

Produção de Textos Paralelos em Língua Portuguesa e uma Interlíngua de LIBRAS

Guilherme Spolavori Santos¹, Milene Selbach Silveira¹, Sandra Maria Aluísio²

¹ PPGCC - Faculdade de Informática – PUCRS - Porto Alegre – RS – Brasil

² Instituto de Ciências Matemáticas e de Computação, USP - São Carlos - SP - Brasil

{guilherme.santos,milene.silveira}@pucrs.br, sandra@icmc.usp.br

***Abstract.** This paper describes the adaptation of a text simplification architecture for text production in a Brazilian Sign Language Interlingua. Therefore were done studies about Sign Language and Text Simplification, as well about a tool used as work platform. As result, besides the adapted architecture, the discussion about its use is presented.*

***Resumo.** Este artigo descreve a adaptação de uma arquitetura de simplificação textual para produção de textos em uma interlíngua da língua de sinais usada no Brasil. Para isso, foram estudados conceitos sobre língua de sinais e simplificação de textos, bem como um sistema usado como plataforma de trabalho. Como resultado, além da arquitetura adaptada, é apresentada uma discussão sobre seu uso.*

1. Introdução

Simplificação Textual (ST) é uma área de pesquisa emergente do Processamento de Língua Natural (PLN) chamada geração texto-para-texto. O objetivo da ST é maximizar a compreensão de textos escritos pela simplificação de sua estrutura lingüística. Como resultado, espera-se que todo o texto possa ser mais facilmente compreendido (SIDDHARTHAN, 2003; MAX, 2006). Inicialmente, o processo de simplificação de textos foi pensado para diminuir a complexidade de um texto e, assim, diminuir o custo computacional do seu processamento. Entretanto, a aplicação desse processo estendeu-se para a resolução de problemas de cunho social mais acentuado. Se um texto simplificado pode ser mais adequadamente aplicado a uma máquina que a sua versão original, a compreensão desse exigirá menos de pessoas com dificuldades em relação à língua. Isso caracteriza essa tarefa como uma forma de promover o acesso à informação para pessoas com baixos níveis de letramento ou com problemas cognitivos, por exemplo.

Assim, os resultados das pesquisas nessa linha podem ajudar a promover o acesso participativo do cidadão brasileiro ao conhecimento, o qual é um dos grandes desafios da computação no Brasil para os próximos anos (SBC, 2006). No âmbito desse desafio, Campos e Silveira (2007), por exemplo, apresentam idéias que dizem respeito à Comunidade Surda, especificamente, e suas necessidades quanto à apropriação e divulgação de sua língua. Diversas iniciativas já foram desenvolvidas para difundir as línguas de sinais (CAMPOS, 2001; SOUZA e VIEIRA, 2006, por exemplo). Normalmente, o objetivo dessas aplicações é familiarizar o surdo com a escrita de sinais, sem uma preocupação aprofundada com a estrutura lingüística da língua de

sinais. A língua de sinais, como qualquer linguagem natural, pode ser tratada computacionalmente, apresentando problemas semelhantes aos de uma linguagem oral qualquer. Explorar essas características estruturais pode enriquecer a aprendizagem de outras línguas – como o português falado, por exemplo.

Nesse sentido, uma ferramenta computacional que aborde a complexidade da língua de sinais vem a auxiliar a comunidade surda na inclusão digital e no acesso à informação. Abordagens estatísticas para a construção desse tipo de ferramenta, por exemplo, necessitam de um grande conjunto de textos paralelos sobre os quais trabalhar. Considerando a Comunidade Surda, e os textos por ela produzidos para uso como base para esta abordagem, surge o desafio abordado aqui: realizar a aquisição dos textos no idioma desejado, ou seja, em LIBRAS, e a versão paralela destes textos em Língua Portuguesa.

Com isso, esse trabalho trata da produção de textos paralelos (na versão da Língua Portuguesa e na versão em LIBRAS) como forma de abastecer um repositório de dados para trabalhos futuros na área de processamento de língua natural envolvendo a língua dos surdos brasileiros. Para esse fim, uma abordagem baseada na arquitetura de um sistema de apoio à anotação de simplificação de texto – parte do projeto PorSimples (Simplificação Textual do Português para inclusão e Acessibilidade Digital) (ALUÍSIO et al, 2008) – foi adotada para a produção desses textos a partir do texto em português já simplificado. Em consequência disso, agrega-se uma nova funcionalidade que contribui para o grupo de ferramentas envolvidas no PorSimples com a criação de um módulo para reescrita do texto em uma interlíngua da LIBRAS, que possibilitaria adicionar outro grupo de usuários beneficiados no que se refere ao acesso a informação, desde que o público alvo em especial do PorSimples são os analfabetos funcionais e pessoas com problemas cognitivos (como afasia e dislexia).

2. Simplificação Textual

Segundo Max (2006), a simplificação textual é definida como o processo que transforma um texto procurando reduzir a complexidade de seu vocabulário e da estrutura de suas sentenças (complexidade léxica e sintática, respectivamente), tentando preservar o seu significado e o seu conteúdo. Esse conceito pode ser ilustrado com o exemplo a seguir (SPECIA et al, 2008):

- (1) "Compete à CPI condenar ou absolver", disse o presidente da Câmara, que se disse "surpreendido" com a renúncia de Valdemar e elogiou sua "bravura".
- (2) "Compete à CPI condenar ou absolver", disse o presidente da Câmara. O presidente se disse "surpreendido" com a renúncia de Valdemar e elogiou sua "bravura".

