

	<p>UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB CENTRO DE INFORMÁTICA - CI NÚCLEO DE PESQUISA E EXTENSÃO LAVID</p> <p>COORDENADOR: Prof. Dr. Tiago Maritan U. de Araújo CONTATO: tiagomaritan@lavid.ufpb.br</p>	
---	--	---

RELATÓRIO DE AVALIAÇÃO DA SOLUÇÃO VLIBRAS PELA COMUNIDADE SURDA

Projeto Suite VLibras: Tradutor Automático de Conteúdos Digitais para LIBRAS em
Código Aberto e Distribuição Livre

COORDENADOR: Prof. Dr. Tiago Maritan U. de Araújo
VICE-COORDENADOR: Prof. Dr. Guido Lemos Souza Filho

Maio - 2015

1. Introdução

Para avaliar a qualidade do VLibras, vários testes foram realizados com a comunidade surda. Uma dessas principais avaliações mais completa foi a realizada na Fundação Centro Integrado de Apoio à Pessoa com Deficiência (FUNAD) da Paraíba, nos dias 29 de Julho e 02 de Agosto de 2012 na Fundação Centro Integrado de Apoio à Pessoa com Deficiência (FUNAD) da Paraíba.

Esse experimento foi dividido em três etapas. Na primeira, os conteúdos acessíveis gerados pela solução proposta foram avaliados por usuários surdos com relação ao nível de compreensão e a qualidade de tradução desses conteúdos. Com isso, é possível investigar se os usuários são capazes de compreender os conteúdos acessíveis gerados e se sentem confortáveis com os mesmos e, por consequência, se é possível reduzir as suas barreiras de acesso à informação. Na segunda parte do experimento, o desempenho de tradução da solução proposta é avaliado, com o objetivo de investigar se a solução proposta era capaz de traduzir esses conteúdos em cenários que exigem tradução em tempo real. Por fim, na terceira parte, a ferramenta WikiLIBRAS é avaliada por usuários surdos, com o objetivo de investigar a produtividade dos usuários na geração das construções linguísticas da solução proposta.

O processo do experimento envolveu a definição, planejamento e execução desse experimento, além da análise estatística e interpretação dos resultados. As etapas de execução do experimento e de análise e interpretação dos resultados serão apresentadas nas próximas seções.

2. Avaliação do Nível de Compreensão dos Conteúdos

O experimento para avaliação do nível de compreensão dos conteúdos foi realizado nos dias 29 de Julho e 02 de Agosto de 2012 na Fundação Centro Integrado de Apoio à Pessoa com Deficiência (FUNAD) da Paraíba. O experimento contou com a participação de vinte usuários surdos selecionados por conveniência dos cursos de formação em LIBRAS da FUNAD. Esses usuários foram divididos aleatoriamente em dois grupos de dez usuários: um grupo para avaliar os objetos com legendas e o outro grupo para avaliar os objetos com trilhas de LIBRAS gerados pela solução proposta. Durante todo o experimento, intérpretes de LIBRAS participaram do processo, intermediando a comunicação com os usuários.

Inicialmente, os usuários foram preparados para o teste. Na fase de preparação, o objetivo do teste foi apresentado para os usuários (avaliar o nível de compreensão em um conjunto de conteúdos multimídia), mas as hipóteses a serem testadas não foram apresentadas. Em seguida, os usuários, já divididos em dois grupos, receberam uma cópia do questionário, e foram instruídos a preencher a primeira parte dele com suas informações pessoais.

O grupo de usuários era heterogêneo e composto por onze mulheres e nove homens na faixa etária de 13 a 56 anos, e com uma média de idade de 28,6 anos. O grau de escolaridade e o nível de conhecimento em Português e LIBRAS dos usuários também foi observado e é apresentado na Figura 1 e na Tabela 1, respectivamente.

Figura 1: Grau de escolaridade dos usuários participantes do experimento.



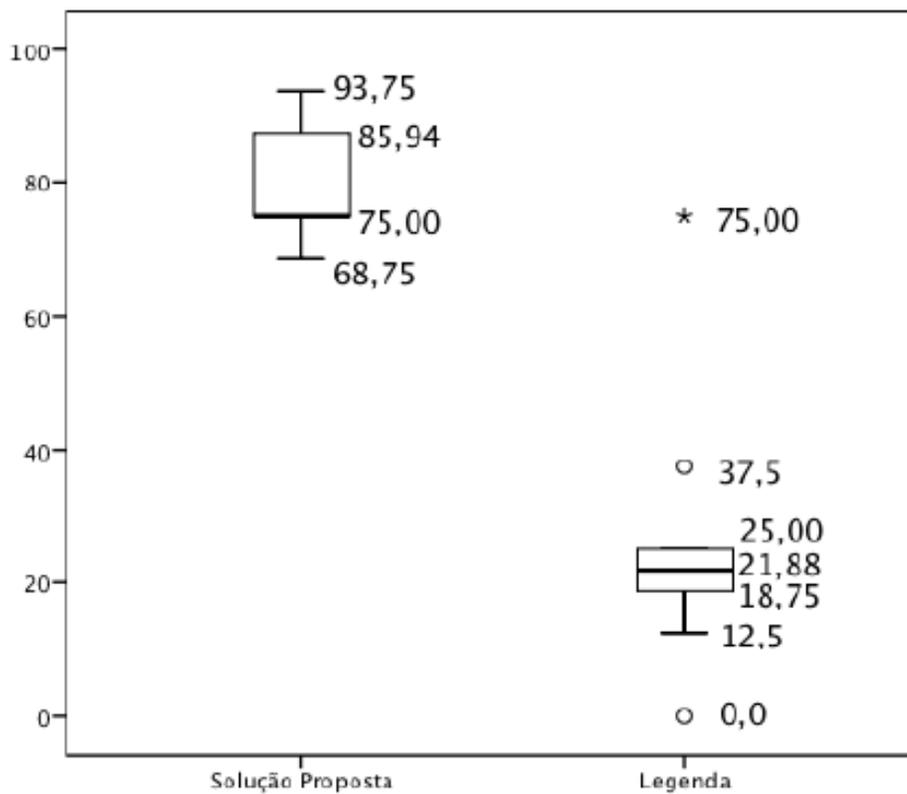
Após o preenchimento da primeira parte do questionário, os usuários (divididos em dois grupos) foram convidados a assistir 4 (quatro) conteúdos multimídia com o seu tratamento (legendas ou trilhas de LIBRAS geradas pelo VLibras). Cada conteúdo (vídeo) foi apresentado duas vezes aos usuários e após a apresentação de cada um desses conteúdos, os usuários foram convidados a responder um conjunto de 16 (dezesesseis) questões relacionadas ao conteúdo apresentado (segunda parte do questionário). Por fim, os usuários foram convidados a responder as perguntas da terceira parte do questionário, avaliando subjetivamente alguns aspectos da solução. Algumas fotos deste experimento são apresentados na Figura 3.

Tabela 1: Nível de conhecimento em Português e LIBRAS dos usuários.

Nível de Conhecimento	Valor Médio	Desvio Padrão (%)
Português (1 a 6)	4,00	1,80
LIBRAS (1 a 6)	5,67	1,54

Os resultados dos testes de compreensão de conteúdos deste experimento são apresentados na Figura 4. Conforme pode ser observado nessa figura, para todos os conteúdos avaliados, os usuários que assistiram conteúdos com legendas tiveram uma média de acertos menor que os usuários que assistiram conteúdos com trilhas de LIBRAS geradas pela solução proposta. Considerando todos os conteúdos, a taxa média de acertos para os usuários que assistiram vídeos com trilhas de LIBRAS (solução proposta) foi de 79,38% com um desvio padrão de 9,34%, enquanto que a taxa média de acertos para os usuários que assistiram vídeos com legendas foi de 25,63% com um desvio padrão de 19,86%.

