram utilizados três textos: **Texto 1** cujo tema central é Combate ao Vírus HIV; **Texto 2** cujo tema é Vulcão; e **Texto 3** cujo tema é *Bullying*. As imagens dos textos foram excluídas. Os textos têm a quantidade máxima de uma página, sendo a média de parágrafos e de palavras igual, respectivamente, a quatro parágrafos e a trezentas palavras. Os textos são referentes a textos jornalísticos publicados entre 2012 e 2013, escolhidos aleatoriamente da revista *Veja* e do site G1. O surdo teve acesso aos três textos do início ao fim do experimento. A ordem de aplicação dos textos foi sequencial: Texto 1 seguido dos Textos 2 e 3. Características das Perguntas Objetivas: A quantidade média de pergunta dos textos foi de cinco perguntas. As perguntas tinham características **gerais**, por exemplo, “*Do que se trata o texto?*” e **específicas** de acordo com o conteúdo do texto, por exemplo, “*Qual a consequência da fumaça da erupção do vulcão?*” ou “*Quais as principais formas de transmissão do vírus HIV para a criança?*”. Todas as respostas das perguntas puderam ser encontradas no texto, não havendo necessidade de conhecimento prévio para responder as mesmas.

4.3. Análise do Experimento

De forma geral, em ambos os entrevistados, verificou-se que houve uma grande dificuldade na compreensão dos textos em Português (**Momento 1**), seja pelos textos longos ou pela terminologia que não fazia parte do vocabulário dos avaliados. As sínteses em nenhum dos três textos estavam de acordo com o tema abordado, e muitas vezes nem era mencionado o assunto principal. As respostas das perguntas objetivas foram incoerentes e incorretas, mostrando que os avaliados não leram ou não compreenderam os textos. Nos textos convertidos pelo LIBROL (**Momento 2**), verificou-se que os avaliados se sentiram mais à vontade com a leitura e mais dispostos a elaborar a síntese e a responder o questionário. Observou-se que nestes textos, a síntese e as respostas foram mais coerentes, apesar de ainda haver dificuldades de compreensão e algumas respostas terem sido incorretas. Estas diferenças podem ser percebidas no exemplo a seguir. Pergunta do questionário referente ao Texto 2: “*A que altura as cinzas do vulcão foram lançadas?*”. Resposta de um entrevistado após leitura do texto em Português: “*Muito alta*”. Resposta à mesma pergunta após leitura do texto traduzido pelo LIBROL: “*Cinzas seis quilômetros altura*”. Esta é a resposta esperada para a questão supracitada, visto que o termo “*muito alta*” não foi utilizado no texto em momento algum, mostrando que o avaliado respondeu incoerentemente. Nos textos em LIBRAS (**Momento 3**), observou-se que as respostas aos questionários não foram diferentes dos textos em LIBROL, exceto em algumas questões em que eles responderam mais próximo do esperado. Por exemplo, a questão referente ao Texto 3 “*Quem mais sofre Bullying, homem ou mulher?*”, teve como resposta “*Agressores maior número são homens. Meninas também sofrem mas é mais difícil*”, mostrando que o estudante respondeu corretamente de acordo com o texto lido. Percebeu-se também a utilização de conhecimentos prévios em alguns momentos do questionário. Por exemplo, na questão “*Qual a consequência da fumaça da erupção?*”, o estudante respondeu “*Vários problemas de saúde*”. Porém, este problema não foi mencionado no texto, levando a crer que possivelmente o estresse, a correria do dia a dia e a aplicação repetitiva dos textos levaram o entrevistado a induzir sua resposta baseada em conhecimentos anteriores. Por fim, as sínteses dos textos em LIBRAS foram divergentes. Enquanto o surdo estudante do ensino superior escreveu textos totalmente fora do tema, o

XXXIV Congresso da Sociedade Brasileira de Computação – CSBC 2014
estudante do técnico escreveu uma síntese mais concreta, contendo informações que só poderiam ser citadas se houvesse um entendimento real do texto. Observou-se ainda que a síntese em LIBROL foi convergente com a síntese em LIBRAS.