A sentença apresentada em (1) é formada por uma oração principal e uma oração subordinada relativa, sendo ambas ligadas por pronome relativo. A simplificação, nesse caso, tomou como base regras pré-definidas para separar a sentença no seu ponto de articulação – o pronome “que” – gerando, em (2), uma sentença apenas com a oração principal e outra sentença com a oração subordinada, a qual terá como sujeito o núcleo do termo da oração principal (SPECIA et al, 2008).

Dentre as utilidades relacionadas à tarefa da simplificação textual, Siddharthan (2002) destaca o pré-processamento e o apoio a usuários. No primeiro caso, o processamento do texto (em um *parser*) pode tornar-se mais leve quando apoiado pela simplificação em um estágio anterior, considerando que sentenças longas são problemáticas nessa tarefa ao contrário de sentenças menores como o caso das geradas na simplificação. O segundo caso trata a simplificação textual como ferramenta de apoio a usuários no que se refere à leitura de textos. Esses usuários podem apresentar algum tipo de limitação que dificulte a compreensão do conteúdo do texto como, por exemplo, é o caso de pessoas afásicas (SIDDHARTHAN, 2002). Há ainda usuários que estão aprendendo um segundo idioma, para os quais, inicialmente, textos simplificados podem facilitar o acesso à língua alvo, dado que estruturas de texto mais complexas podem dificultar a compreensão do texto (PETERSEN & OSTENDORF, 2007).

Pesquisas relacionadas, como a de Chandrasekar et al (1996), tratam a simplificação textual como um processo de dois estágios: análise e transformação. O primeiro faz uma análise das sentenças devolvendo uma representação estrutural dessas. O segundo, por sua vez, aplica uma seqüência de regras que identificam e extraem os elementos a serem simplificados a partir da representação obtida no estágio de análise. Entretanto, pesquisas mais recentes, como a publicada por Siddharthan (2002 e 2006), apresentam um terceiro estágio: regeneração, visto que a transformação em si não garante a coesão entre o texto original e o texto simplificado. A Figura 1 apresenta estes estágios em uma arquitetura para o processo de simplificação.

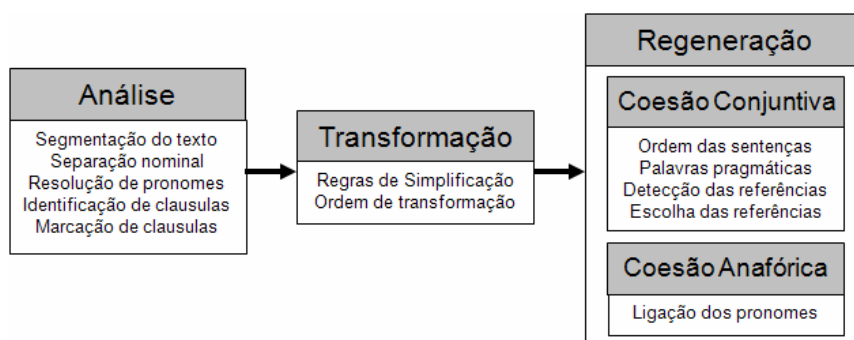


Figura 1. Estágios da simplificação textual conforme Siddharthan (2006)

Considerando que a simplificação do texto opera individualmente sobre cada sentença, o estágio de análise deve identificar as estruturas que serão simplificadas (as cláusulas, grupo de palavras ou sintagmas) e definir adequadamente os pontos de articulação onde as cláusulas, por exemplo, serão logicamente marcadas e referenciadas. A identificação das *cláusulas* ocorre com base nas informações etiquetadas no nível morfosintático (*part-of-speech tagging*) ou sintático (*parsing*). Nesses são definidas as marcas referentes às expressões que delimitam sintagmas nominais e sintagmas verbais, como conjunções, pronomes, nomes, etc.

A partir disso, o estágio de transformação toma como base as sentenças pré-processadas no estágio anterior. É a partir da marcação dos termos que delimitam as orações e dos termos que relacionam as orações que se define o conjunto de regras necessárias a serem aplicadas para simplificar a sentença (CHANDRASEKAR et al, 1996). Exemplo disso é a regra citada em (3):

$$(3) \quad X:NP, \text{ RelPron } Y, Z \rightarrow W X:NP Z. X:NP Y$$

Essa regra pode ser interpretada considerando que, se uma sentença inicia com um sintagma nominal (X:NP) e é seguida de uma estrutura com pronome relativo, na forma de (, RelPron Y ,) e seguida por (Z), sendo (Y) e (Z) seqüências quaisquer de palavras, então a sentença pode ser simplificada gerando duas novas sentenças: (X) seguida por (Z) e (X) seguida por (Y). É possível que ocorra a necessidade de simplificar mais de uma construção na mesma sentença. Nesse caso, as transformações devem seguir uma ordem predefinida, considerando as regras mais freqüentemente usadas como prioritárias, evitando, assim, transformações improdutivas.

Mesmo com a correta aplicação das regras de transformação, existem riscos provenientes da alteração da estrutura sintática da sentença que podem prejudicar a clareza do significado para o leitor. Em alguns casos, esses riscos são uma questão de estilo de escrita, como a forma de reescrever os sintagmas nominais referenciados quando uma sentença é dividida, por exemplo. Há ainda casos mais complexos em que é necessário detectar e escolher o marcador discursivo (*cue-word*) mais adequado para manter a relação entre duas partes de uma sentença dividida. Sendo assim a fase de transformação é apoiada por um processo de coesão conjuntiva (conforme a Figura 1) com a intenção de manter a coesão das sentenças reescritas. Uma vez ocorridas as substituições dos sintagmas nominais, pode haver uma retomada, total ou parcial, de outras estruturas no texto. Sendo assim, a etapa final da fase de regeneração trata da reorganização da estrutura anafórica dos textos, visto que durante o processo de reescrita do texto pode ocorrer uso de pronomes de forma equivocada.