Figura 2: Gráfico de caixa (box plot) para os testes de compreensão dos conteúdos



Para uma melhor compreensão desses dados, um gráfico de caixa (ou *boxplot*) foi gerado com os resultados desses testes. Esse gráfico é apresentado na Figura 2. É importante ressaltar que os bigodes (*whiskers*) representados desse gráfico foram definidos de acordo com a proposta de Frigge et al. (1989), que definem os bigodes superior e inferior como o comprimento da caixa (i.e., a diferença entre o terceiro e o primeiro quartil) multiplicado por 1,5 e somado e subtraído terceiro (bigode superior) e primeiro quartil (bigode inferior), respectivamente.

De acordo com a Figura 2, é possível observar que existe um padrão claro de que os usuários que assistiram os conteúdos com trilhas de LIBRAS baseados na solução proposta obtiveram um melhor desempenho nos testes de compreensão de conteúdos do que os usuários que avaliaram os conteúdos com legendas. Para os conteúdos baseados na solução proposta, os valores da mediana, do primeiro e terceiro quartil da distribuição foram de 75,00%, 75,00% e 85,94%, respectivamente. Isso significa que 50% dos usuários obtiveram uma taxa de acertos na faixa de 75,00% a 85,94%. Além disso, nenhum *outlier* foi identificado na distribuição, o que indica que todos os usuários obtiveram uma taxa média de acertos entre 68,75% e 93,75%.

Figura 3: Algumas fotos do experimento com usuários (nível de compreensão) (a) Alguns usuários do experimento, (b) Fase de preparação dos usuários, (c) Usuários avaliando conteúdos legendados, (d) Usuários avaliando conteúdos com trilhas de LIBRAS geradas pela solução proposta.

(a)



(b)



(c)

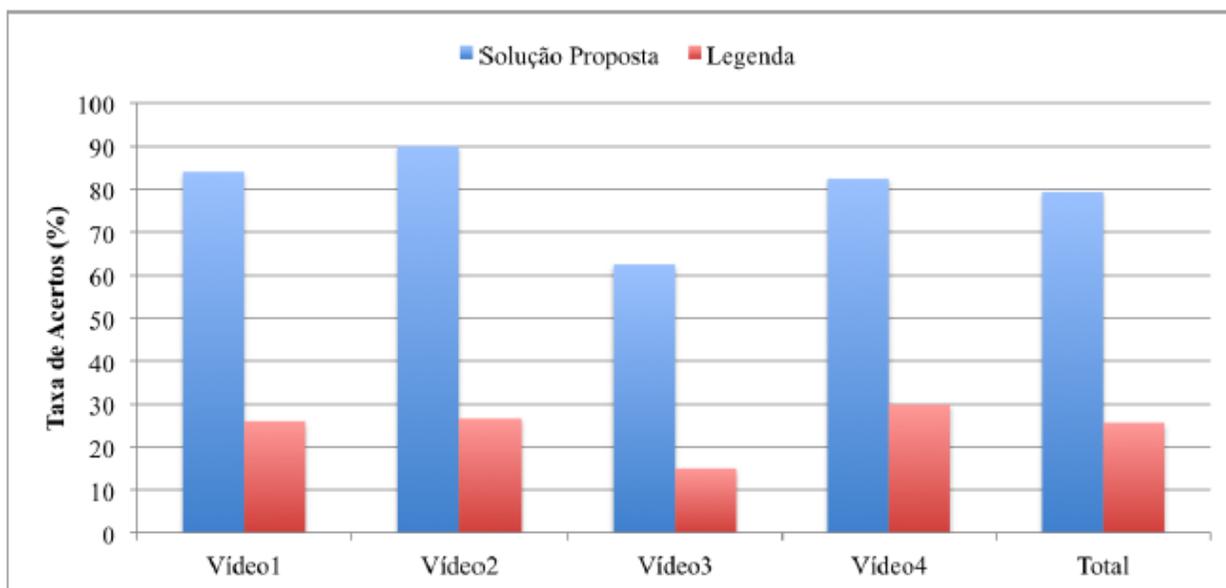


(d)



Para os conteúdos com legendas, os valores da mediana, do primeiro e terceiro quartil da distribuição foram, respectivamente 21,88%, 18,75% e 25,00%. Isso significa que menos do que um quarto dos usuários obtiveram uma taxa de acertos superior a 25,00% das questões do experimento. Além disso, três outliers foram identificados nesta amostra. Um outlier negativo, representando um usuário que não obteve nenhum acerto (0,00%) e dois outliers positivos, representando dois usuários que obtiveram uma taxa de acertos de 37,5% e 75,00% para esses conteúdos. Como os outliers, neste caso, representam casos que podem realmente ocorrer na prática (isto é, usuários surdos que não sabem ler absolutamente nada - *outlier* negativo - e usuários surdos que sabem ler um pouco melhor *outlier* positivo com 75% de acertos), essas amostras não foram removidas da análise.

Figura 4: Resultado dos testes de compreensão dos conteúdos



Para avaliar se essa diferença de desempenho é estatisticamente representativa, e, conseqüentemente, testar a hipótese relacionada ao nível de compreensão dos conteúdos (Hipótese 1), um teste-t foi então aplicado sobre todo o conjunto de amostras considerando um intervalo de confiança de 95% e 18 graus de liberdade (número de usuários ou amostras menos dois, ou seja, $20-2=18$). Os resultados da aplicação desse teste são apresentados na Tabela 2.

Tabela 2: Resultados do teste-t para a Hipótese 1 (compreensão dos conteúdos)

μ compreensao (SP)	79,38%	σ compreensao (SP)	9,34%
μ compreensao (Leg)	25,63%	σ compreensao (Leg)	19,86%
Grau de Confiança (p-valor)	0,95	Graus de Liberdade	18
t-valor	7,74	t0,0025,18	2.12

De acordo com a Tabela 2, uma vez que o t -valor obtido para o teste foi maior do que o valor crítico para o teste-t com intervalo de confiança de 95% e 18 graus de liberdade ($t_{0,0025,18}$) [Wohlin et al. 2000], é possível concluir que a hipótese H_0 pode ser rejeitada. Dessa forma, é possível afirmar com um grau de confiança de 95% que existe uma diferença significativa de compreensão dos conteúdos quando usuários surdos são submetidos a conteúdos com trilhas de LIBRAS baseadas na solução proposta com relação ao nível de compreensão quando os usuários são submetidos a conteúdos com legendas.

Por fim, alguns aspectos da solução como, por exemplo, a qualidade de tradução, naturalidade da apresentação, dentre outros, também foram avaliados de forma subjetiva pelos usuários. Nessa avaliação, após assistir todos os conteúdos, os usuários atribuíram conceitos de 1 a 6 para cada um desses aspectos. Alguns deles, no entanto, que não fazem sentido para os conteúdos com legendas, como, por exemplo, naturalidade da apresentação, qualidade dos movimentos das mãos e qualidade das expressões faciais, e não foram avaliados pelos usuários do grupo de conteúdos com legendas. Os resultados dessa avaliação são apresentados na Tabela 3.

De acordo com a Tabela 3, o nível de compreensão e a qualidade da tradução, em média, obtiveram valores médios moderados (4,60 e 4,60, respectivamente) nos conteúdos gerados pela solução proposta. Além disso, esses valores foram maiores do que os valores correspondentes nos conteúdos com legendas (3,70 e 4,13, respectivamente). Os altos valores de desvio padrão para os dois tipos de conteúdos, no entanto, mostram que as opiniões dos usuários divergiram com relação a avaliação desses aspectos.

