

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**PRODUÇÃO DE TEXTOS PARALELOS EM LÍNGUA
PORTUGUESA E UMA INTERLÍNGUA DE LIBRAS**

GUILHERME SPOLAVORI DOS SANTOS

Dissertação apresentada como requisito parcial a obtenção do grau de mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientadora: Prof^a. Dr^a. Milene Selbach Silveira

Co-orientadora: Prof^a. Dra. Renata Vieira

Porto Alegre

2009

Dados Internacionais de Catalogação na Publicação (CIP)

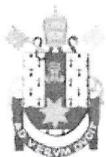
S237p Santos, Guilherme Spolavori dos
Produção de textos paralelos em língua portuguesa e uma
interlíngua de Libras / Guilherme Spolavori dos Santos. – Porto
Alegre, 2009.
94 f.

Dissertação (Mestrado) – Faculdade de Informática, PUCRS.
Orientador: Prof.^a Dr.^a Milene Selbach Silveira
Co-orientadora: Prof.^a Dr.^a Renata Vieira

1. Informática. 2. Processamento de Linguagem Natural.
3. Língua Brasileira de Sinais. 4. Acessibilidade. I. Silveira, Milene
Selbach. II. Vieira, Renata. III. Título.

CDD 006.35

**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "**Produção de Textos Paralelos em Língua Portuguesa e uma Interlíngua de LIBRAS**", apresentada por Guilherme Spolavori dos Santos, como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Sistemas Interativos de Visualização, aprovada em 05/03/09 pela Comissão Examinadora:

Milene Selbach Silveira

Profa. Dra. Milene Selbach Silveira – PPGCC/PUCRS
Orientadora

Renata Vieira

Profa. Dra. Renata Vieira – PPGCC/PUCRS
Co-orientadora

Vera Lúcia Strube de Lima

Profa. Dra. Vera Lúcia Strube de Lima – PPGCC/PUCRS

Márcia de Borba Campos

Profa. Dra. Márcia de Borba Campos – FACIN/PUCRS

Sandra Maria Aluísio

Profa. Dra. Sandra Maria Aluísio – USP-São Carlos

Homologada em 21/07/2016, conforme Ata No. 014 pela Comissão Coordenadora.

Fernando Gehm Moraes
Prof. Dr. Fernando Gehm Moraes
Coordenador.

PUCRS

Campus Central

Av. Ipiranga, 6681 – P32 – sala 507 – CEP: 90619-900

Fone: (51) 3320-3611 – Fax (51) 3320-3621

E-mail: ppgcc@pucrs.br

www.pucrs.br/facin/pos

PRODUÇÃO DE TEXTOS PARALELOS EM LÍNGUA PORTUGUESA E UMA INTERLÍNGUA DE LIBRAS

RESUMO

Esse trabalho situa-se nas áreas de Interação Humano-Computador e Processamento de Linguagem Natural, atuando em conceitos sobre acessibilidade, a partir de uma arquitetura de simplificação textual e uma estrutura de corpus. O resultado da pesquisa trata da produção de textos paralelos – na versão da Língua Portuguesa e na versão da Língua Brasileira de Sinais (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. Sendo assim, foi elaborado um referencial teórico em que, além de estudos sobre aspectos linguísticos da LIBRAS e aquisição da segunda língua pelo surdo, faz-se referência a conceitos da simplificação textual. De forma complementar, trabalhos relacionados são apresentados como contribuições no que diz respeito à aplicação desses conceitos. A partir disso, o trabalho estuda a arquitetura e funcionalidades do Editor de Anotações de Simplificação, dentro do contexto do projeto PorSimples, como plataforma para o desenvolvimento de um módulo de produção textual em uma interlíngua da LIBRAS baseado em um processo de simplificação. Por fim, as funcionalidades do módulo são exploradas por um especialista da língua de sinais em questão e os textos produzidos são analisados buscando melhorias no processo de produção dos textos.

Palavras-chave: Processamento de Linguagem Natural, Língua Brasileira de Sinais, Simplificação Textual, Análise de Corpus Paralelo.

PRODUCTION OF PARALLEL TEXTS IN PORTUGUESE LANGUAGE AND AN LIBRAS INTERLINGUA

ABSTRACT

This work embraces the Human-Computer Interaction and the Natural Language Processing areas of knowledge, touching accessibility concepts, based on one textual simplification architecture and on one *corpus* structure. The result of this research is the generation of parallel texts – in the Portuguese Language version and in Brazilian Sign Language (LIBRAS) version – with the objective of supplying a data repository created for future works in the Natural Language Processing area that have relation with the language of the Brazilian deaf people. This work presents a theoretical revision where are discussed questions like studies about LIBRAS linguistic aspects, acquisition of a second language in the deaf community as well as concepts of textual simplification. Complementing this discussion, related works are presented as contributions concerning to the application of the referred concepts. Based on this revision, this work studies the architecture and the functionalities of the *Editor de Anotações de Simplificação*, in the context of *PorSimples* project, as a platform for the development of a module of textual generation in a LIBRAS interlingua which is based in a simplification process. At last, the functionalities of the module are explored by a LIBRAS expert and the generated texts are analyzed searching for future improvements in the process of text production.

Keywords: Natural Language Processing. Brazilian Sign Language. Text Simplification. Parallel Corpus Analysis.

LISTA DE FIGURAS

Figura 1 - Exemplos de sinais.....	21
Figura 2 - Exemplo de classificadores	21
Figura 3 - Concordância Verbal	22
Figura 4 - Estágios da simplificação textual conforme	26
Figura 5 - Exemplo de uso da notação	33
Figura 6 - Ferramenta ELAN.....	34
Figura 7 - Processo de simplificação	40
Figura 8 - Lista de Produções no Editor de Anotação de Simplificação.....	41
Figura 9 - Interface de anotação de simplificação.....	41
Figura 10 - Interface de Simplificação Forte	42
Figura 11 - Estrutura de dados do Editor de Anotação de Simplificação	45
Figura 12 - Arquivos em anotação XCES do corpus paralelo	49
Figura 13 - Novo fluxo de criação de produções.....	51
Figura 14 - Acesso a criação da versão em LIBRAS	51
Figura 15 - Opção de Reescrita em LIBRAS para produções já existentes.....	52
Figura 16 - Interface para reescrita em LIBRAS	53
Figura 17 - Interface para anotação de substituição	62

LISTA DE TABELAS

Tabela 1 -	Fragmentos de textos analisados	18
Tabela 2 -	Relação entre as conjunções e as <i>cue-words</i>	29
Tabela 3 -	Fragmentos de textos	35
Tabela 4 -	Estatísticas sobre o corpus analisado	36
Tabela 5 -	Fragmentos de textos analisados	37
Tabela 6 -	Cardinalidade das operações de simplificação	44
Tabela 7 -	Tipos de arquivos de anotação XCES	48
Tabela 8 -	Exemplo de uso da operação de Reescrita em LIBRAS	54
Tabela 9 -	Exemplo de uso da operação de Reescrita Forte	54
Tabela 10 -	Exemplo de uso da operação de Mudança de voz	55
Tabela 11 -	Exemplo de inversão de sentença	55
Tabela 12 -	Exemplo de uso da operação de Divisão de sentenças	55
Tabela 13 -	Exemplo de uso da operação de União de sentenças	56
Tabela 14 -	Exemplo de uso da operação de Remoção de parte da sentença	56
Tabela 15 -	Exemplo de uso da operação de Substituição Lexical	57
Tabela 16 -	Exemplo de uso da operação de Substituição Datilológica	57
Tabela 17 -	Cardinalidade das operações de simplificação	58
Tabela 18 -	Definição de estado do atributo <i>visible</i>	59
Tabela 19 -	Relação de visibilidade para cada operação	59
Tabela 20 -	Exemplo de união de sentenças já divididas	63
Tabela 21 -	Exemplo de união de sentenças	63
Tabela 22 -	Exemplo de substituição lexical	63
Tabela 23 -	Exemplo de substituição lexical	64
Tabela 24 -	Exemplo de substituição lexical	64
Tabela 25 -	Exemplo de topicalização	65
Tabela 26 -	Exemplo de inversão para discurso indireto	65
Tabela 27 -	Exemplo de discurso indireto sem inversão	65