3. Língua de Sinais

A Língua de Sinais é a língua natural do surdo e trata da substituição de sons por gestos, os quais compõem a principal forma de comunicação utilizada entre pessoas surdas ou entre pessoas com algum outro tipo de problema auditivo. Quanto à estrutura, essa língua é composta por níveis lingüísticos variados, tais como: fonologia, morfologia, sintaxe e semântica. Da mesma forma que nas línguas oral-auditivas existem palavras, nas línguas de sinais também existem itens lexicais, que recebem o nome de sinais, mas que se diferenciam por pertencerem à modalidade visual-espacial. Com isso, para comunicar-se usando Língua de Sinais, não basta apenas conhecer os sinais: é necessário conhecer a sua gramática para estruturar as frases, estabelecendo, dessa maneira, a comunicação.

Da mesma forma que existem diferentes idiomas para línguas faladas oralmente, existem diferentes línguas de sinais correspondentes aos falados em cada país, havendo igualmente variações dentro delas, assim como há regionalismos e dialetos em línguas orais. Essas variações se devem a culturas diferentes e a influências diversas no sistema de ensino, por exemplo. No Brasil, isso não é diferente, e a Língua Brasileira de Sinais (LIBRAS) é reconhecida como meio legal de comunicação e expressão entre as pessoas surdas (FENEIS, 2002).

3.1. Aspectos Lingüísticos da LIBRAS

A LIBRAS apresenta uma formação gramatical tão completa e complexa quanto à do português falado e escrito. Essa formação vem desde a estruturação de sinais por meio

de itens léxicos específicos, bem como a aplicação dessas estruturas por mecanismos morfológicos, sintáticos e semânticos. Se pensarmos em um sinal, que usa recursos gestuais e espaciais e não sonoros, as unidades mínimas são referentes às configurações de mão, movimentos, pontos de articulação ou expressões faciais.

Entretanto, formar um sinal não trata apenas de combinar movimentos com configurações de mão ou expressões faciais. Existem restrições que permitem uma formação de unidades mínimas qualquer ser aceita ou não e isso varia de acordo com cada língua de sinais. A formação de sinais em LIBRAS pode ocorrer de forma semelhante ao português como é o caso da derivação. Exemplo disso é o caso do sinal que representa a palavra “bonitinho”, que deriva da palavra “bonito”, a partir do acréscimo da expressão facial referente a “olhos arregalados”. Ainda é possível a formação de sinais por meio da junção de sinais mais simples em formas compostas. Os sinais “casa” e “cruz”, que quando juntos formam o sinal “igreja”, ilustram a formação de sinais por composição.

Em relação à classe verbal, existem duas classificações: verbos que possuem uma marca de concordância e verbos que não a possuem. Um verbo que apresenta concordância verbal necessita de uma marcação de forma a identificar quem é sujeito e quem é objeto em uma frase, enquanto um verbo que não possui concordância não utiliza esse recurso posicional. A Figura 2 ilustra a marcação no espaço de sinalização que define a concordância de um determinado verbo.

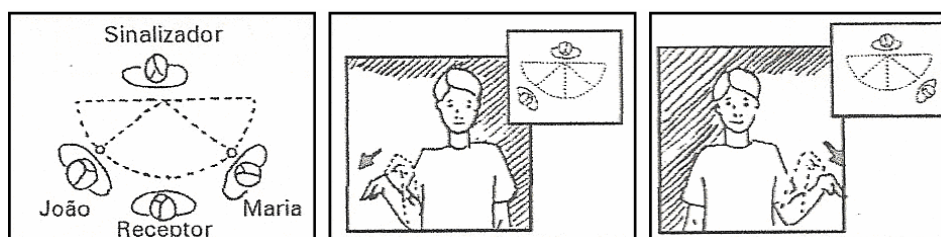


Figura 2. Concordância Verbal (QUADROS e KARNOPP, 2004).

Conforme Quadros e Karnopp (2004), em relação à ordem das frases na LIBRAS, há a ocorrência de uma ordenação mais básica que as demais: a ordem Sujeito-Verbo-Objeto (SVO). Essa ordem fica evidenciada em orações simples ou em orações complexas, com ocorrência de subordinação, como ocorre nos exemplos a seguir:

JOÃO GOSTAR MARIA - João gosta da Maria

EU ACHAR MARIA IR EMBORA - Eu acho que a Maria foi embora

Sendo a LIBRAS uma linguagem com base em atributos visuais, é comum a ocorrência de marcas não manuais como direção do olhar, da cabeça ou do corpo como recurso para a concordância nas frases. Segundo Quadros e Karnopp (2004), a marcação não manual de concordância torna a frase mais carregada, forçando mudanças na ordem da frase e gerando estruturas diferentes. Exemplo disso são as ordens SOV e OSV que só ocorrem quando a concordância é feita por marcadores não manuais. Caso contrário, essas construções são consideradas agramaticais e, como essas marcas não são obrigatórias na ordem SVO, justifica-se o fato dessa ser apontada como a ordem básica da frase em LIBRAS. Os exemplos a seguir apresentam as ordens SOV e OSV respectivamente:

JOÃO FUTEBOL GOSTAR (SOV) - João gosta de futebol

FUTEBOL JOÃO GOSTAR (OSV) - João gosta de futebol

Além disso, um mecanismo gramatical denominado de topicalização é considerado responsável por uma flexibilização na ordem da frase. Esse recurso é muito comum na LIBRAS e consiste em evidenciar um determinado tema, posicionado no início da frase, e o suceder com comentários a respeito. Na LIBRAS é possível topicalizar tanto o sujeito quando o objeto (eventualmente ambos). A seguir são apresentados alguns exemplos de topicalização:

<FRANÇA> EU IR (apenas o objeto topicalizado)

<EU FRANÇA> IR (sujeito e objeto topicalizados)

<MARIA> JOÃO GOSTA ELA (tópico sem ligação)

3.2. Aquisição da Linguagem e da Escrita

É importante ressaltar que o surdo não está necessariamente restrito a uma única forma de comunicação. Uma educação bilíngüe permite ao surdo compreender, ao menos, duas línguas que podem coexistir no mesmo ambiente de aprendizagem. Além disso, esta modalidade de educação não apenas define qual língua deve ser usada como primeira língua (L1) e qual deve ser usada como segunda língua (L2), como também indica as funções de cada uma delas no ambiente social do surdo (QUADROS, 1997).

Segundo QUADROS e SCHMIEDT (2006), atualmente, a aquisição do português escrito pelos surdos ocorre por um processo semelhante à aquisição do escrito por ouvintes que já dominam o falado. Contudo, conforme Quadros (1997), a criança surda, ao estabelecer relações com as letras e as palavras do português, sofre uma interrupção nesse processo, já que tal sistema escrito não consegue expressar a língua de sinais. Dessa forma, o aprendizado de português pelo surdo requer primeiro a apropriação de sua própria língua (LIBRAS), sendo isto fundamental nesse processo de aprendizagem.

Nesse sentido, um sistema de escrita que se aproxime da língua de sinais é uma porta que se abre no processo de letramento do surdo que domina a língua de sinais usada no país. Existem sistemas não alfabéticos capazes de representar as unidades morfológicas para escrita de línguas de sinais – tais como o SignWriting (SIGNWRITING, 2008) e o Elis (ESTELITA, 2008), por exemplo – e que ainda não são muito difundidos. Isso faz com que a atividade da escrita em português, como parte da aprendizagem da segunda língua, implique a ocorrência de um fenômeno natural interpretado como a formação de uma interlíngua entre a L1 e a L2. Finau (2007) analisa a formação dessa interlíngua como sistema lingüístico organizado em uma “Gramática Mental” que sofre influência da língua nativa (L1) do aprendiz e da língua alvo (L2). Um exemplo do uso dessa versão escrita é observado em Estelita (2008), que descreve um sistema de notação para escrita de sinais e usa uma interlíngua para relacionar essa escrita (em LIBRAS) e escrita na Língua Portuguesa. De uma forma geral, essa interlíngua da LIBRAS seria o uso de um vocabulário lematizado do português organizado na estrutura da sintaxe da LIBRAS. A seguir, um exemplo da interlíngua utilizada pela autora.

[Interlíngua LIBRAS]

Lara ir junto mamãe praia. Lara brincar areia ver assustar tartaruga grande, sair correr chorar junta mamãe.

[Português]

Lara foi junto com sua mãe para a praia. Lara estava brincando na areia quando se assustou ao ver uma tartaruga muito grande e saiu chorando correndo para junto de sua mãe.

4. Ferramentas para Simplificação Textual de Português

Conforme citado na Introdução, uma das utilidades da simplificação textual é facilitar o acesso à informação para pessoas com algum tipo de necessidade especial no que se refere à leitura. Nesse contexto, o projeto PorSimples (ALUÍSIO et al, 2008) diz respeito ao desenvolvimento de uma tecnologia de apoio ao acesso à informação, através de simplificação de textos na língua portuguesa do Brasil. Dentre as diversas ferramentas em desenvolvimento neste projeto, nesse capítulo serão apresentados o Editor de Anotações de Simplificação e o Portal de Corpora Simplificados, por estarem diretamente relacionados ao objetivo do trabalho aqui apresentado.

4.1. Editor de Anotação de Simplificação

O objetivo do Editor de Anotação de Simplificação é dar suporte à criação de um corpus paralelo de simplificação. Especificamente, esta ferramenta trata do estudo de características que tornam um texto complexo nos aspectos léxico e sintático, realizando um levantamento dessas características por meio de um *parser* e listas de palavras simples, oferecendo apoio ao ser humano na tarefa de anotação de simplificação. A Figura 3 mostra os passos pelos quais o texto passa em seu processo de simplificação pelo usuário.