Um aspecto importante a ser observado, no entanto, é que a avaliação do aspecto "nível de compreensão" para os conteúdos com legendas não foi compatível com os testes de compreensão realizados com os usuários. Embora os usuários que avaliaram estes conteúdos tenham indicado que eles foram razoavelmente compreendidos (3,70), os resultados dos testes de compreensão mostraram que a maioria dos usuários não haviam compreendido bem os conteúdos (taxa média de acerto de 25% das questões).

Tabela 3: Valores médios para os aspectos avaliados (escalas de 1 a 6)

Aspecto avaliado	Solução Proposta		Legenda	
	Valor Médio	Desvio Padrão	Valor Médio	Desvio Padrão
Nível de compreensão	4,60	1,68	3,70	2,33
Fluência Gramatical	4,60	1,56	4,13	2,05
Naturalidade	4,40	1,74	-	-
Qualidade dos movimentos	4,80	1,40	-	-
Qualidade das expressões faciais	4,56	1,89	-	-

É possível verificar essa incompatibilidade, analisando a correlação entre as duas variáveis: (1) resultados dos testes de compreensão e (2) avaliação do "nível de compreensão". Ao calcularmos os valores do coeficiente de correlação de Pearson e do coeficiente de correlação de postos de Spearman [Wohlin et al. 2000] para essas variáveis, os valores obtidos foram de 0,033 e -0,182, respectivamente, o que indica uma baixa correlação entre as variáveis. Uma das possíveis explicações para isso, é que segundo Wohlin et al. (2000), os humanos têm medo de serem avaliados e quando passam por um processo de avaliação tentam parecer melhor do que realmente são, o que pode atrapalhar a saída do experimento.

Além disso, com relação a fluência gramatical, outra inconsistência nos resultados é que muitos usuários indicaram que os conteúdos com legendas eram compatíveis com a gramática de LIBRAS (4,13 de valor médio), enquanto, na realidade, os conteúdos estavam sendo transmitidos na gramática da língua portuguesa.

A naturalidade e qualidade da apresentação também foram avaliados. A naturalidade da apresentação também obteve um valor médio moderado (4,40), mas foi o aspecto avaliado que obteve o pior valor médio. Esse resultado é compatível e próximo dos valores obtidos para a qualidade dos movimentos das mãos (4,80) e de qualidade das expressões faciais (4,56). Contudo, de forma similar, as outras medidas, os valores de desvio padrão também foram altos, o que sugere que existem alguns indícios de divergência entre os usuários.

Assim como em San-Segundo et al. (2011), algumas prováveis causas para essas divergências foram observadas durante este experimento. Por exemplo, durante o experimento, os usuários divergiram sobre a estrutura de algumas sentenças em LIBRAS. De forma similar as outras línguas de sinais (como por exemplo, na LSE [San-Segundo et al. 2011]), a LIBRAS possui um alto grau de flexibilidade na estruturação das sentenças. Essa flexibilidade é algumas vezes não é bem compreendida e algumas das possibilidades foram consideradas como sentenças incorretas. Além disso, alguns usuários também divergiram

com relação a correta sinalização de alguns sinais, como, por exemplo, os sinais CAFÉ e MERCADO.

Uma alternativa para reduzir essas divergências no LibrasTV, por exemplo, seria customizar os Dicionários de LIBRAS nos receptores de TV dos usuários. O desenvolvimento desses dicionários customizados, no entanto, é uma tarefa muito trabalhosa. Outra alternativa seria investir mais esforços para padronizar a LIBRAS. Nesse caso, uma maior disseminação de LIBRAS nas TIC poderia auxiliar nesse processo de padronização, como também vem ocorrendo em outras línguas menores na Espanha [San-Segundo et al. 2011].

Em consequência dessa forte divergência, ao aplicarmos o teste-t (ver Tabela 4), não é possível rejeitar a segunda hipótese deste experimento relacionada a adequação a gramática de LIBRAS (Hipótese 2) e, conseqüentemente, não é possível afirmar que os conteúdos gerados pela solução proposta se adéquam melhor a gramática de LIBRAS do que os conteúdos com legendas. Os bons resultados nos testes de compreensão dos conteúdos, no entanto, são um indício de que a qualidade de tradução é adequada. Uma análise qualitativa mais profunda, contudo, precisa ser desenvolvida para avaliar esse aspecto, sendo uma das propostas de trabalho futuro.

Tabela 4: Resultados do teste-t para a Hipótese 2 (adequação a gramática de LIBRAS)

μ traducao (SP)	4,60	σ traducao (SP)	1,56
μ traducao (Leg)	4,13	σ traducao (Leg)	2,05
Grau de Confiança (p -valor)	0,95	Graus de Liberdade	18
t -valor	0,58	$t_{0,0025,18}$	2.12

3. Avaliação do Desempenho de Tradução

O experimento para avaliar o desempenho de tradução foi realizado utilizando testes objetivos com um sinal real de TV Digital durante um período de 24 horas (das 9 horas do dia 20 de Dezembro de 2011 até as 09 horas do dia 21 de Dezembro de 2011). Durante esse período, o sinal da "TV Correio", uma emissora de TV Digital Brasileira afiliada da "TV Record", foi sintonizado pelo "Retransmissor IP MPEG-2 TS" e o fluxo MPEG-2 TS desse sinal foi transmitido para o "Tradutor de LIBRAS" e para o Multiplexador. Sempre que pacotes de legendas eram recebidos pelo "Tradutor de LIBRAS", o processo de tradução e geração das trilhas de LIBRAS era executado pelo protótipo e os atrasos de cada componente eram medidos e armazenados, resultando na tradução de cerca de 2200 sentenças.

Na Tabela 5 são apresentados os valores dos atrasos médio, máximo, mínimo e o desvio padrão do experimento para cada um dos componentes. De acordo com a Tabela 5, o atraso médio para executar todos os componentes do LibrasTV foi menor do que 44 milissegundos. O atraso máximo obtido (considerando o atraso máximo de cada um dos componentes) foi de 142 milissegundos, enquanto que o atraso mínimo obtido foi de 20,51 milissegundos.

Tabela 5: Medidas do desempenho de tradução para os componentes do protótipo LibrasTV

Componentes	Valor Médio (ms)	Desvio Padrão (ms)	Valor Máximo (ms)	Valor Mínimo (ms)
Filtragem e Extração de Legendas	0,024	0,022	0,554	0,017
Tradução Automática	0,975	2,957	80,126	0,220
Codificação	0,215	0,089	1,061	0,072
Decodificação	0,170	0,143	0,519	0,020
Animação e Distribuição	42,445	8,747	59,998	20,00
Total	43,805	9,434	142,21	20,509

Para avaliar se esse atraso de tradução atende o critério de tradução em tempo real (isto é, μ atraso (SP) 4 segundos), um teste-t foi aplicado sobre essa amostra considerando um grau de confiança de 95% e 2192 graus de liberdade (número de amostras menos dois, ou seja, $2194-2=2192$). Os resultados da aplicação desse teste são apresentados na Tabela 6. De acordo com a Tabela 6.6, uma vez que t-valor obtido para o teste foi maior do que o valor crítico para o teste-t com intervalo de confiança de 95% e 2192 graus de liberdade ($t_{0,0025,2192}$) [Wohlin et al. 2000], é possível concluir que a hipótese H_0 dessa parte do experimento pode ser rejeitada (Hipótese 1). Dessa forma, considerando que esse experimento foi executado com um vocabulário aberto e representativo e, em um cenário real, é possível afirmar com um grau de confiança de 95% que a solução proposta é capaz de traduzir conteúdos em tempo real.