LISTA DE SIGLAS E DE ABREVIATURAS

ANVISA	-	Agência Nacional de Vigilância Sanitária
CPI	-	Comissão Parlamentar de Inquérito
DESIRE	-	Deaf and Sign Language Research Team
ELAN	-	EUDICO Linguistic Annotator
DGS	-	German Sign Language (Deutsche Gebärdensprache)
EUDICO	-	European Distributed Corpora Project
FENEIS	-	Federação Nacional de Educação e Integração dos Surdos
IBGE	-	Instituto Brasileiro de Geografia e Estatística
L1	-	Primeira língua
L2	-	Segunda Língua
LIBRAS	-	Língua Brasileira de Sinais
MPEG	-	<i>Motion Picture Expert Group</i>
MVC	-	<i>Model View Controller</i>
NILC	-	Núcleo Interinstitucional de Linguística Computacional
OSV	-	Objeto-Sujeito-Verbo
PLN	-	Processamento de Linguagem Natural
SBC	-	Sociedade Brasileira de Computação
SOV	-	Sujeito-Objeto-Verbo
SVO	-	Sujeito-Verbo-Objeto
ST	-	Simplificação Textual
XCES	-	<i>Corpus Encoding Standard</i>
XML	-	<i>eXtensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivos	13
1.2	Organização da Dissertação	13
2	LÍNGUA BRASILEIRA DE SINAIS	15
2.1	Aquisição da Linguagem e da Escrita	16
2.2	Aspectos Linguísticos da LIBRAS	20
3	SIMPLIFICAÇÃO TEXTUAL	25
3.1	Estágio de Análise	26
3.2	Estágio de Transformação	28
3.3	Estágio de Regeneração	29
4	TRABALHOS RELACIONADOS	32
4.1	Corpus Bilíngue para a DGS – Língua de Sinais Alemã	32
4.2	Análise de um corpus para extração de regras de simplificação	35
5	FERRAMENTAS PARA SIMPLIFICAÇÃO TEXTUAL DE PORTUGUÊS	39
5.1	Editor de Anotação de Simplificação	39
5.1.1	Interface e Funcionalidades	40
5.1.2	Operações e Alinhamento	43
5.1.3	Arquitetura	44
5.2	Portal de Corpora Simplificados	47
6	MÓDULO DE REESCRITA EM LIBRAS	50
6.1	Funcionalidades e Interface	51
6.2	Operações	53
6.2.1	Reescrita em LIBRAS	53
6.2.2	Reescrita Forte	54
6.2.3	Mudança de voz	54
6.2.4	Inversão na ordem da sentença	55
6.2.5	Divisão de sentenças	55
6.2.6	União de sentenças	56
6.2.7	Remoção da sentença e Remoção de parte da sentença	56
6.2.8	Substituição Lexical	57
6.2.9	Substituição Datilológica	57
6.3	Cardinalidade	58

6.4	Arquitetura	58
7	A EXPERIÊNCIA DE USO.....	60
7.1	Atividade realizada.....	60
7.2	Problemas encontrados	61
7.3	Uso das operações	62
7.4	Depoimento do especialista.....	66
8	CONSIDERAÇÕES FINAIS.....	67
8.1	Trabalhos Futuros.....	68
	REFERÊNCIAS BIBLIOGRÁFICAS.....	70
	APÊNDICE – EXEMPLO DO FORMATO XCES GERADO NO SISTEMA	73

1 INTRODUÇÃO

A comunicação humana é essencialmente diferente e superior a todas as outras formas de comunicação conhecidas. Todos os seres humanos nascem com os mecanismos da linguagem, específicos da espécie, e todos os desenvolvem normalmente, independente de qualquer fator racial, social ou cultural [QUA97]. Ao mesmo tempo, os resultados do Censo 2000 mostram que aproximadamente 24,6 milhões de pessoas, ou 14,5% da população total brasileira, apresentam algum tipo de necessidade especial. E que, dentre os 5,7 milhões de brasileiros com algum grau de problema auditivo, um pouco menos de 170 mil declararam-se surdos, segundo o IBGE [INSG07]. Para todas essas pessoas, a necessidade de comunicação é imprescindível para expressar a sua identidade e a sua personalidade.

Durante sua existência, o ser humano desenvolveu a capacidade de comunicação social através da linguagem. Nesse sentido, entende-se por linguagem natural a forma como o ser humano faz uso de um sistema de comunicação – uma linguagem – para trocar informação entre si. Sendo assim, a ideia de estudar o Processamento de Linguagem Natural (PLN) para o tratamento de língua de sinais (principal forma de comunicação do surdo) tem contribuição fundamental ao acesso à informação e à inclusão digital do surdo.

No âmbito da área de PLN, a Simplificação Textual (ST) é uma área de pesquisa emergente chamada geração texto-para-texto. O objetivo da ST é maximizar a compreensão de textos escritos pela simplificação de sua estrutura linguística. Como resultado, espera-se que todo o texto possa ser mais facilmente compreendido [SID03] [MAX06]. Atualmente, essa linha de pesquisa aponta soluções para o processamento de variadas e complexas estruturas de sentenças, de forma a resolver ambiguidades e dependências contextuais (consequência da riqueza de uma língua natural) e as transformar em sentenças simplificadas.

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. Há ainda estudos, como em Petersen e Ostendorf [PET07], que defendem a simplificação textual como

ferramenta de apoio no ensino de uma segunda língua, considerando que nem sempre os alunos tem contato com textos cujo nível de leitura exigido corresponde ao nível intelectual e interesse deles. Isso caracteriza essa tarefa como uma forma de promover acessibilidade à informação a 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 [SBC06]. E, no âmbito desse desafio, a comunidade científica é estimulada a apresentar novas questões e propostas no intuito de melhor atendê-lo. Em Campos e Silveira [CAM07], por exemplo, são apresentadas ideias que dizem respeito à Comunidade Surda, especificamente, e suas necessidades quanto à apropriação e divulgação de sua língua. Dentre elas, encontram-se estudos na área de língua de sinais e na área de PLN, com o desenvolvimento de sistemas de tradução automática entre língua de sinais e línguas orais.

Diversas iniciativas já foram desenvolvidas para difundir as línguas de sinais, inclusive através do desenvolvimento de softwares como forma de apoiar a inclusão digital e social das pessoas surdas [CAM01] [SOU06] [TAV05], por exemplo. Normalmente, o objetivo dessas aplicações é familiarizar o surdo com a escrita de sinais, sem uma preocupação aprofundada com a estrutura linguística da língua de sinais. A LIBRAS, 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 português falado, por exemplo – assim como a aprendizagem da LIBRAS pelos surdos e pelos ouvintes. Nesse sentido, uma ferramenta computacional que aborde a complexidade da língua de sinais em forma de um tradutor automático vem a auxiliar a comunidade surda na inclusão digital e no acesso à informação.

Essas ideias, no entanto, englobam diversas linhas de pesquisas dentro da área de PLN, bem como decisões no que se referem à abordagem mais adequada para uma ferramenta de tradução. Essas abordagens variam entre modelos fundamentais (regras para tradução) e modelos estatísticos (análise sobre textos já traduzidos). Uma abordagem estatística, em longo prazo, possui a vantagem de evoluir conforme a qualidade dos textos traduzidos. Entretanto, para isso, é necessário 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

deste trabalho: realizar a aquisição dos textos no idioma desejado, ou seja, em LIBRAS, e a versão paralela destes textos em Língua Portuguesa.

1.1 Objetivos

O objetivo geral desse trabalho é a 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) [PER08]. – foi adotada para a produção desses textos a partir do texto em português já simplificado.

Esse projeto, desenvolvido pelo Núcleo Interinstitucional de Linguística Computacional (NILC), propõe o desenvolvimento de uma tecnologia para facilitar o acesso à informação, por meio da simplificação de textos em português do Brasil [ALU08]. Sendo assim, como objetivos específicos, busca-se 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. 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 à informação.

1.2 Organização da Dissertação

O texto está organizado em oito capítulos.