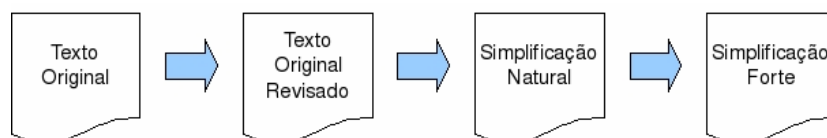


Figura 3. Processo de simplificação (SPECIA et al 2008)

A ferramenta funciona de forma que, em um primeiro momento, dado um texto original como entrada, o usuário tem a opção de fazer ajustes nele para corrigir eventuais erros gramaticais ou de pontuação a fim de não prejudicar a análise sintática a ser feita. É importante ressaltar que uma entrada que necessita de muitos ajustes deve ser descartada, considerando a má qualidade do texto. No passo seguinte, o usuário faz a simplificação manual do texto seguindo recomendações gerais (como encurtar as sentenças e seguir a ordem Sujeito Verbo Objeto (SVO) nas sentenças, por exemplo), mas sem qualquer compromisso quanto às regras que serão aplicadas na simplificação. A esse processo é dado o nome de simplificação natural. Neste passo, o linguísta pode fazer uso do “Léxico” como apoio para ressaltar as palavras consideradas complexas e os marcadores discursivos ambíguos ou não usuais. Essa classificação ocorre a partir de um dicionário infantil (Biderman, 2005), de uma lista de palavras concretas de (Janczura et al., 2007) e do corpus de textos do jornal Zero Hora, coluna “Para seu filho ler”,

podendo assim ser customizada com outros lexicos. A ferramenta dá suporte à anotação de cada operação aplicada pelo usuário, de acordo com as regras predefinidas, e armazena essa informação junto com cada sentença (na versão original e simplificada). O último passo é denominado de Simplificação Forte, quando o usuário é obrigado a usar as regras de simplificação sintática descritas em um manual de simplificação criado no escopo do projeto (SPECIA et al, 2008). Nesse processo, além do “Léxico”, o usuário tem o auxílio do parser Palavras (BICK, 2000) que tem a funcionalidade de ressaltar os pontos em que o texto deve ser alterado para torná-lo mais simples.

A interface de Anotação de Simplificação (Figura 4) permite que o usuário simplifique as sentenças separadamente, dado que cada conjunto de sentenças simplificadas (correspondente a uma sentença original) deve ser introduzido manualmente na caixa de texto destinada ao texto simplificado (à direita). O usuário é auxiliado por uma funcionalidade que resalta palavras complexas no texto, sugerindo uma substituição. No canto inferior direito da tela (podendo mudar de posição), fica disponível a caixa de texto com as informações referentes à lista de regras já aplicadas para a simplificação de uma determinada sentença. De forma semelhante, a simplificação forte pode ser anotada e registrada a partir da versão de simplificação natural. Uma diferença é a funcionalidade que resalta os trechos que prejudicam a compreensão de um texto e que precisam ser simplificados.



Figura 4. Interface de Anotação de Simplificação.

4.2. Portal de Corpora Simplificados

O Portal de Corpora Simplificados compreende dois objetivos principais: o primeiro consiste na criação de um Portal web para consulta de informações sobre o processo de simplificação no corpus paralelo, e o segundo diz respeito à geração de arquivos com anotação linguística no formato de intercâmbio internacional XCES¹ para o corpus

¹ <http://www.xces.org/>

paralelo. Os arquivos gerados possuem informações gerais sobre o texto e sobre o corpus, sobre como estão organizadas as partes que compõem o texto (como parágrafos e sentenças), sobre *tokens* (palavras) e suas características gramaticais, sobre os alinhamentos entre sentenças de textos simplificados, sobre operações de simplificação ocorridas, bem como sobre o próprio texto em si.

5. Módulo de Reescrita em LIBRAS

Com base na arquitetura do Editor de Anotação de Simplificação descrito no capítulo anterior, foi realizada uma adaptação desse sistema, inserindo-se um módulo de produção de textos em LIBRAS². Assim, fazendo uso dessa arquitetura já homologada, a produção do texto é abordada como um processo de transformação do texto em português para uma versão em uma interlíngua de LIBRAS, com as devidas anotações referentes às modificações de cada sentença.

Esse módulo receberá como entrada um texto já simplificado e, a partir de operações realizadas, irá gerar uma nova versão do texto considerando os aspectos lingüísticos da Língua Brasileira de Sinais. Uma vez que o desafio deste módulo envolve a abordagem de simplificação textual para reescrita em LIBRAS, a interlíngua adotada para representar a língua de sinais tem o objetivo de aproximar aspectos da LIBRAS à forma como se organizam textos escritos em língua oral, para não prejudicar a análise do corpus.

Para isso, o processo elaborado é definido em duas fases: Análise e Transformação. A fase de Análise não diverge do processo que já ocorre no Editor de Anotação de Simplificação, ou seja, o texto é enviado ao *parser* Palavras e o retorno, devidamente analisado, é armazenado na base de dados. A diferença é que esse processo ocorrerá sobre a versão de Simplificação Forte, que não possui versões posteriores de texto. Essa escolha baseia-se na idéia de que é a versão que necessita de menos operações de simplificação, influenciando, assim, que o uso das operações ocorra em um contexto de reescrita para a LIBRAS. Já a fase de Transformação registra as regras que modificam cada sentença de forma que estas sigam a estrutura sintática da LIBRAS em uma representação linear com a lematização das palavras da sentença. Com o objetivo de não alterar o fluxo original de criação das produções no Editor de Anotação de Simplificação, o uso do módulo de reescrita em LIBRAS é opcional no sistema. Dessa forma, o fluxo de criação de produções foi alterado conforme a Figura 5.

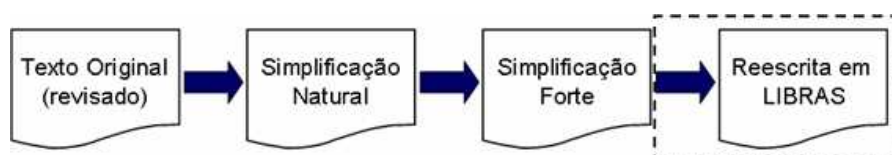


Figura 5. Novo Fluxo para criação de produções.