Tabela 6: Resultados do teste-t para a Hipótese desse experimento (desempenho ou atraso da tradução)

μ atraso (SP)	43,805 miliseg.	σ atraso (SP)	9,434 miliseg.
Atraso de referência	4 seg.		
Grau de Confiança (p -valor)	0,95	Grau de Liberdade	2192
t -valor	19.632,8	$t_{0,0025,2192}$	1,96

Além disso, o tempo (atraso) médio de tradução obtido é provavelmente bem inferior ao tempo gasto por uma tradução humana, embora com uma qualidade de tradução inferior também.

4. Avaliação do WikiLibras

O experimento para avaliação da ferramenta de computação humana (WikiLIBRAS) também foi realizado nos dias 29 de Julho e 02 de Agosto de 2012 na FUNAD-PB. Ele contou com a participação de onze usuários surdos e três intérpretes de LIBRAS selecionados por conveniência dos cursos de formação em LIBRAS da FUNAD-PB.

Inicialmente, os usuários foram preparados de forma similar ao experimento de avaliação do nível de compreensão dos conteúdos e preencheram a primeira parte do questionário com suas informações pessoais.

O grupo de usuários também era heterogêneo e composto por sete mulheres e sete homens na faixa etária de 12 a 42 anos, e com uma média de idade de 25,4 anos. O grau de escolaridade e o nível de conhecimento em Português e LIBRAS dos usuários também foi observado e é apresentado na Figura 5 e na Tabela 7, respectivamente.

Figura 5: Grau de Escolaridade dos usuários do Teste

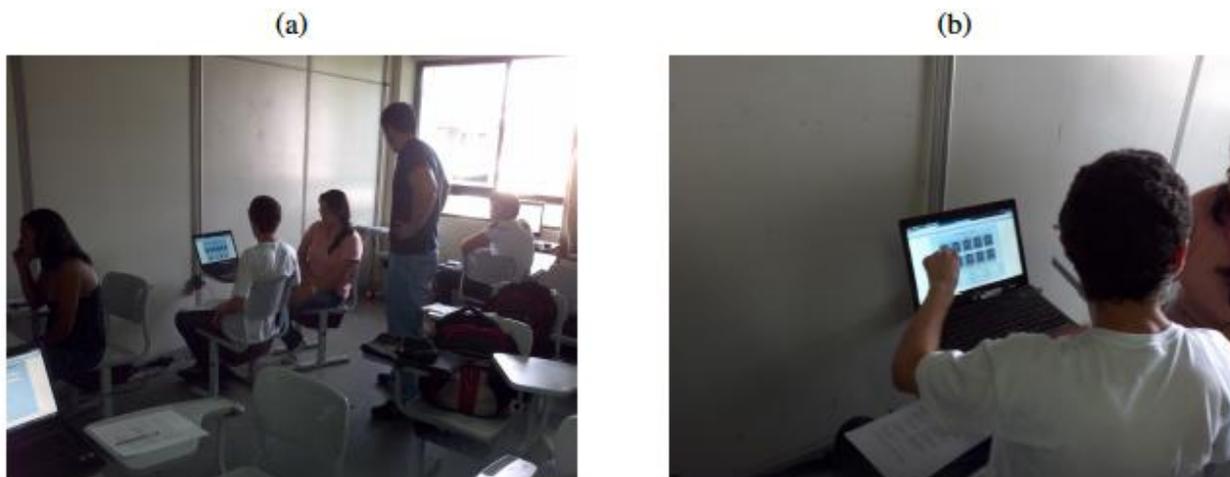


Tabela 7: Nível de conhecimento em Português e LIBRAS dos Usuários

Nível de Conhecimento	Valor Médio	Desvio Padrão
Português (1 a 6)	4,43	1,24
LIBRAS (1 a 6)	5,07	1,16%

Os usuários foram então convidados a gerar os sinais de LIBRAS propostos utilizando o WikiLIBRAS e a preencher a segunda e a terceira parte do questionário. Durante o experimento, o tempo médio gasto pelos usuários para gerar os sinais (eficiência) também foi armazenado no WikiLIBRAS. Na Figura 6 são apresentadas algumas fotos do experimento realizado com surdos e intérpretes de LIBRAS na FUNAD-PB.

Figura 6: Algumas fotos do experimento com usuários (WikiLIBRAS).



Para confrontar com o desempenho dos usuários no WikiLIBRAS, um experimento similar também foi executado com designers-3D no LAViD-UFPB. Nesse experimento, três experientes designers-3D foram convidados a gerar o mesmo conjunto de sinais utilizando a ferramenta de modelagem e animação Blender3. Para auxiliar os designers-3D na animação dos sinais, um vídeo com cada sinal representado por um intérprete de LIBRAS foi gerado e passado para os designers para ser usado como referência na geração dos sinais. Os três designers que participaram do experimento são experientes e participam de projetos de pesquisa no LAViD envolvendo modelagem e animação-3D, sendo dois deles alunos de graduação e o outro aluno de pós-graduação da UFPB.

Os principais resultados desses dois experimentos são apresentados nas Figuras 7 e 8. De acordo com a Figura 7, é possível observar que os designers-3D conseguiram gerar todos os sinais corretamente na ferramenta de animação, enquanto que os surdos e intérpretes de LIBRAS conseguiram gerar, em média, cerca de 81,43% dos sinais no WikiLIBRAS. Com relação ao tempo médio para gerar os sinais corretamente (ver Figura 8), no entanto, o tempo médio gasto pelos usuários no WikiLIBRAS (93,96 segundos) foi bem inferior ao tempo médio gasto pelos designers na ferramenta de animação (711,33 segundos).

Analisando a geração de cada sinal individualmente, é possível observar que os usuários tiveram, em média, mais dificuldades para gerar os sinais PROFESSOR e PRESIDENTE no WikiLIBRAS. O sinal PROFESSOR foi gerado corretamente por 57,14% dos usuários, enquanto que o sinal PRESIDENTE foi gerado corretamente por 75,86% dos usuários. Para os outros sinais (LÁBIO, CALAR e TIO), os usuários encontraram menos dificuldades e obtiveram uma taxa média de acertos superior a 85%. Esses resultados são compatíveis com o tempo médio gasto pelos usuários para gerar os sinais. Nos sinais que os usuários tiveram mais dificuldades, como, por exemplo, nos sinais PROFESSOR e PRESIDENTE, o tempo médio para gerar os sinais (118,72 segundos e 126,33 segundos, respectivamente) foi maior do que o tempo médio necessário para gerar os sinais LÁBIO, CALAR e TIO (87,88 segundos, 58,91 segundos e 53,31 segundos, respectivamente).

Figura 7: Taxa de sinais gerados corretamente pelos usuários (WikiLIBRAS vs Manual)