Os capítulos dois e três constituem o referencial teórico em que está baseada a pesquisa. Primeiramente, são descritos aspectos linguísticos da Língua Brasileira de Sinais, bem como a importância dessa língua na formação bilíngue do surdo (LIBRAS e Língua Portuguesa). Após, são apresentados conceitos sobre a simplificação textual bem como arquiteturas abordadas na literatura.

Os capítulos quatro e cinco referem-se ao estudo de trabalhos envolvidos diretamente com os objetivos dessa pesquisa. No capítulo quatro, dois trabalhos são analisados com o objetivo de entender às necessidades envolvidas na criação de um corpus paralelo. Na sequência, são apresentadas características do editor de anotações para simplificação textual (vinculado ao projeto PorSimples) como plataforma para o desenvolvimento do trabalho.

A descrição do desenvolvimento do trabalho, em relação às alterações necessárias para inclusão do Módulo de Reescrita em LIBRAS, e da experiência de uso, bem como as alterações e observações realizadas, são apresentadas, respectivamente, nos capítulos seis e sete.

Por fim, o capítulo oito conclui essa dissertação apresentando considerações finais que se referem ao que foi pesquisado e desenvolvido durante a pesquisa.

2 LÍNGUA BRASILEIRA DE SINAIS

A Língua de Sinais é a língua natural do surdo e, de uma forma geral, trata da substituição de sons por gestos e expressões faciais, os quais compõem a principal forma de comunicação utilizada entre pessoas surdas ou entre pessoas com algum outro tipo de problema auditivo. No que se refere à estrutura, essa língua é composta por níveis linguísticos 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 se comunicar usando língua de sinais, não basta apenas conhecer os sinais, é necessário conhecer a sua gramática para estruturar as frases, estabelecendo, desta forma, a comunicação.

Além disso, a língua de sinais tem papel importante no desenvolvimento cognitivo e social do indivíduo e lhe permite a aquisição de conhecimentos sobre o mundo que o cerca. Assim, a aprendizagem da língua de sinais é um instrumento indispensável para a verdadeira integração entre as pessoas surdas de forma que elas usufruam sua plena cidadania. Entretanto, o fato da língua de sinais ser, normalmente, a principal forma de comunicação do surdo, acaba tornando-se uma barreira para aquisição de conhecimento sobre o mundo externo a sua comunidade. Em outras palavras, uma vez que a informação está registrada em uma língua oral, o surdo precisa ter acesso a ambas as formas de comunicação para garantir o acesso ao ensino e ser independente no seu processo de aprendizagem.

Da mesma forma que existem diferentes idiomas para línguas faladas, existem diferentes línguas de sinais correspondentes aos idiomas 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 [FEN02].

O presente capítulo tem como objetivo de abordar a importância da LIBRAS na formação bilíngue dos surdos, em relação à escrita da Língua Portuguesa, e demonstrando o fenômeno da formação de uma interlíngua durante esse processo de aprendizado. Também, neste capítulo, é apresentado, em uma visão geral, as

particularidades linguísticas e a complexidade da Língua Brasileira de Sinais, relacionados aos seus aspectos morfológicos e sintáticos.

2.1 Aquisição da Linguagem e da Escrita

Vale destacar que o surdo não está necessariamente restrito a uma única forma de comunicação. Uma educação bilíngue 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 [QUA97]. Torna-se mais fácil compreender a importância da abordagem bilíngue no processo de aquisição de linguagem, ao pensar que, no Brasil, a constituição define o idioma português como oficial, ou seja, é o idioma em que se registram os bens, os compromissos, a identificação de pessoas e o próprio ensino. É importante ressaltar que é direito de todo o cidadão aprender e utilizar o idioma oficial definido pela constituição. Assim, a aprendizagem do idioma português pelo surdo – como sua segunda língua – é imprescindível para o exercício de sua cidadania.

Segundo Quadros e Schmiedt [QUA06], 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 [QUA97], 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. Um exemplo de método de alfabetização do surdo é o sistema artificial de português sinalizado, que usa os sinais da LIBRAS na estrutura do português, desconsiderando a complexidade linguística da língua de sinais brasileira. O domínio sobre a língua de sinais é adquirido de acordo com as experiências de interação com usuários dessa mesma língua. Isso é válido também para a escrita - tanto da língua de sinais quanto da língua oral - por parte do surdo, no sentido de que praticar a escrita é tão importante para o aprendiz, quanto estudar a estrutura linguística que a sustenta. Dessa forma, o aprendiz de português pelo surdo requer primeiro a apropriação de sua própria língua (LIBRAS), sendo isso fundamental nesse processo de aprendizagem. Esse processo de

alfabetização bilíngue vai sendo delineado com base no processo de descoberta e de amadurecimento da própria língua (no caso do surdo, da língua de sinais).

Nesse sentido, um sistema de escrita que se aproxime da língua de sinais é uma porta que se abre no processo de alfabetização do aluno surdo que domina a língua de sinais usada no País. Este sistema envolve a composição das unidades mínimas de significado da língua, formando estruturas em forma de textos [QUA06]. Existem sistemas não alfabéticos que representam as unidades morfológicas para escrita de línguas de sinais – tais como o *SignWriting* [SIG08] e o Elis [EST08], por exemplo –, mas não são difundidos ainda. Sendo assim, a atividade da escrita em português, como parte da aprendizagem da segunda língua, implica à ocorrência de um fenômeno natural interpretado como a formação de uma interlíngua entre a L1 e a L2. Finau [FIN07] analisa a formação dessa interlíngua como sistema linguístico organizado em uma “Gramática Mental” que sofre influência da língua nativa (L1) do aprendiz e da língua alvo (L2). Além disso, essa gramática também pode ser influenciada pelo próprio processamento interno, o que gera a transferência de regras, a omissão de estruturas e as generalizações. Portanto, ela muda com o tempo, e isso pode ser percebido quando o usuário adiciona ou apaga regras aumentando gradativamente a complexidade de seu conhecimento em relação à L2.

Segundo Chan-Vianna [CHA07], reconhece-se que a gramática da interlíngua difere de várias maneiras da gramática dos ouvintes, que têm uma língua materna (L1) oral-escrita. A interlíngua caracteriza-se pela variabilidade no uso de flexão verbal e nominal e em itens lexicais relacionados. Morfologia relacionada à concordância, número, gênero, caso e tempo, entre outras categorias, estão às vezes presentes, e às vezes ausentes na produção dos aprendizes de L2, quando seriam obrigatórios para as pessoas que têm como L1 uma língua oral-escrita. Em Brochado [BRO06], é analisada a produção textual de alunos surdos fluentes na língua de sinais, com idade de 8 a 11 anos, do ensino fundamental de uma escola da rede estadual de educação. São descritas características sobre diferentes estágios dessa interlíngua durante o processo de aquisição do português pelo surdo. A Tabela 1 apresenta alguns exemplos destes diferentes estágios.

Tabela 1 - Fragmentos de textos analisados [BRO06]

Interlíngua I	Interlíngua II	Interlíngua III
<u>A raposa e as uva</u> Colhas come muito uva Colhos fugiu escuro árvore Raposa quero vontade come uva Raposa pulou não pegeu conseguiu	<u>Chapeuzinho Vermelho</u> Mãe fala chapeuzinho vermelho A vovó muito doena [doente?] Chapeuzinho Vermelho foi vê flor muito bonita Chapeuzinho Vermelho assauto lobo. Lobo corre muito casa vovó lobo come vovó Chapeuzinho Vermelho lobo quem chapeuzinho Vermelho porque olho grande, porque nariz grande, porque orelha grande, porque boca grande come chapeuzinho Vermelho.	<u>Chapeuzinho Vermelho</u> A mamãe falou: – Chapeuzinho por favo você vai casa da vovó. Chapeuzinho falou – Porque eu vou casa da vovó? Mamãe falou – Porque a vovó está doente entendeu. Chapeuzinho falou – da eu vou casa da vovó porque eu tenho soudade da vovó eu do feliz. A mamãe falou – Por favor cuidado mato é perigoso.