Para novo módulo, o sistema carrega uma interface semelhante à apresentada no capítulo anterior, contudo modificada para atender às necessidades da reescrita em LIBRAS. Sendo assim, o usuário realiza a reescrita considerando uma sentença original por vez. A Figura 6 exemplifica o processo de reescrita do texto em LIBRAS. Mais

² <http://caravelas.icmc.usp.br/anotlibras/>

especificamente, cada sentença original em 1, deve ser reescrita em 2 conforme a notação de interlíngua adotada para a produção dos textos. Para complementar, o usuário deve informar as operações realizadas sobre a sentença em 1, sendo essas apresentadas em 3. Além disso, a interface permite reescrever o título do texto (campo superior em 2) da mesma forma que o processo de reescrita do texto descrito anteriormente.



Figura 6. Interface para reescrita em LIBRAS.

Diferente do processo de simplificação, a reescrita em uma interlíngua da LIBRAS aproxima-se de uma tradução entre a Língua Portuguesa e a LIBRAS. Assim, foi necessário um levantamento – com apoio de um especialista em LIBRAS – de características que relacionam as duas línguas nesse processo de tradução. A partir daí, foi possível observar que operações já utilizadas no processo de simplificação também seriam utilizadas para relacionar a reescrita em LIBRAS. Além disso, decorrente desse levantamento, novas operações específicas para o uso da reescrita em LIBRAS foram criadas. A seguir, são descritas as operações definidas para produção de textos na interlíngua da LIBRAS.

- **Reescrita em LIBRAS:** é sugerida uma nova versão para a sentença, processando a sentença original e reescrevendo-a automaticamente na versão em LIBRAS. O processamento dá-se retirando as palavras referentes a preposições e a artigos, usando a forma lematizada (informação retornada do PALAVRAS) das palavras restantes. É importante ressaltar que o usuário fica livre para alterar a sugestão gerada automaticamente assim como pode incluir comentários sobre essas alterações;
- **Reescrita Forte:** são feitas modificações complexas na sentença de forma que a sentença mantém apenas o sentido da sentença original. A reescrita forte apresenta características de outras operações como a divisão da sentença ou a substituição lexical, por exemplo;
- **Mudança de Voz:** é registrado mudança de voz (sempre de passiva para ativa), dado que a estrutura de frase da LIBRAS não permite ocorrência de voz passiva;

- Inversão na ordem da sentença: é modificada a ordem de alguns elementos da sentença, visto que, em alguns casos, a versão em LIBRAS pode ser melhor expressada;
- Divisão de sentença: são divididas sentenças em que o tipo de relação entre as frases torna complicado expressar a informação original na versão em LIBRAS.
- União de sentença: são unidas sentenças de forma que a informação possa ser expressada de uma forma direta, seguindo a tendência das estruturas de LIBRAS;
- Remoção de sentença e Remoção de parte da sentença: são removidas sentenças ou parte delas quando há ocorrência de redundância mantendo a forma direta de se expressar em LIBRAS;
- Substituição Lexical: são substituídas palavras com a finalidade de adequar o sentido da sentença reescrita em LIBRAS dado que seu vocabulário é consideravelmente menor que o da Língua Portuguesa,
- Substituição Datilológica: são representados em forma de soletração manual (através de configurações de mão que representam as letras do português) termos técnicos ou nomes que não existem em forma de sinais.

Cada operação possui um comportamento no que se refere ao número de ocorrências da sentença no texto durante o processo de alinhamento (a cardinalidade). Uma vez finalizada a reescrita do texto, o sistema roda o procedimento de alinhamento considerando a cardinalidade de cada operação atribuída a cada sentença do texto. A partir dessa informação se define quantas sentenças são produzidas no texto reescrito a partir da sentença original.

6. Discussão sobre o Uso das Operações

Buscando validar o uso de uma abordagem de simplificação textual para produção textual em língua de sinais, tendo como resultado um corpus alinhado sentencialmente e devidamente anotado, uma experiência de uso foi realizada por uma especialista em LIBRAS. Ela possui nível superior (Bacharelado em Artes Plásticas), tendo ampla formação como intérprete de LIBRAS, além de uma experiência de 20 anos, enquanto intérprete, junto às comunidades surdas, para a formação de jovens e crianças surdos, na educação infantil e nos ensinos fundamental e médio. Também atua junto a adultos ouvintes nos cursos de formação ministrados na área da surdez, no ensino superior.

A atividade definida na experiência de uso trata da reescrita de textos a partir de uma versão de simplificação forte do texto para uma interlíngua da LIBRAS. Para este fim, o especialista reescreveu 10 textos selecionados aleatoriamente do corpus do jornal Zero Hora, disponível na base de dados do sistema. Todos os textos já possuíam suas versões original, natural e forte e, para cada texto, o especialista o reescreveu sentença por sentença e registrou quais das operações disponíveis se aplicaria àquela reescrita.

No que se refere ao uso das operações, diferentes aspectos foram observados. Primeiramente, houve a preocupação em verificar se a opção por limitar a reescrita em LIBRAS apenas a versões de simplificação forte não seria prejudicial à produção dos textos. Isso se deve ao fato de que houve muitas ocorrências de união de sentenças no

momento da reescrita, que já tratavam do resultado de uma divisão no processo de simplificação forte. A Tabela 1 exemplifica esse tipo de ocorrência.

Tabela 1. Exemplo de união de sentenças.