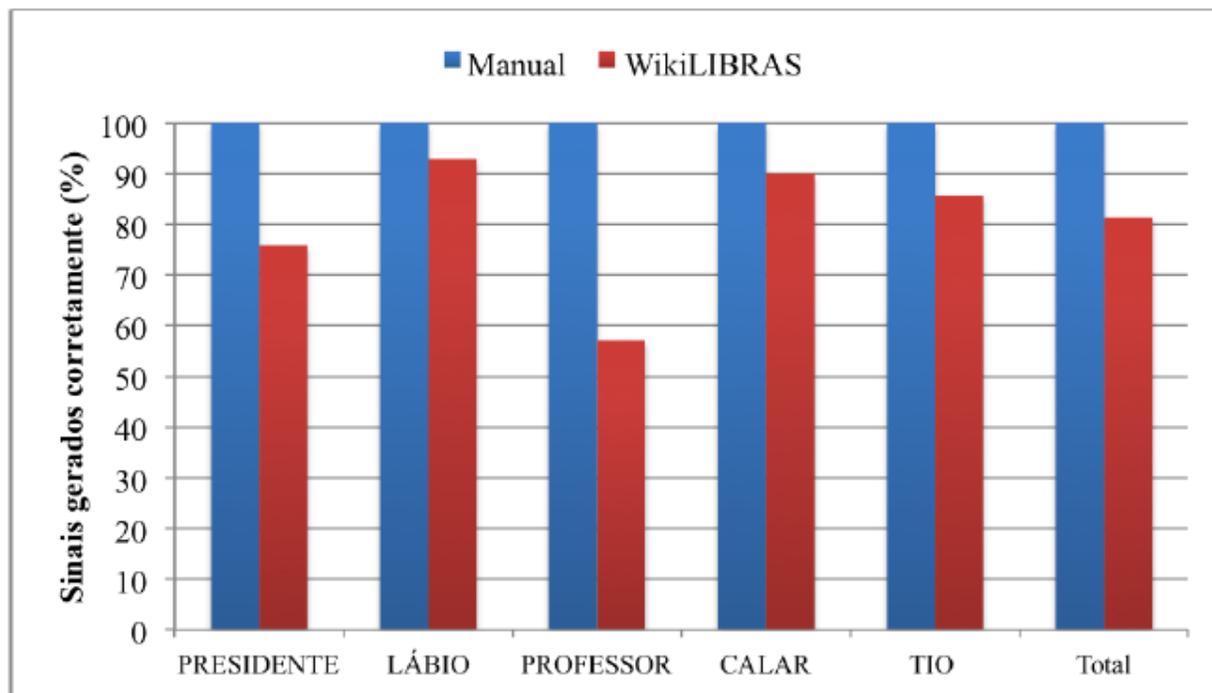
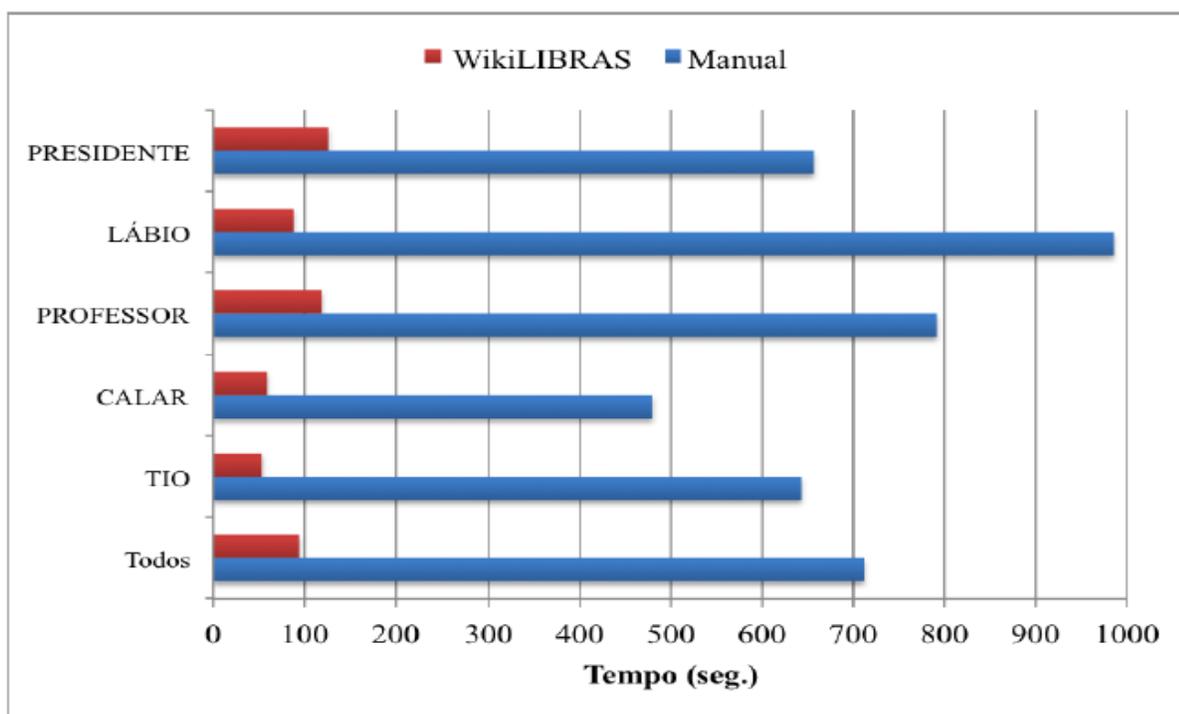


Figura 8: Tempo médio para gerar os sinais corretamente (WikiLIBRAS vs Manual).



Dentre as dificuldades apontadas pelos usuários para gerar os sinais PROFESSOR e PRESIDENTE, destaca-se a dificuldade de alguns usuários em entender alguns parâmetros utilizados na interface Web do WikiLIBRAS. Dessa forma, uma proposta de trabalho futuro é incluir nesta interface, vídeos com intérpretes de LIBRAS para auxiliar os usuários durante a navegação.

Para avaliar se esta diferença com relação ao tempo médio de geração dos sinais no WikiLIBRAS é estatisticamente representativa e, conseqüentemente, testar a hipótese dessa parte do experimento, um teste-t foi novamente aplicado considerando um grau de confiança de 95% e 15 graus de liberdade (número de usuários menos dois, ou seja, $17-2=15$). Os resultados da aplicação desse teste são apresentados na Tabela 8.

Tabela 8: Resultados do teste-t para a Hipótese desse experimento (tempo médio para gerar os sinais)

$\mu_{tempo} (WL)$	93,96 seg.	$\sigma_{tempo} (WL)$	74,57 seg.
$\mu_{tempo} (MAN)$	711,33 seg.	$\sigma_{tempo} (MAN)$	260,33 seg.
Grau de Confiança (p -valor)	0,95	Grau de Liberdade	15
t -valor	12,53	$t_{0,0025,15}$	2,13

De acordo com a Tabela 8, uma vez que o t -valor obtido para o teste foi maior do que o valor crítico para o teste-t com intervalo de confiança de 95% e 15 graus de liberdade ($t_{0,0025,15}$) [Wohlin et al. 2000], é possível concluir que a hipótese H_0 pode ser rejeitada. Dessa forma, é possível afirmar com um grau de confiança de 95% que existe uma diferença significativa no tempo médio de geração dos sinais quando especialistas em LIBRAS utilizam o WikiLIBRAS, quando comparado com o tempo médio de geração desses mesmos sinais através de ferramentas de animação.

Além de produzir os conteúdos num intervalo de tempo significativamente menor, o número de surdos e intérpretes de LIBRAS que existe no Brasil é muito maior do que o número de designers-3D disponíveis, além do que os designers-3D necessitam de vídeos de referências dos sinais ou precisam aprender os sinais em LIBRAS. Dessa forma, é possível criar um Dicionário de LIBRAS de forma mais produtiva utilizando o WikiLIBRAS, especialmente considerando que um Dicionário de LIBRAS possui cerca de 6.000 a 8.000 sinais [Acesso Brasil 2012].

Por fim, a Tabela 9 apresenta a avaliação subjetiva dos principais aspectos do WikiLIBRAS por usuários surdos e intérpretes de LIBRAS. De acordo com a Tabela 9, a usabilidade do WikiLIBRAS obteve o maior valor médio (4,77) dentre os aspectos avaliados. Esse resultado é compatível com os valores médios de avaliação das telas de configuração de mão (4,69), expressão facial (4,92) e ponto de articulação (4,69) do WikiLIBRAS. A tela de orientação da

palma da mão, no entanto, provavelmente causou alguma confusão entre os usuários e obteve um dos piores valores médios (3,92). Conforme mencionado anteriormente, uma das propostas de trabalho futuro para melhorar esse aspecto, é incluir na interface do WikiLIBRAS, vídeos com intérpretes de LIBRAS, para auxiliar os usuários durante a navegação.