Conforme a autora, no primeiro estágio observa-se o emprego predominante de estratégias de transferência da língua de sinais (L1) para a escrita da língua portuguesa (L2) desses alunos, caracterizando-se por:

- Predomínio de construções frasais sintéticas;
- Estrutura gramatical de frase muito semelhante à LIBRAS (L1), apresentando poucas características da Língua Portuguesa (L2);
- Predomínio da construção de topicalização sobre a ordem SVO;
- Predomínio de palavras de conteúdo (substantivos, adjetivos, verbos);
- Falta de elementos funcionais (artigos, preposição, conjunção);
- Uso de verbos, preferencialmente, no infinitivo;
- Uso raro de verbos de ligação (ser, estar, ficar);
- Falta de flexão dos nomes em gênero, número e grau,
- Pouca flexão verbal em pessoa, tempo e modo.

No segundo estágio, há, na escrita de alguns alunos, uma intensa mescla das duas línguas, na qual se observa o uso de estruturas linguísticas da LIBRAS e o uso de elementos da língua portuguesa, na tentativa de se apropriar da língua alvo [BRO06]. Essa mescla caracteriza-se por:

- Justaposição intensa de elementos da L1 e da L2;
- Estrutura da frase ora com características da língua de sinais brasileira, ora com características gramaticais da frase do português;
- Uso de verbos no infinitivo e também flexionados;
- Uso de palavras de conteúdo (substantivos, adjetivos e verbos);

- Uso, pouco frequente, de verbos de ligação com correção;
- Uso de elementos funcionais de modo inadequado;
- Uso de artigos, algumas vezes com concordância correta;
- Uso de algumas preposições, nem sempre adequado,
- Uso de conjunções, quase sempre inadequado.

No último estágio, a escrita demonstra o emprego da gramática da língua portuguesa em todos os níveis, principalmente no sintático, caracterizando-se por apresentar:

- Uso de palavras funcionais (artigos, preposição, conjunção);
- Uso de categorias funcionais, predominantemente, com adequação;
- Uso de artigos definidos e, algumas vezes, de indefinidos;
- Uso de preposições com mais acertos;
- Uso de algumas conjunções coordenativas, além das subordinativas;
- Uso dos nomes flexionados;
- Uso de verbo flexionado, com maior adequação;
- Uso de flexão dos nomes em gênero, número e grau;
- Uso de flexão verbal em pessoa (1ª e 3ª pessoas), número (1ª e 3ª pessoas do singular e 1ª pessoa do plural) e tempo (presente e pretérito perfeito), com consistência,
- Uso de verbos de ligação ser, estar e ficar com maior frequência e correção.

Esse estudo possibilita observar como a produção escrita desenvolve-se durante o processo de aquisição da segunda língua (nesse caso, da Língua Portuguesa). Além disso, demonstra, de forma geral, como se organiza essa versão intermediária da língua de sinais, combinando a estrutura da LIBRAS com palavras do português escrito. Um exemplo do uso dessa versão escrita é observado em Estelita [EST08], 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.

- a) [Interlíngua LIBRAS]: Lara ir junto mamãe praia. Lara brincar areia ver assustar tartaruga grande, sair correr chorar junta mamãe. Mamãe explicar chorar-não porque tartaruga morder-não, poder carinho. Lara resolver carinho tartaruga, admirada igual altura.

b) [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. Sua mãe lhe explicou que não precisava chorar porque tartaruga não morde que podia fazer carinho. Lara resolveu ir fazer carinho na tartaruga e ficou muito admirada em ver que eram da mesma altura.

O fato de ser de uma linguagem natural ao surdo, e por isso menos utilizada, não implica menor complexidade em termos de estrutura. Na próxima seção, serão apresentados características linguísticas que evidenciam a importância da LIBRAS.

2.2 Aspectos Linguísticos da LIBRAS

Pode acontecer de uma língua que não é usada em todos os setores da sociedade, ou que é usada em uma cultura distinta da que conhecemos não apresentar vocábulos ou palavras para determinado campo semântico, todavia, isso não significa que essa língua seja pobre, uma vez que, potencialmente, ela tem todos os mecanismos para criar ou gerar palavras para qualquer conceito que vier a ser utilizado pela comunidade que a usa [BRI96]. 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 da aplicação dessas estruturas por mecanismos morfológicos, sintáticos e semânticos. De acordo com Campos [CAM01], os sinais são formados através das seguintes combinações (Figura 1):

- Configuração de mão: forma que a mão assume na execução do sinal, incluindo a orientação da sua palma e do seu dorso;
- Ponto de Articulação: posição das mãos no instante da sinalização;
- Movimento: movimentos das mãos, dos dedos ou do posicionamento do corpo;
- Expressão: expressões faciais, as quais indicam dúvida, espanto ou mesmo graus superlativos de um dado adjetivo.



Figura 1 - Exemplos de sinais [PEI06]

Entretanto, formar um sinal não trata apenas de combinar movimentos com configurações de mão ou expressões faciais. Existem regras que permitem a formação de unidades mínimas, 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” [BRI96]. 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” exemplificam a formação de sinais por composição.

Outra possível formação é definida através dos classificadores. Segundo Brito [BRI96], os classificadores são um tipo de morfema gramatical que é afixado a um morfema lexical ou sinal para mencionar a classe à qual pertence esse sinal, para descrevê-lo quanto à forma e tamanho, ou para descrever a maneira como ele é segurado ou se comporta na ação verbal. Assim, a partir de um sinal acrescentado de um classificador, pode-se formar um novo sinal com um novo significado e uma nova classe gramatical. Como exemplo, temos os classificadores B e Y. O primeiro é usado para descrever superfícies planas como mesa ou parede, por exemplo, enquanto o segundo descreve objetos com formas irregulares. A Figura 2 ilustra os classificadores citados.

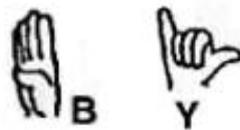


Figura 2 - Exemplo de classificadores [QUA97]

Independente do tipo de formação, a maioria dos sinais não varia sua forma quanto às categorias de verbo, de nome, de adjetivo ou de advérbio, no entanto, podem ser

classificados de acordo com a função exercida por eles na sentença. Segundo Felipe [FEL97], as categorias gramaticais são as classes de palavras de uma língua. De uma forma geral, todas as línguas possuem palavras que podem ser classificadas como pertencentes a um determinado tipo ou classe no que se refere aos seus aspectos morfológicos, sintáticos e semânticos. Embora muitas línguas não possuam algumas classes gramaticais, isso não significa que sejam inferiores a outras, pois elas expressam de forma diferente os mesmos conceitos. Na LIBRAS, por exemplo, não existem artigos, enquanto que no inglês falado existe apenas uma ocorrência para artigo definido, o “*the*”.

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 3 ilustra a marcação no espaço de sinalização que define a concordância de um determinado verbo.

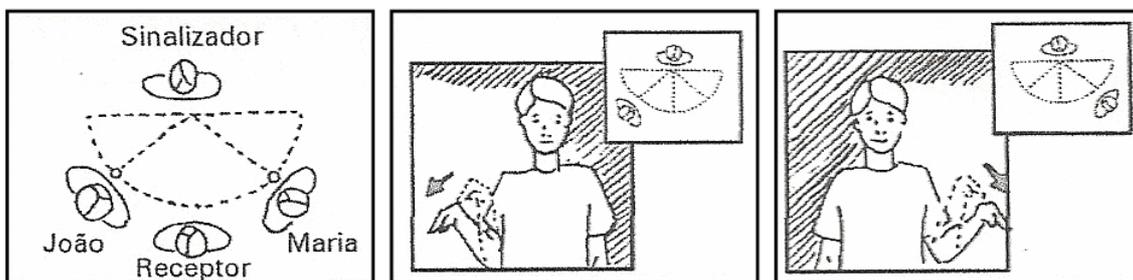


Figura 3 - Concordância Verbal [QUA97]

Uma característica sobre a estrutura da frase refere-se à ocorrência de pronomes e à flexão dos verbos. Em LIBRAS, assim como em Português, o sujeito da frase, muitas vezes, pode ser suposto pelo contexto ou, quando faz referência à primeira pessoa, é, supostamente, identificado pelo receptor. Em alguns casos, como os verbos ver, perguntar e responder, por exemplo, a identificação de sujeito ou objeto não ocorre em forma de um sinal específico, mas, sim, na forma de um movimento. Este movimento tem o seu ponto de origem referindo-se ao sujeito e o seu ponto final ao objeto. À definição de direção com esses pontos é dado o nome de flexão verbal [QUA04].