Simplificação Natural	O policial militar Danubio Lisboa, 33 anos, por exemplo, precisava estar às 10h em Brasília para uma reunião.
Simplificação Forte	O policial militar Danubio Lisboa precisava estar em Brasília às 10h para uma reunião. Danubio Lisboa tem 33 anos.
Interlíngua LIBRAS	<i>Policial militar DANUBIO LISBOA precisar Brasília 10h reunião, ele idade 33.</i>

Ao mesmo tempo, há ocorrência de sentenças que foram divididas na reescrita em LIBRAS (Tabela 2). Entretanto, levando em consideração que trata-se de uma ferramenta para simplificação de textos, adjungindo a reescrita em LIBRAS e a simplificação textual.

Tabela 2. Exemplo de divisão de Sentenças.

Simplificação Forte	A neblina não ficou só na capital: a Central de Meteorologia afirma que a neblina cobriu as cidades de Santa Maria, Santo Angelo e Pelotas.
Interlíngua LIBRAS	<i>Neblina só capital não. CENTRAL METEOROLOGIA avisou neblina também Santa Maria, Santo Ângelo Pelotas.</i>

O segundo aspecto observado refere-se ao uso considerado adequado das operações conforme a descrição apresentada. Nesse contexto, a operação de substituição lexical foi utilizada de duas formas distintas. A primeira refere-se a uma substituição para adequar uma palavra a um sinal, ou seja, a palavra simplesmente não possui um sinal correspondente em LIBRAS e é substituída por um sinônimo que preserve o conteúdo da sentença, e que possui um sinal. (Tabela 3).

Tabela 3 . Exemplo de substituição lexical.

Simplificação Forte	A neblina escondeu o amanhecer dos <u>moradores</u> de Porto Alegre.
Interlíngua LIBRAS	<i>Neblina escondeu amanhecer <u>peessoas moram</u> Porto Alegre.</i>

Outra ocorrência para a substituição lexical diz respeito a uma palavra que possui sinal em LIBRAS, mas não é usada para expressar a mesma idéia do texto original. Esse o caso é exemplificado na Tabela 4.

Tabela 4. Exemplo de substituição lexical.

Simplificação Forte	<u>Leandro Puchalski é</u> da Central de meteorologia.
Interlíngua LIBRAS	<i><u>Ele trabalha</u> CENTRAL METEOROLOGIA.</i>

É interessante diferenciar o tipo dessa ocorrência em forma de duas novas operações. A primeira serve para identificar uma substituição por falta de vocabulário e a segunda para adequar o vocabulário ao conteúdo do texto. O último caso ocorre, muitas vezes, em consequência da simplificação forte que repete o sujeito original da sentença.

A operação de remoção de parte da sentença é utilizada para suprimir palavras que representariam sinais desnecessários na representação em LIBRAS. Entretanto, em alguns casos de sentenças mais curtas, por exemplo, apenas uma palavra é removida da sentença (Tabela 5). A criação de uma operação específica para registrar a remoção de uma palavra específica contribuirá com a qualidade das operações.

Tabela 5. Exemplo de substituição lexical.

Simplificação Forte	O secretário da Segurança Pública disse que o aumento está ligado à legislação pouco rigorosa contra desmanches, ao aumento da quantidade de carros e ao <u>chamado</u> golpe do seguro.
Interlíngua LIBRAS	<i>Secretário segurança público dizer aumento culpa legislação pouco rigoroso contra DESMANCHES, aumento número carro também golpe seguro.</i>
Detalhamento	Foi retirada a palavra "chamado" por não existir esta expressão em LIBRAS.

Dois comportamentos distintos caracterizaram a operação de inversão da ordem da sentença. No primeiro caso, é modificação a ordem básica da sentença com o objetivo iniciá-la com a idéia principal da sentença. Em LIBRAS, colocar em evidência um determinado trecho da sentença é denominado Topicalização. Sendo assim, é necessária a criação de uma operação específica para reescrita em LIBRAS: reescrever topicalização. A Tabela 6 apresenta um caso para uso dessa operação.

Tabela 6. Exemplo de topicalização.

Simplificação Forte	O tempo testou novamente a paciência de centenas de passageiros com vôos marcados no Aeroporto Internacional Salgado Filho ontem pela manhã.
Interlíngua LIBRAS	<i>De novo teste paciência pessoas ontem manhã vôo marcar aeroporto internacional SALGADO FILHO.</i>

No segundo caso de inversão, destaca-se a ocorrência da estrutura de discurso indireto na sentença. No entanto, há uma peculiaridade relacionada ao tamanho da sentença que define se essa inversão ocorre ou não. Apenas se uma sentença curta é estruturada na forma de um discurso indireto, a ordem da sentença é invertida na reescrita em LIBRAS. Já em sentenças maiores, ou em um conjunto de sentenças organizadas em forma de um discurso indireto, a reescrita não exige uma inversão. As Tabelas 7 e 8 exemplificam o segundo caso de inversão.

Tabela 7. Exemplo de inversão para discurso indireto.

Simplificação Forte	Agora, vamos atacar para valer - admitiu Bacci.
Interlíngua LIBRAS	<i>Bacci falou agora atacar forte.</i>

Tabela 8. Exemplo de discurso indireto sem inversão.