5. Conclusão e Trabalhos Futuros

Os resultados observados no experimento respondem a pergunta mencionada no início da Seção 4: A compreensão do texto em LIBROL é mais fácil de compreender do que os textos em Português (*Baseline*), porém mais difícil do que os textos em LIBRAS (*Gold Standard*). Portanto, o software LIBROL, que converte automaticamente textos no idioma Português para o idioma LIBRAS, é um meio esperançoso para que os estudantes surdos aprendam a estudar sem a presença de um tradutor humano e adquiram cada vez mais independência na vida acadêmica e profissional. Como trabalhos futuros, pode-se acrescentar o recurso de visualização de sentenças (texto) em sinais de LIBRAS (imagem). Usar imagens para ilustrar o texto é uma solução para diminuir a dificuldade do surdo em ler textos originados do Português. Também ajudará o surdo a aprender novas terminologias. Estas melhorias serão capazes de estimular o surdo a aprender novas palavras, facilitando a compreensão de textos originados do Português.

Agradecimentos

Agradecemos o apoio do Núcleo de Acessibilidade e Inclusão para Pessoas com Deficiência (NAIPD) da UESB *campus* Vitória da Conquista/BA, e a participação dos estudantes surdos e dos intérpretes que contribuíram positivamente com a realização do experimento.

Referências

1. Ballantyne, J., Martin, M.C. and Martin, A. (1995) **Surdez**. Artes Médicas Sul, RS.
2. Chihaya, K. (2012) **Deficiência auditiva atinge 9,7 milhões de brasileiros**, <http://www.winaudio.com.br/produtos-e-servicos/noticias-em-audiologia/3704-deficiencia-auditiva-atinge-98-milhoes-de-brasileiros.html>.
3. FENEIS. (2014) **Federação Nacional de Educação e Integração dos Surdos**, <http://www.feneis.org.br/>.
4. Carvalho, R.S. et al. (2013) LIBROL: Software Tradutor de Português para LIBRAS. In: **ENCOMPIF**, pages 2098-2101. CSBC.
5. Nunes, M.G.V. et al. (2003) **Projeto Pulo**, USP, São Carlos, <http://www.nilc.icmc.usp.br/nilc/download/PULO-Jun-2003.zip>.
6. Oliveira, J.P.S. (2012) ProDeaf: Uma ferramenta colaborativa para a produção de conhecimento em Libras. In: **INES**, pages 1-6.
7. Januário, G.C., Leite, L.A.F. and Koga, M.L. (2010) **Poli-Libras: Um Tradutor de Português para Libras**. 93 f. TCC (Bacharelado em Eng. da Comp.) – USP, São Paulo.
8. HandTalk. (2014) **Hand Talk Tradutor para Libras**, <http://www.handtalk.me/>.
9. Moura, G. et al. (2013) Luz, Câmera, Libras!: How a Mobile Game Can Improve the Learning of Sign Languages. **DUXU/HCI**. LNCS, Springer, pages 266-275.
10. Rocha, D.F.S., Bittencourt, I.I. and Brito, P. (2013) AssistLibras: Um Assistente Gráfico para Construção de Sinais 3D da LIBRAS. In: **WCBIE**, pages 132-141. CBIE.
11. Aires, R.V.X. (2000) **Implementação, adaptação, combinação e avaliação de etiquetadores para o português do Brasil**. 154 f. Dissertação – ICMC, USP, São Carlos.
12. Dias, M.A.L. and Malheiros, M.G. (2006) Automatic extraction of keywords for the portuguese language. In: **PROPOR**, pages 204-207. Springer-Verlag.
13. Achando. (2014) **Verbos em Português**, <http://www.achando.info/verbos/conjugar/dicionario-a.html>.