Ainda no que se refere ao aspecto verbal, de acordo com Brito [BRI96], a LIBRAS, assim como várias línguas de sinais e línguas orais, modula o movimento dos sinais para distinguir entre os aspectos pontual, continuativo e iterativo. O aspecto pontual se

caracteriza por se referir a uma ação ou a um evento ocorrido e terminado em algum ponto bem definido no passado, enquanto que o aspecto continuativo refere-se a uma ação que tem uma continuidade no tempo, já o aspecto iterativo refere-se à ação ou evento que se dá repetidas vezes. Por outro lado, a LIBRAS não apresenta marcação de tempo para suas formas verbais. Dessa forma, o que define a conjugação do verbo são os sinais adverbiais referentes ao tempo, o que evita a ocorrência de ambiguidades uma vez que iniciada a narração com uma marca no passado, enquanto não for introduzida outra marca de tempo, o contexto interpretado não muda. Nesse sentido, a ausência de marcação de tempo indica a conjugação do verbo no tempo presente.

Conforme Quadros e Karnopp [QUA04], 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 nas expressões a seguir. Os exemplos abaixo apresentam a ordem SVO em orações simples e uma ocorrência de oração subordinada.

JOÃO FALAR MARIA

(João falou com Maria)

JOÃO VIAJAR GRAMADO ONTEM

(João viajou para Gramado ontem).

EU ACHAR MARIA IR EMBORA

(Eu acho que a Maria vai embora)

Além dessa, existem ordenações secundárias resultantes da interação de outros mecanismos gramaticais. Sendo a LIBRAS uma língua 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 as autoras, 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)

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). Porém, pode ocorrer topicalização sem haver ligação, necessariamente, com qualquer parte da frase. Nesse caso, é comum existir uma cópia do tópico, em forma de um pronome para referência, por exemplo. 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 SIMPLIFICAÇÃO TEXTUAL

Segundo Max [MAX06], a simplificação textual é definida como o processo que transforma um texto, reduzindo 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. Originalmente, essa tarefa foi pensada como um pré-processamento do texto para diminuir a complexidade das sentenças, tornando-as mais simples em seu processamento. Com o tempo, essa abordagem convergiu para uma tecnologia de apoio a pessoas com dificuldades para a aquisição e para o uso da linguagem. Esse conceito pode ser ilustrado com o exemplo a seguir [SPE08]:

- (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 adjetiva relativa, sendo ambas ligadas por uma conjunção relativa ("que"). A simplificação, nesse caso, tomou como base regras pré-definidas para separar a sentença no seu ponto de articulação – a conjunção "que" – gerando, em (2), uma sentença apenas com a oração principal e outra sentença com a oração subordinada, que terá como sujeito o núcleo do sujeito da oração principal [SPE08].

Dentre as utilidades relacionadas à tarefa da simplificação textual, Siddharthan [SID02] 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 de pré-processamento, considerando que sentenças longas são problemáticas nessa tarefa ao contrário de sentenças menores como o caso das geradas na simplificação. Ainda nessa linha, as sentenças longas podem prejudicar a performance de tarefas como a de tradução automática e sumarização, uma vez que sentenças curtas podem ser traduzidas com mais exatidão e sentenças simplificadas possuem menos informação agregada a ser extraída. 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 [SID02]. 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 [PET07].

Pesquisas relacionadas, como a de Chandrasekar *et al.* [CHA96], 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 sequência 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 [SID02] [SID06], 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 4 apresenta estes estágios em uma arquitetura para o processo de simplificação.

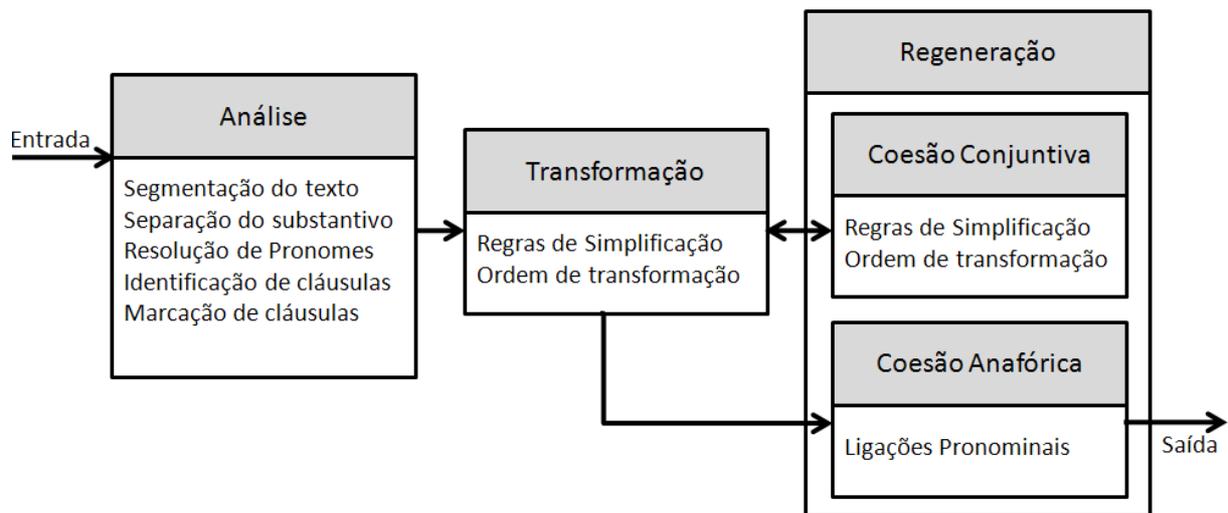


Figura 4 - Estágios da simplificação textual conforme [SID06]

3.1 Estágio de Análise

Segundo Siddharthan [SID06], existem diferentes níveis de análise possíveis no processamento do texto, podendo variar de uma análise detalhada e computacionalmente cara, até uma análise superficial baseada em marcações referentes à categoria gramatical e formação morfológica (etiquetagem ou *part-of-speech tagging*). Na

simplificação textual, em alguns casos, como em sentenças longas e ambíguas, por exemplo, é necessária uma análise mais aprofundada para as informações serem melhor aproveitadas nos estágios seguintes.

Considerando que a simplificação do texto opera individualmente sobre cada sentença, o estágio de análise deve identificar as estruturas que serão simplificadas (os *chunks*, grupo de palavras ou sintagmas) e definir adequadamente os pontos de articulação onde os *chunks* serão logicamente marcados e referenciados. A identificação dos *chunks* ocorre com base nas informações etiquetadas em cada palavra do texto. Nessas estruturas (*chunks*) são definidas as marcas referentes às expressões que delimitam sintagmas nominais e sintagmas verbais, como conjunções, pronomes, nomes, etc.

Em alguns casos, há a necessidade de resolver ambiguidades na relação entre esses termos e o *chunk* que eles referenciam, de forma que o estágio seguinte possa operar corretamente. Em Siddharthan [SID02] são descritos problemas de ambiguidade para determinar os termos que delimitam uma oração principal e suas subordinações. Primeiramente é abordada a decisão sobre como definir corretamente o termo referenciado pela conjunção relativa no caso da ocorrência de orações apositivas (3) e da estrutura de frase noun preposition noun (4). As orações relativas estão sublinhadas e os termos referentes à conjunção relativa estão entre colchetes:

- (3) *"She was an inspirational lady", says [Laura Dobson], a freshman at the University of South Carolina, who had Mrs. Yeargin in the teacher-cadet class last year.*
- (4) *The pace of life was slower in those days, says [51-year-old Cathy Tinsall] from [South London], who had five children.*

A solução para o primeiro caso baseia-se na ideia de que uma sequência de orações apositivas aponta para uma mesma referência (um determinado *chunk*). Assim, basta decidir qual é o *chunk*, entre todos da sentença, que melhor preserva o sentido em cada parte desta. Para o exemplo (3), significaria escolher *Laura Dobson* como o termo referenciado pelos demais *chunks*. No segundo caso, uma abordagem de aprendizagem de máquina, referente ao tipo de relação com nomes que cada pronome relativo (*who* e *which*) pode exercer, resolve o problema. Isso é usado na fase de regeneração quando necessário substituir um pronome referente a uma expressão preservando a coesão anafórica. Verificar essas relações, no entanto, exige o acesso à informação hierárquica

de um sistema léxico de referência (uma *WordNet*). No exemplo (4), a decisão sobre a referência do pronome *who* apontaria para o termo “51-year-old Cathy Tinsall” uma vez que o sistema estivesse treinado para relacionar esse pronome a uma entidade com personalidade.