Tabela 9: Valores médios para os aspectos avaliados (escalas de 1 a 6)

Questões	Valor Médio	Desvio Padrão
Usabilidade	4,77	1,59
Naturalidade da animação	3,62	1,71
Tela de Configuração de Mão	4,69	1,84
Tela de Orientação	3,92	1,93
Tela de Expressão Facial	4,92	1,04
Tela de Ponto de Articulação	4,69	1,55

A naturalidade da animação 3D obteve o menor valor médio (3,62) dentre os aspectos avaliados. Isso pode ser explicado porque a naturalidade da sinalização baseada em avatares ainda não é comparável com a sinalização humana, especialmente porque os sinais são gerados de forma semi-automática. Além disso, essa pouca naturalidade pode ter influenciado nos resultados dos testes de compreensão e de avaliação da adequação gramatical. Dessa forma, é necessário investir mais esforços para melhorar a flexibilidade e naturalidade do avatar-3D. Uma possível alternativa é integrar ferramentas de captura de movimento, como, por exemplo, o Microsoft Kinect, para capturar os movimentos dos sinais.

5. Considerações

Nesse relatório foi apresentado um processo de experimentação com a comunidade surda para o VLibras e WikiLibras. Nesse processo, foram formuladas hipóteses mais específicas para avaliar o nível de compreensão de usuários surdos nos conteúdos gerados pela solução proposta, para avaliar o desempenho (tempo médio de tradução) e para avaliar a produtividade do WikiLIBRAS no desenvolvimento das construções linguísticas da solução.

Os resultados desse experimento mostraram que três das quatro hipóteses formuladas puderam ser rejeitadas, indicando que a solução proposta é capaz de reduzir as barreiras de acesso a informação, uma vez que ela conseguiu melhorar o nível de compreensão dos conteúdos com relação aos conteúdos com legendas; é capaz de traduzir conteúdos também em cenários que exigem tradução em tempo real, como, por exemplo, na TV; e sua ferramenta de computação humana é capaz de reduzir o tempo médio de produção das suas construções linguísticas, quando comparado com a geração manual dessas construções.

Adicionalmente, conforme evidenciado em trabalhos anteriores [Cox et al. 2002][Kipp et al. 2012][San-segundo et al. 2011], embora as soluções baseadas em avatares não sejam a primeira opção dos surdos que preferem tradução humana, a solução proposta é apresentada como uma alternativa prática e viável, capaz de tornar os conteúdos multimídia acessíveis para surdos, especialmente quando intérpretes humanos não estão disponíveis.

6. Referências

ABNT (2005), 'Abnt nbr 15290 acessibilidade em comunicação na televisão'.

ABNT (2007), 'Abnt nbr 15606-1 televisão digital terrestre - codificação de dados e especificações de transmissão para radiodifusão digital - parte 1: codificação de dados'.

Acesso Brasil (2012), 'Dicionário da língua brasileira de sinais'. URL: <http://www.acessobrasil.org.br/libras/>

Adobe (2012a), 'Adobe flash player 11'. URL: <http://www.adobe.com/br/products/flashplayer.html>

Adobe (2012b), 'Cinemática inversa'. Adobe Flash Professional. URL: http://help.adobe.com/pt_BR/flash/cs/using/WS58BD3A02-DA25-488f-B534-AE5463A24833.html

Ahn, L. V., B. Maurer, C. McMillen, D. Abraham & M. Blum (2008), 'Recaptcha: humanbased character recognition via web security measures', *Science* 321(5895), 1465–1468.

Ahn, L. V. & L. Dabbish (2004), Labeling images with a computer games, em 'Proceedings of the SIGCHI conference on Human factors in computing systems - CHI'04', Vienna, Áustria, pp. 319–326.

Ahn, L. V. & L. Dabbish (2008), 'Designing games with a purpose', *Communications of the ACM* 54(8), 58–67.

Ahn, L. V., R. Liu & M. Blum (2006), Peekaboom: a game for locating objects in images, em 'Proceedings of the SIGCHI conference on Human factors in computing systems - CHI'04', Montreal, Canadá, pp. 55–64.

Anuja, K., S. Suryapriya & S. M. Idicula (2009), Design and development of a frame based mt system for english-to-isl, em 'Proceedings of the World Congress on Nature & Biologically Inspired Computing- NaBIC 2009', Coimbatore, Índia, pp. 1382–1387.

Araújo, T. M. U., F. L. S. Ferreira, D. A. N. S. Silva, E. L. Falcão, L. Dantas, L. A. Domingues, L. H. Lopes, Y. Sato, H. R. Lima, A. N. Duarte & G. L. Souza Filho (2012), 'Accessibility as a service: Augmenting multimedia content with sign language video tracks', *Journal of Research and Practice in Information Technology*.

Batista, L. V. & M. M. Meira (2004), 'Texture classification using the lempel-ziv-welch algorithm', *Lecture Notes in Computer Science* 3171, 444–453.

Biolchini, J., P. G. Milan, A. C. C. Natali & G. H. Travassos (2005), Systematic review in software engineering, Relatório técnico, System Engineering and Computer Science Department COPPE/UFRJ.

Blakowski, G. & R. Steinmetz (1996), 'A media synchronization survey: reference model, specification and case studies', *IEEE Journal on Selected Areas in Communications* 14(1), 5–35.

Bratko, A., B. Filipic, G. V. Cormack, T. R. Lynam & B. Zupan (2006), 'Spam filtering using statistical

data compression models', *Journal of Machine Learning Research* 7(12), 2673–2698.

Brito, L. F. (1995), *Por uma gramática de língua de sinais*, Editora Tempo Brasileiro, Rio de Janeiro, Brasil.

Brown, P. F., V. J. D. Pietra, S. A. D. Pietra & R. L. Mercer (1993), 'The mathematics of statistical machine translation: parameter estimation', *Computational Linguistics* 19(2), 263–311.

Buttussi, F., L. Chittaro & M. Coppo (2007), Using web3d technologies for visualization and search of signs in an international sign language dictionary, em 'Proceedings of the International Conference on 3D Web Technology', Perugia, Itália, pp. 61–70.

Campos, M. B. & L. M. M. Girafa (2000), Signsim: uma ferramenta para auxílio à aprendizagem da língua brasileira de sinais, em 'Proceedings of V Congresso Ibero- Americano de Informática na Educação - RIBIE', Vinãdelmar, Chile, pp. 1–13.

Chiu, Y. H., C. H. Wu, H.Y. Su & C.J. Cheng (2007), 'Joint optimization of word alignment and epenthesis generation for chinese to taiwanese sign synthesis', *IEEE Transactions on Advances in Computers Pattern Analysis and Machine Intelligence* 29(1), 28–39.

Cleary, J.G. & I. H. Witten (1984), 'Data compression using adaptive coding and partial string matching', *IEEE Transactions on Communications* 32(4), 396–402.

Coimbra, T. C. (2011), 'Computação humana aplicada ao problema do caixeiro viajante'. Monografia (Bacharelado em Ciências da Computação), Universidade Federal de Pelotas.

Cox, S., M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt & S. Abbott (2002), Tessa, a system to aid communication with deaf people, em 'Proceedings of the Fifth International ACM Conference on Assistive Technologies - Assets'02', Edinburgh, Escócia, pp. 205–212.

Dangsaart, S., K. Naruedomkul, N. Cercone & B. Sirinaovakul (2008), 'Intelligent thai text - thai sign translation for language learning', *Computers & Education* 51, 1125–1141.