Simplificação Forte	Os ataques se tornaram mais freqüentes. Por isso, aconselhamos mais cuidado - diz o subprefeito Nei Pinto.
Interlíngua LIBRAS	<i>Ataque mais. Aconselhar muito cuidado dizer subprefeito NEI PINTO.</i>

7. Considerações Finais

É a partir dos relatos apresentados anteriormente, que se pretende expor as principais contribuições relacionadas a essa pesquisa. Com base no fato que, durante o período de desenvolvimento do trabalho, não foi encontrada qualquer referência a algum conjunto de textos escritos em LIBRAS e organizados em uma estrutura qualquer, a possibilidade de produzir um corpus bilíngüe alinhado por sentenças é um avanço no sentido do tratamento computacional para a LIBRAS. Além disso, o Módulo de Reescrita em LIBRAS é considerado uma contribuição importante para o objetivo do projeto PorSimples – Simplificação Textual do Português para inclusão e Acessibilidade Digital, por permitir o acesso à informação por parte do surdo no sentido de oferecer a possibilidade de uma melhor inteligibilidade dos textos durante o aprendizado de sua segunda língua.

No que se refere a trabalhos futuros, após a execução deste trabalho, vislumbra-se algumas possibilidades. Por exemplo, o desenvolvimento e exploração do corpus paralelo como forma de melhorar a qualidade das anotações incluindo novas operações. A partir disso, é importante propor um método para validar se a produção textual resultante do uso da ferramenta contribuir para a melhor compreensão da leitura para um surdo em processo de alfabetização no sentido de garantir acesso à informação e inclusão digital ao surdo.

Bibliografia

- ALUÍSIO, S. M. ; SPECIA, L. ; PARDO, T. A. S. ; MAZIERO, E. ; FORTES, R. P. M.. Towards Brazilian Portuguese Automatic Text Simplification Systems. In: ACM Proceedings of the 2008 ACM Symposium on Document Engineering. New York: ACM Digital Library, 2008. p.240-248.
- BICK, E.. The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. PhD thesis – Aarhus University. Aarhus, Denmark: Aarhus University Press, 2000.
- BIDERMAN, M. T. C. (2005). Dicionário Ilustrado de Português. São Paulo, Editora Ática. 1ª Edição. São Paulo: Ática.
- CAMPOS, M. B.; SILVEIRA, M. S.. Promoção da Cidadania da Comunidade Surda: o Uso das TICs na Apropriação de sua Língua Materna. In: Anais do XXXIV Seminário Integrado de Software e Hardware. Porto Alegre: Sociedade Brasileira de Computação, 2007. p.2232-2246.
- CHANDRASEKAR, R.; DORAN, C.; SRINIVRA, B. Motivations and Methods for Text Simplification. 1996. Proc. of COLING96, 1041-1044.
- ESTELITA, M. B.. ELis - Escrita das Línguas de Sinais: Proposta teórica e verificação prática. Tese (Doutorado em Linguística) - Universidade Federal de Santa Catarina, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. 2008.
- FENEIS. Lei n.º 10.436, de 24 de abril de 2002. Dispõe sobre a Língua Brasileira de Sinais - Libras e dá outras providências. Presidência da República – Casa Civil, Brasília. DF, 24 abr. 2002. Disponível em: <http://www.feneis.com.br/page/legislacao_resultado.asp?1> Acesso em: 22 de mai. 2007.

- FINAU, R. A.. O processo de formação de interlíngua na aquisição de língua portuguesa por surdos e as categorias tempo e aspecto. In: Heloísa Salles. (Org.). Bilingüismo dos Surdos. Questões Lingüísticas e Educacionais. 1 ed. : Cãnone, 2007, v. 1, p. 161-191.
- MAX, A. Writing for Language-impaired Readers. In the Proceedings of Seventh International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), pp. 567-570. Mexico City, Mexico, (2006).
- PETERSEN, S. E., OSTENDORF, M.. Text Simplification for Language Learners: A Corpus Analysis. Speech and Language Technology for Education workshop. Pennsylvania, USA. 2007
- QUADROS, Ronice Müller de. Educação de surdos: a aquisição da linguagem. Porto Alegre: Artes Médicas, 1997. 126 p.
- QUADROS, R. M. KARNOPP, L. B. Língua de Sinais Brasileira: Estudos Lingüísticos. Porto Alegre. Artmed, 2004. 222 p.
- QUADROS, Ronice Müller de; SCHMIEDT, Magali L. P. Idéias para ensinar português para alunos surdos. Brasília : MEC, SEESP, 2006.
- SBC. Grandes Desafios da Pesquisa em Computação no Brasil – 2006 – 2016. SBC - São Paulo. 2006.
- SPECIA, L.; ALUISIO, S.M.; PARDO, T.A.S. Manual de Simplificação Sintática para o Português. Technical Report NILC-TR-08-06, 27 p. 2008, São Carlos-SP.
- SIDDHARTHAN A.. An Architecture for a Text Simplification System. In Proceedings of the Language Engineering Conference 2002 (LEC 2002). Pages 64-71.
- SIDDHARTHAN, A. Syntactic simplification and text cohesion. Phd thesis. University of Cambridge. 2003.
- SIDDHARTHAN, A.. Syntactic Simplification and Text Cohesion. In Research on Language and Computation, Volume 4, Issue 1, 2006, Pages 77--109, Springer Science, the Netherlands.
- SIGNWRITING. Sutton's SignWriting Site. Disponível em: <<http://www.signwriting.org>>. Acessado em: junho/2008.
- SOUZA, V. C. e VIEIRA, R.. Uma Proposta para Tradução Automática entre Libras e Português no Sign WebMessage. IV Workshop de Tecnologia da Informação e Linguagem Humana TIL, 2006.