Há ainda a tarefa de determinar os limites de uma sentença para que as orações e suas relações sejam extraídas corretamente. Em Siddharthan [SID02] é definida uma heurística para detecção dos limites de orações relativas (no inglês) com base na análise de 1136 sentenças. Esse estudo sobre a configuração das sentenças permite observar comportamentos como a ocorrência das orações apositivas. Estas orações, por exemplo, referenciam o sintagma nominal, sempre precedendo a oração relativa. Com base neste comportamento, é definido o tratamento adequado para a detecção de limites da sentença.

3.2 Estágio de Transformação

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 [CHA96]. Exemplo disso é a regra citada em (5):

$$(5) \quad X:NP, \text{ RelPron } Y, Z \rightarrow W X:NP Z. X:NP Y$$

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 ($, \text{ RelPron } Y ,$) e seguida por (Z), sendo (Y) e (Z) sequencias 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 frequentemente usadas como prioritárias, evitando, assim, transformações improdutivas.

O conjunto de regras para simplificação de textos é definido de acordo com um determinado domínio. Assim como existem regras que podem ser comuns a domínios

disjuntos, existem regras que só se aplicam a um domínio específico, o que torna a criação manual das regras um processo custoso. Em Chandrasekar e Srinivas [CHA97], por exemplo, é apresentada uma implementação para indução automática das regras de simplificação a partir de um *corpus* alinhado e anotado de textos originais e suas versões simplificadas. Esses dados são submetidos a um *parser*, que determina as relações de dependência entre as sentenças. As regras de simplificação são posteriormente induzidas por meio de técnicas semelhantes às de aprendizagem de máquina.

3.3 Estágio de Regeneração

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. Alguns desses riscos são meramente uma questão de estilo de escrita, como a forma de se reescrever os sintagmas nominais referenciados. Outros, como a ordem da frase, a preservação das relações retóricas e o tratamento de referências anafóricas são vitais para preservar o significado do texto.

No momento de separar uma sentença, por exemplo, é necessário garantir que o tipo de relação que ocorre (representado pela conjunção) não será alterado. Para garantir isso é necessário inserir uma nova *cue-word*¹ no texto. Essa questão é resolvida associando um grupo de conjunções de um determinado tipo de relação a uma conjunção do mesmo tipo e que será usada para manter o sentido nas frases separadas [SID06]. A Tabela 3.1 apresenta exemplos de relação entre uma lista de conjunções e uma determinada *cue-word* correspondente

Tabela 2 - Relação entre as conjunções e as *cue-words* [SID06]

Conjunções	Relação	Cue-Word
<i>although, though, whereas, but, however</i>	<i>(a, Concession, b)</i>	<i>But</i>
<i>or, or else</i>	<i>(a, Anti-Conditional, b)</i>	<i>Otherwise</i>
<i>because</i>	<i>(a, Cause, b), (b, Result, a)</i>	<i>So</i>
<i>And</i>	<i>(a, And, b)</i>	<i>And</i>

¹ Palavra usada para sinalizar alguma informação relevante. Nesse caso, sinaliza a conjunção que será usada para manter a relação entre as sentenças.

Em geral, a ordem das sentenças é preservada durante o estágio de transformação. Entretanto, podem ocorrer transformações em que a ordem precise ser alterada para manter a coerência do texto. Os casos de mudança de voz podem modificar a função gramatical das frases referenciadas e alterar a ordem que essas serão apresentadas. Tanto a mudança de voz, quanto a mudança da ordem das sentenças podem quebrar as referências anafóricas existentes e afetar a habilidade do leitor em identificar a relação dos pronomes corretamente. Sendo assim, é necessário decidir qual a ordem em que serão apresentadas as novas sentenças geradas.

Em Siddharthan [SID06], são abordadas algumas dessas exceções. A primeira diz respeito a escolha da *cue-word*, considerando que a relação da sentença original pode ser de fácil representação e, assim, a ordem das sentenças seria invertida (relação de causa e resultado apresentada na imagem anterior). Outra ocorrência é o caso de decidir a posição de uma oração adjetiva quando extraída da sentença no processo de simplificação. Para manter o sentido, a oração separada deve ser reescrita na sequência da oração principal, como representado em (5), por exemplo.

Além disso, algumas regras de simplificação, como orações relativas ou apositivas, por exemplo, exigem a separação da sentença e a duplicação do sintagma nominal referenciado – como apresentado em (2), por exemplo. Dependendo da estrutura do sintagma nominal, é necessário gerar uma nova expressão (baseada nele) para que a frase torne-se mais simples à compreensão do leitor. Nesse sentido, Siddharthan [SID06] propõe um algoritmo capaz de gerar expressões de referência (sintagma nominal referenciado) em domínios abertos como textos de jornal, por exemplo. Essa proposta não segue o método tradicional, sendo independente um esquema de classificação de adjetivos, mas sim utilizando uma *WordNet* de antônimos e sinônimos para decidir entre as possibilidades para fazer a referência ao sintagma nominal.

Quando o texto é simplificado, algumas estruturas podem ser substituídas por expressões (sintagmas nominais, por exemplo) que resultam em uma retomada, total ou parcial, de outros segmentos no texto. Sendo assim, a etapa final da fase de regeneração trata da reorganização de referências anafóricas dos textos, visto que durante o processo de reescrita do texto pode ocorrer uso de pronomes de forma equivocada. Para resolver essa questão, Siddharthan [SID06] descreve a aplicação de um algoritmo para decisão do uso de pronomes baseado em uma janela de três sentenças (a sentença simplificada que está sendo analisada e duas anteriores a essa). A sentença simplificada é, então, comparada à sentença original no que se refere as suas relações no uso dos pronomes.

Para cada diferença encontrada, o algoritmo que gera expressões de referência é aplicado.

4 TRABALHOS RELACIONADOS

Conforme o conteúdo apresentado anteriormente, é aqui relatado a análise de dois trabalhos referentes ao tratamento e análise de corpora paralelos. Sendo que no primeiro caso, os textos envolvem língua de sinais e no segundo caso os textos são formados por simplificações.

4.1 Corpus Bilíngue para a DGS – Língua de Sinais Alemã

O trabalho apresentado em Bungeroth *et al.* [BUN06] faz parte de um esforço maior de pesquisa que visa à construção de uma ferramenta de tradução automática bidirecional entre a Língua Alemã e a Língua de Sinais Alemã (DGS - Deutsche Gebärdensprache – German Sign Language). Nesse sentido, apresenta a criação de um corpus cujos textos são provenientes de produções visuais (vídeos) de previsões do tempo que combinam a Língua Alemã falada e a DGS.

A escolha de um domínio restrito como a previsão de tempo se deve a alguns fatores que colaboram com qualidade do corpus:

- sentenças semelhantes ocorrem com bastante frequência, visto que as estruturas dos textos de previsão de tempo não possuem muita variação;
- o léxico da língua visual-espacial considerada (DGS) é pequeno, se for comparado com o léxico da Língua Alemã (que tem cerca do dobro do tamanho);
- as estruturas gramaticais específicas de línguas visual-espacial, tais como o uso de expressões faciais, que são bastante difíceis de serem modeladas, são menos comuns nesse domínio.