Dasgupta, T. & A. Basu (2008), Prototype machine translation system from text-to-indian sign language, em 'Proceedings of of the 13th International Conference on Intelligent User Interfaces- IUI'08', Gran Canaria, Canada, pp. 313–316.

D'Haro, L. F., R. San-Segundo, R. Córdoba, J. Bugeroth, D. Stein & H. Ney (2008), Language model adaptation for a speech to sign language translation system using web frequencies and a map framework, em 'Proceedings of the Interspeech 2008', Brisbane, Austrália, pp. 2199–2202.

Digital Cinema Initiatives, LLC (2008), 'Digital cinema system specification'.

Dorr, B., P. Jordan & J. Benoit (1999), 'A survey of current paradigms in machine translation', *Advances in Computers* 49, 1–68.

Elliott, R., J. R. Glauert & J. R. Kennaway (2004), A framework for non-manual gestures in a synthetic signing system, em 'Proceedings of the of the Cambridge Workshop Series on Universal Access and Assistive Technology', Cambridge, Reino Unido, pp. 127–136.

FACTory, The (2012), 'The factory'. URL: <http://game.cyc.com>

Felipe, T. A. (2007), *Libras em Contexto : Curso Básico*, 8a edição, WalPrint Gráfica e Editora, Rio de Janeiro, Brasil.

FGV/ABERT (2012), 'Pesquisa sobre tv digital no brasil'. URL:
<http://www.abert.org.br/site/images/stories/pdf/TVProgramacao.pdf>

Fotinea, S. E., E. Efthimiou, G. Caridakis & K. Karpouzi (2008), 'A knowledge-based sign synthesis architecture', *Universal Access in the Information Society* 6, 415–418.

Freitas, C., P. Rocha & E. Bick (2008), Floresta sintá(c)tica: bigger, thicker and easier, em 'Proceedings of the 8th international conference on Computational Processing of the Portuguese Language, PROPOR'08', Aveiro, Portugal, pp. 216–219.

Frigge, M., D. C. Hoaglin & B. Iglewicz (1989), 'Some implementations of the boxplot', *The American Statistician* 43(1), 50–54.

Fusco, E. (2004), *X-libras: Um ambiente virtual para a língua brasileira de sinais*, Dissertação de mestrado, Centro Universitário "Eurípedes de Marília", UNIVEM, Marília, SP.

Gallo, B., R. San-Segundo, J. M. Lucas, R. Barra, L. F. D'Haro & F. Fernández (2009), 'Telecom+d04: Speech into sign language statistical translation system for deaf people', *IEEE Latin America Transactions* 7(3), 400–404.

Gibet, S., T. Leborque & P. F. Marteau (2001), 'High-level specification and animation of communicative gestures', *Journal of Visual Languages and Computing* 12, 657–687.

Goes, M. C. R. (1996), *Linguagem, Surdez e Educação*, Editora Autores Associados, Campinas, Brasil.

Groves, R. M., F. J. Fowler Jr, M. P. Couper, J. M. Lepkowski, E. Singer & R. Tourangeau (2009), *Survey Methodology*, 2a edição, Willey Blackwell, Hoboken, EUA.

Haddon, L. & G. Paul (2001), *Technology and the Market: Demand, Users and Innovation*, ASEAT Conference Proceedings Series, Edward Elgar Publishing, Cheltenham, Reino Unido, capítulo Design in the ICT industry: the role of users., pp. 201–215.

Hanke, T. (2004), Hamnosys - representing sign language data in language resources and language processing contexts, em 'Proceeding of Workshop on the Representation and Processing of Sign Languages - LREC'04', Lisboa, Portugal, pp. 1–6.

Huenerfauth, M. (2008), 'Generating american sign language animation: overcoming misconceptions and technical challenges', *Universal Access in the Information Society* 6(4), 419–434.

Huenerfauth, M., L. Zhao, E. Gu & J. Allbeck (2007), Evaluating american sign language generation through the participation of native asl signers, em 'Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility', Tempe, EUA, pp. 211–218.

IBGE (2000), *Censo demográfico 2000: Características gerais da população*, Relatório técnico, Instituto Brasileiro de Geografia e Estatística. URL:
http://www.ibge.gov.br/home/estatistica/populacao/censo2000/populacao/censo2000_populacao.pdf

IBGE (2010), Censo demográfico 2010: Características gerais da população, religião e pessoas com deficiência, Relatório técnico, Instituto Brasileiro de Geografia e Estatística.

URL:

ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Caracteristicas_Gerais_Religiao_Deficiencia/tab1_3.pdf

IEEE (2012), 'Overview articles'. IEEE Signal Processing Society. URL:

<http://http://www.signalprocessingsociety.org/publications/overview-articles>

IGDA (2012), '2008-2009 casual games white paper'. URL:

http://www.igda.org/sites/default/files/IGDA_Casual_Games_White_Paper_2008.pdf

ISO/IEC (1996), 'So/iec13818-1 tr information technology - generic coding of moving pictures and associated information: part 1: systems'.

ISO/IEC (1998), 'So/iec13818-6 tr information technology - generic coding of moving pictures and associated information: part 6: extension for digital storage media command and control'.

Jain, R. (1991), *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*, Willey, Nova York, EUA.

Kim, S.W., Z. H. Li & Y. Aoki (2004), On intelligent avatar communication using korean, chinese and japanese sign-languages: an overview, em 'Proceeding of 8th International Conference on Control, Automation, Robotics and Vision, ICARCV 2004', Kunming, China, pp. 747–752.

Kipp, M., Q. Nguyen, A. Heloir & S. Matthes (2012), Assessing the deaf user perspective on sign language avatars, em 'Proceeding of 13th International ACM SIGACCESS Conference on Computers and Accessibility', Dundee, Escócia, pp. 1–8.

Kit, C., H. Pan & J. J. Webster (2002), *Translation and Information Technology*, Chinese University of HK Press, Hong Kong, China, capítulo Example-based machine translation: A new paradigm, pp. 57–78.

Kitchenham, B. & S. Charters (2005), *Guidelines for performing systematic literature reviews in software engineering*, Relatório técnico, Keele University and University of Durham.

Knight, K. (1997), 'Automating knowledge acquisition for machine translation', *AI Magazine* 18(4), 81–96.

Koehn, P., F. J. Och & D. Marcu (2003), Statistical phrase-based translation, em 'Proceedings of the Human Language Technology Conference 2003, HLT-NAACL'03', Edmonton, Canadá, pp. 127–133.

Kumar, S., Y. Deng & W. Byrnek (2004), 'A weighted finite state transducer translation template model for statistical machine translation', *Natural Language Engineering* 1(1), 1–41.

Lee, D. G., D. I. Fels & J. P. Udo (2007), 'Emotive captioning', *Computers in Entertainment* 5(2), 3–15.

Lee, S., V. Henderson, H. Hamilton, T. Starner, H. Brashear & S. Hamilton (2005), A gesture based american sign language game for deaf children, em 'Proceeding of the Conference on Human Factors in Computing Systems - CHI'2005', Portland, EUA, pp. 1589–1592.