O processamento e armazenamento dos textos em língua de sinais ocorre com a transcrição do conteúdo apresentado nos vídeos para uma notação especificada pelo grupo de pesquisa em surdez e língua de sinais de Aachen (Deaf and Sign Language Research Team - DESIRE). Como exemplo desta notação, dá-se a seguinte sentença, tomada do corpus descrito neste trabalho relacionado:

HOCH++ ATLANTIK WACHSEN-(mehr)-hn

Os três sinais em DGS são descritos pelas seguintes marcações: *HOCH*, *ATLANTIK* e *WACHSEN*, em alemão. Símbolos repetidos são marcados com um duplo sinal de adição (++). A marcação “-(mehr)” refere-se ao formato assumido pela boca do interlocutor durante a sinalização, enquanto “-hn” refere-se à marcação de assentimento feito pelo interlocutor com a cabeça, durante a enunciação da sentença. A figura 5 representa o uso dessa notação para o registro dos textos

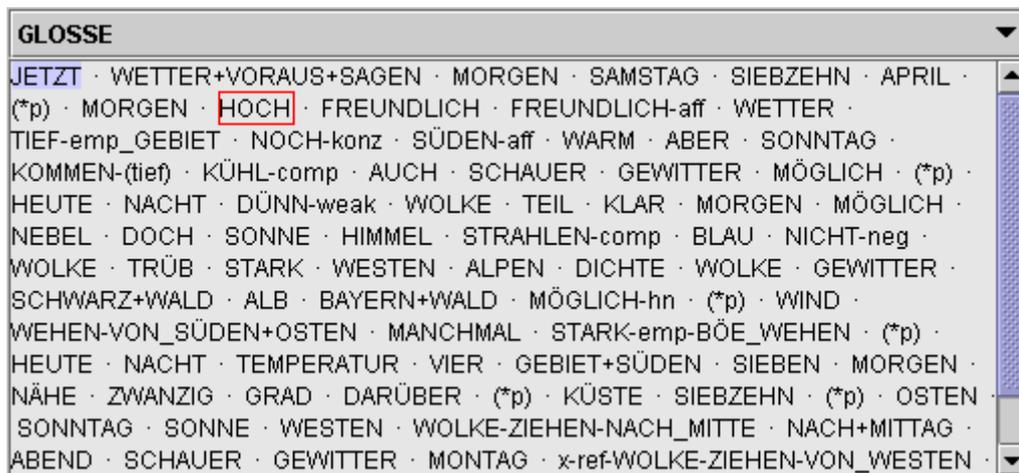


Figura 5 - Exemplo de uso da notação [BUN06]

O corpus gerado consiste de três elementos: os arquivos de vídeo com anotações, um corpus bilíngue de sentenças e um corpus monolíngue em Língua Alemã.

Os arquivos de vídeo estão em formato *Motion Picture Expert Group* (MPEG) e são anotados por um surdo, cuja língua materna é DGS, com o auxílio da ferramenta ELAN². Essas anotações são armazenadas em um formato XML específico da ferramenta o que permite diferentes tipos de marcação em um mesmo intervalo de tempo. O processamento do texto ocorre com seis tipos de marcação

- Os sinais são marcados na notação específica para marcação de escrita de sinais, juntamente com uma marcação referenciando o tempo de início e fim da execução de cada sinal no vídeo;

² Disponível em: <http://www.mpi.nl/tools/elan.html>.

- As sentenças em DGS são marcadas e delimitadas por etiquetas específicas da ferramenta, como (p*) para parágrafo, por exemplo;
- As sentenças em Língua Alemã e DGS são alinhadas, segundo as marcas de delimitação executadas no estágio anterior;
- As sentenças faladas em Língua Alemã recebem marcação específica de início e fim de execução;
- Para 45 vídeos, é feita a anotação de características de cada palavra marcada e alinhada no corpus (part of speech tagging);
- Para 20 vídeos, é feita a marcação do local do sinal no espaço de sinalização.

A Figura 6 apresenta a ferramenta ELAN, usada para anotação, que atua especificamente em cinco estágios do tratamento do corpus: reescrita na notação específica, classificação das palavras, segmentação das sentenças na língua de sinais, segmentação das sentenças em Alemão e segmentação das sentenças faladas em Alemão.

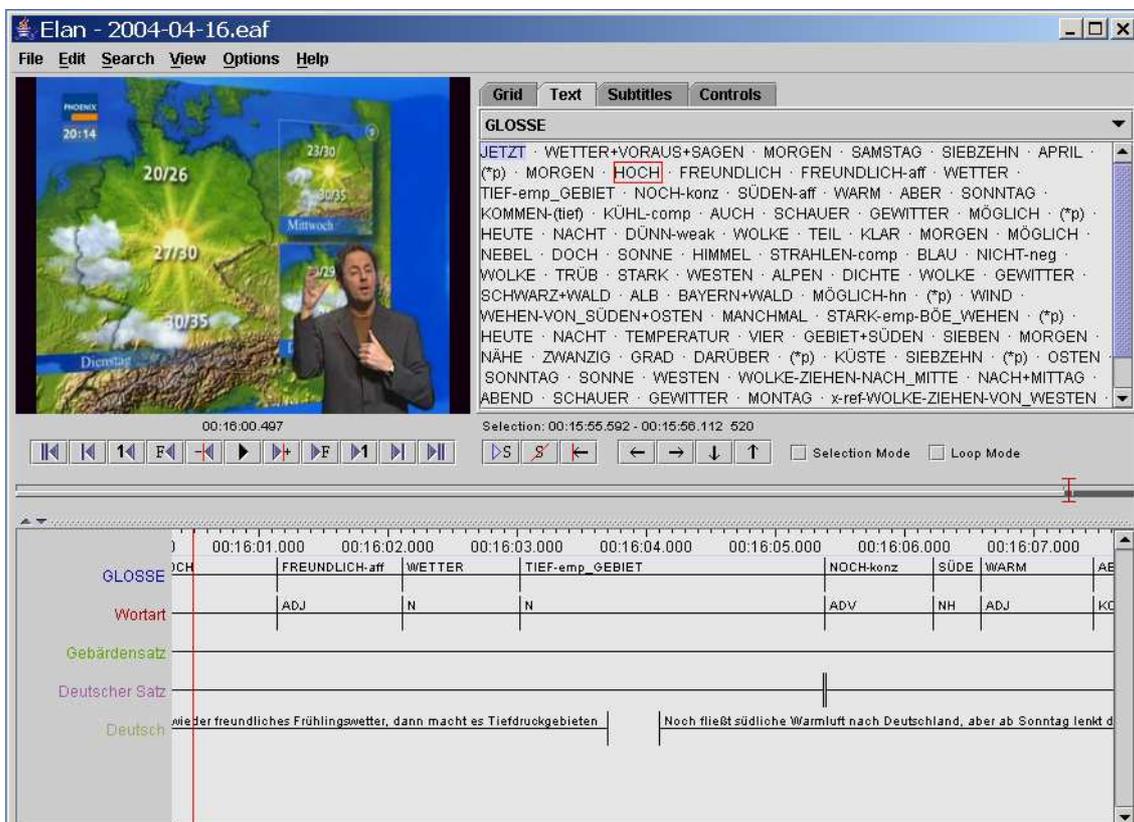


Figura 6 - Ferramenta ELAN [BUN06]

A partir do processo de anotação, um corpus de textos bilíngues é extraído dos arquivos gerados pelo ELAN. Em uma fase de pós-processamento desses textos, as sentenças longas em Alemão são divididas em partes, de forma que se aproximem das sentenças em DGS. O limite das sentenças em DGS é determinado empiricamente pelo baixar das mãos ou por uma definição prévia do intérprete. Além disso, tudo que não foi traduzido entre a Língua Alemã e a DGS foi removido do texto. A Tabela 3 mostra exemplos do corpus bilíngue.

Tabela 3 - Fragmentos de textos [BUN06]

DGS	<i>JETZT WETTER+VORAUS+SAGEN MORGEN DONNERSTAG ZWO" LF MAI.</i>
Alemão	<i>Und nun die Wettervorhersage f"ur morgen, Donnerstag, den zw"olften Mai.</i>
Português	E agora a previsão do tempo para amanhã, 12 de maio.
DGS	<i>ABER-konz WETTER FREUNDLICH LANG-neg.</i>
Alemão	<i>Das freundliche Wetter ist aber nicht von Dauer.</i>
Português	Mas o tempo amigável tem curta duração.