- Lopez-Ludena, V., R. San-Segundo, R. Martín, D. Sánchez & A. Garcia (2011), 'Evaluating a speech communication system for deaf people', *IEEE Latin America Transactions* 9(4), 565–570.
- Macedo, D.R. (1999), *Sign dic: Um ambiente multimídia para a criação e consulta de dicionários bilíngües de línguas de sinais e línguas orais*, Dissertação de mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, PUC-RS, Porto Alegre, RS.
- Mafra, S. N. & G. H. Travassos (2006), *Estudos primários e secundários apoiando a busca por evidências em engenharia de software*, Relatório técnico, Programa de Engenharia de Sistemas e Computação - COPPE/UFRJ. URL: <http://www.cos.ufrj.br/uploadfiles/1149103120.pdf>
- Mahoui, M., W. J. Teahan, W. J. T. Sekhar & S. Chilukuri (2008), Identification of gene function using prediction by partial matching (ppm) language models, em 'Proceedings of the 17th ACM conference on Information and knowledge management', Napa Valley, EUA, pp. 779–786.
- Medeiros, T. F. L., A. B. Cavalcanti, E. V. C. Lima Borges, I. L. P. Andrezza, B. E. S. Cavalcante & L. V. Batista (2011), Heart arrhythmia classification using the ppm algorithm, em 'Proceedings of the Biosignals and Biorobotics Conference', Vitória, Brasil, pp. 1–5.
- Moffat, A. (1990), 'Implementing the ppm data compression scheme', *IEEE Transactions on Communication* 38(11), 1917–1921.
- Morrissey, S. (2008), *Data-driven machine translation for sign languages*, Tese de doutorado, Dublin City University, Dublin, Irlanda.
- Murakami, M., T. Kuroda, Y. Manabe & K. Chihara (2002), Generation of modifier representation in sign animation, em 'Proceedings of the Fourth International Conference Series on Disability, Virtual Reality and Associated Technologies', Veszprém, Hungria, pp. 27–32.
- Nagao, M. (1984), *Artificial and Human Intelligence*, Elsevier Science Publishers, Amsterdam, Holanda, capítulo A framework of a mechanical translation between Japanese and English by analogy principle, pp. 173–180.
- Othman, A. & M. Jemni (2011), 'Statistical sign language machine translation: from english written text to american sign language gloss', *International Journal of Computer Science Issues* 8(5), 65–73.
- Papineni, K., S. Roukos, T. Ward & W. Zhu (2001), Bleu: a method for automatic evaluation of machine translation, em 'Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics', Philadelphia, EUA, pp. 311–318.
- Quadros, R. M. (1997), *Educação de Surdos - Aquisição da Linguagem*, Editora Artes Médicas, Porto Alegre, Brasil.
- Quinn, A. J. & B. B. Bederson (2011), Human computation: a survey and taxonomy of a growing field, em 'Proceedings of the 2011 annual conference on Human factors in computing systems - CHI'11', Vancouver, Canada, pp. 1403–1412.
- Russell, S. & P. Norvig (2004), *Inteligência Artificial*, 2a edição, Editora Campus, Rio de Janeiro, Brasil.
- Sagawa, H., M. Ohki, E. Oohira, H. Ikeda & H. Fujisawa (1996), 'Pattern recognition and synthesis for

a sign language translation system', *Journal of Visual Languages & Computing* 7(1), 109–127.

San-segundo et al., R. (2008a), 'Proposing a speech to gesture translation architecture for spanish deaf people', *Journal of Visual Languages and Computing* 19(5), 523–538.

San-segundo et al., R. (2008b), 'Speech to sign language translation system for spanish', *Speech Communication* 50(11), 1009–1020.

San-Segundo, R., A. Pérez, D. Ortiz, L.F. D'Haro, M.I. Torres & F. Casacuberta (2007), Evaluation of alternatives on speech to sign language translation, em 'Proceedings of the Interspeech 2007', Antwerp, Bélgica, pp. 2529–2532.

San-segundo, R., J. M. Montero, R. Córdoba, V. Sama, F. Fernández, L. F. D'Haro, V. López-Ludeña, D. Sánchez & A. Garcia (2011), 'Design, development and field evaluation of a spanish into sign language translation system', *Pattern Analysis and Applications* 15(2), 203–224.

San-Segundo, R., R. Barra, L.F. D'Haro, J. M. Montero, R. Córdoba & J. Ferreiros (2006), A spanish speech to sign language translation system for assisting deaf-mute people, em 'Proceedings of the Interspeech 2006', Pittsburgh, EUA, pp. 1399–1402.

Santos, G. S., M. S. Silveira & S. M. Aluisio (2009), Produção de textos paralelos em língua portuguesa e uma interlíngua em libras, em 'Proceedings of XXXVI Seminário Integrado de Software e Hardware - SEMISH'09', Bento Gonçalves, Brasil, pp. 371–385.

Sato, S. (1993), Example-based translation of technical terms, em 'Proceedings of Fifth International Conferente on Theoretical and Methodological Issues in Machine Translation', Kyoto, Japão, pp. 58–68.

Soares, L. F. G. & S. D. J. Barbosa (2005), Programando em NCL 3.0: Desenvolvimento de aplicações para o middleware Ginga, Editora Campus, Rio de Janeiro, Brasil.

Souza Filho, G. L., L. E. C. Leite & C. E. C. F. Batista (2007), 'Ginga-j: The procedural middleware for the brazilian digital tv system', *Journal of the Brazilian Computer Society* 12(4), 47–56.

Starner, T., A. Pentland & J. Weaver (1998), 'Real-time american sign language recognition using desk and wearable computer based video', *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(12), 1371–1375.

Stokoe, W. C. (1980), 'Sign language structure', *Annual Review of Anthropology* 9, 365–390.

Stumpf, M. R. (2000), Língua de sinais: escrita dos surdos na internet, em 'Proceedings of V Congresso Ibero-Americano de Informática na Educação', Viñadelmar, Chile, pp. 1–8.

Su, H. Y. & C. H. Wu (2009), 'Improving structural statistical machine translation for sign language with small corpus using thematic role templates as translation memory', *IEEE Transactions on Machine Translation, Audio, Speech, and Language Processing* 17(7), 1305–1315.

Surowiecki, J. (2005), *Wisdom of Crowds*, Anchor, Nova York, EUA. Sutton, V. (1995), *Lessons in Sign Writing, Textbook and Workbook*, 2a edição, Deaf Action Committee for SignWriting, La Jolla, EUA.

va Zijl, L. & G. Olivrin (2008), South african sign language assistive translation, em 'Proceedings of the

IASTED International Conference on Telehealth/Assistive Technologies', Baltimore, EUA, pp. 3–7.

Veale, T., A. Conway & B. Collins (1998), 'The challenges of cross-modal translation: English to sign language translation in the zardoaz system', *Machine Translation* 13(1), 81–106.

Wauters, L. N. (2005), Reading comprehension in deaf children: The impact of the mode of acquisition of word meanings, Tese de doutorado, Radboud University, Nijmegen, Holanda.

WHO (2012), 'Deafness and hearing impairment'. World Health Organization. URL: <http://www.who.int/mediacentre/factsheets/fs300/en/>

Wohlin, C., P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell & A. Wesslén (2000), *Experimentation in Software Engineering: An Introduction*, Kluwer Academic Publisher Norwell, EUA.

Woodward, J.C. (1973), 'Some observation on sociolinguistic variation and american sign language', *Kansas Journal of Sociology* 9(2), 191–200.

Wu, C. H., H. Y. Su, Y. H. Chiu & C. H. Lin (2007), 'Transfer-based statistical translation of taiwanese sign language using pcfgr', *ACM Transactions on Asian Language Information Processing* 16(1), 1–18.

Yuen, M. C., L. J. Chen & I. King (2009), A survey of human computation systems, em 'Proceedings of the International Conference on Computational Science and Engineering -CSE '09', Vancouver, Canadá, pp. 723–728.

Zhang, Q., L. Cheng & R. Boutaba (2010), 'Cloud computing: state-of-the-art and research challenges', *Journal of Internet Services and Applications* 1, 7–18.

Zhao, L., K. Kipper, W. Schuler, C. Vogler, N. Badler & M. Palmer (2000), Machine translation system from english to american sign language, em 'Proceedings of the 4th Conference of the Association for Machine Translation in the Americas', Cuernavaca, Mexico, pp. 54–67.