Como informação adicional, foi elaborado um corpus monolíngue em Língua Alemã, contendo as previsões do tempo do ano de 1999 extraídas do site do canal Phoenix. Esse corpus permitiu o aperfeiçoamento do sistema de tradução de línguas de sinais, o que consiste no objetivo último desse trabalho.

A principal contribuição, no contexto da pesquisa aqui apresentada, diz respeito ao processo de tratamento do corpus paralelo. Isso compreende desde a escolha da notação para representar linearmente a língua de sinais até a alternativa (ferramenta ELAN) para armazenar os textos e suas anotações. Em relação à escrita de sinais, não é necessário descartar outras formas de representação, mas é evidente que uma notação facilita a produção dos textos bem como a forma de armazenamento desses.

4.2 Análise de um corpus para extração de regras de simplificação.

O trabalho apresentado por Petersen e Ostendorf [PET07] tem o objetivo de orientar o trabalho de professores no ensino de uma segunda língua a estudantes. Para isso, é feita uma análise sobre um corpus de textos originais e suas versões simplificadas devidamente alinhadas. Dentro deste escopo, é mostrada a possibilidade de

desenvolvimento de ferramentas que possam ser utilizadas para simplificação automática de textos, eliminando, assim, ao máximo, o trabalho humano nessa atividade. O corpus em questão é formado por notícias diversas, contendo 104 textos originais e simplificados manualmente, através de operações nas sentenças tais como exclusão de sentenças, divisão de sentenças, modificação de vocábulos, etc.

No contexto dessa pesquisa, são discutidas as seguintes questões de pesquisa, dentro da análise do corpus:

- Que diferenças de características de palavras e tipos de frases são encontradas entre as sentenças originais e as simplificadas?
- Quais são as características de uma sentença que é dividida quando um texto é simplificado?
- Quais são as características de uma sentença que é excluída quando um texto é simplificado?

A análise estatística (Tabela 4) mostra que o texto simplificado diminui consideravelmente o número de palavras utilizadas (cerca de 30% a menos que o texto original), mas essa diminuição brusca não é verificada no número de sentenças (cerca de 3% a menos que o texto original).

Tabela 4 - Estatísticas sobre o corpus analisado [PET07]

	Original	Simplificado	Redução
Total de sentenças	2539	2459	3%
Total de palavras	41982	29584	30%
Tamanho médio das sentenças	16.5	12.0	27%

Na etapa de comparação dos textos originais e suas versões simplificadas, para fins de extração de regras para a simplificação automatizada, é necessário fazer o alinhamento dos mesmos. Este alinhamento foi feito por um falante nativo do Inglês e baseado em instruções desenvolvidas por trabalhos de alinhamento de corpora comparáveis. Essas instruções direcionam o anotador humano a marcar as sentenças, na versão original e na simplificada, para que informação em pelo menos uma sentença. Entretanto, há alguns casos em que as sentenças da versão são apagadas ou modificadas gerando um tipo diferente de relação. Em certos casos, inclusive, não há como fazer um mapeamento direto das sentenças, visto que elas são totalmente reformuladas e reescritas, guardando pouca correspondência sintática com a sentença original (porém, a correspondência semântica é obrigatoriamente mantida). A Tabela 5 traz alguns exemplos de sentenças originais, e suas resultantes simplificadas.

Tabela 5 - Fragmentos de textos analisados [PET07]

Divisão	
Original	<i>Keith Johnson is the Makah Tribe Spokesman, *** and he comments, "We made history today".</i>
Simplificada	<i>Keith Johnson is the Makah Tribe Spokesman. He said, "We made history today".</i>
Edição	
Original	<i>Congress gave Yosemite the money to repair damage from the 1997 flood.</i>
Simplificada	<i>Congress gave the money after the 1997 flood.</i>
Reescrita	
Original	<i>The park service says the solution is money.</i>
Simplificada	<i>Why hasn't the National Park Service kept up the park repairs? There is a lack of money.</i>

Na terceira sentença do exemplo, nota-se uma reescrita total da mesma, visto que a sentença original é tão dependente do contexto que todas as informações adicionais para a compreensão do assunto têm de ser incluídas na sentença simplificada.

Desta análise, passou-se a extrair padrões das sentenças simplificadas a fim de tentar definir as regras para a simplificação automática de textos. Um dos passos, dentro deste contexto, é definir as regras para escolha de sentenças a serem divididas (*split sentence*). Com base nas estruturas de frase analisadas, foram extraídas as seguintes regras:

- sentenças com número de palavras menor que 19 tendem a não ser divididas;
- sentenças com número de palavras maior que 24 tendem a ser divididas,
- sentenças com comprimento médio (tamanho menor que 24 palavras), compostas de sintagmas nominais curtos, tendem a não ser divididas.

No caso das sentenças excluídas, a análise detectou que o seu conteúdo, mais que qualquer outro fator, determina a sua exclusão no texto simplificado. Nesse contexto, os seguintes fatores puderam ser considerados como motivos de exclusão de sentenças:

- posição da sentença no documento;
- posição da sentença no parágrafo (caso seja a primeira ou a última, é provável a exclusão);
- presença de citação direta na sentença;
- porcentagem de *stop words* (palavras sem relevância semântica na análise do texto);
- frequência direta e inversa de uma ou mais palavras presentes na sentença.

A partir da comparação entre as sentenças alinhadas definindo tipos de ocorrência de simplificação, a contribuição do segundo trabalho, no âmbito dessa pesquisa, refere-se análise sobre os textos produzidos com a finalidade de melhorar a qualidade das anotações incluindo possibilidades de operações e detalhamento das mesmas.

5 FERRAMENTAS PARA SIMPLIFICAÇÃO TEXTUAL DE PORTUGUÊS

Conforme citado anteriormente, 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 [ALU08] 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 Anotação de Simplificação e o Portal de Corpora Simplificados, por serem diretamente relacionados ao objetivo do trabalho aqui apresentado.

5.1 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 o suporte à anotação de simplificação com a finalidade de apoiar um ser humano nessa função. Nesse sentido, foi desenvolvido um editor (através de uma interface *web*) que apoia a tarefa de simplificação e registra todas as decisões tomadas pelo usuário que está simplificando textos [PER08].

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 pode fazer a simplificação manual do texto seguindo recomendações gerais (como manter a ordem SVO, transformar voz passiva para voz ativa e encurtar sentenças, por exemplo), mas sem qualquer compromisso quanto às regras que serão aplicados na simplificação. A esse processo é dado o nome de simplificação natural. Neste passo, o linguista pode fazer uso do “Léxico” como apoio para ressaltar as palavras consideradas complexas e os

marcadores discursivos ambíguos ou não usuais. A ferramenta dá suporte à anotação de cada regra aplicada pelo usuário 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 *Specia et al.* [SPE08] na forma de um manual criado no escopo do projeto. Nesse processo, além do “Léxico”, o usuário tem o auxílio do *parser* Palavras [BIC00] que tem a funcionalidade de ressaltar os pontos em que o texto deve ser alterado para torná-lo mais simples. A Figura 7 mostra os passos pelos quais o texto passa em seu processo de simplificação pelo usuário.

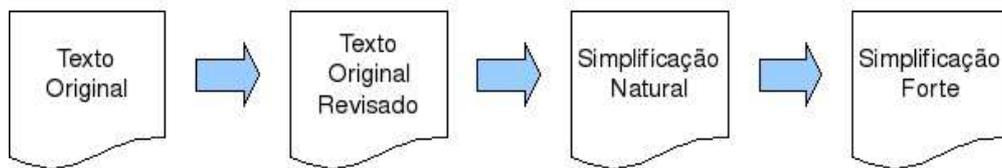


Figura 7 - Processo de simplificação [SPE08]

Nos tópicos a seguir, serão descritas – em maior detalhe – as características deste Editor.

5.1.1 Interface e Funcionalidades

A interface do Editor (Figura 8) é composta basicamente por uma tela principal em que o usuário visualiza a lista das produções já criadas e simplificadas, bem como tem a possibilidade de acessar as demais funcionalidades do sistema: Criar uma nova produção, Visualizar produções e alinhamentos existentes, Simplificar produções e Excluir produções.

PorSimples Editor de Anotação de Simplificação
Simplificando o Português

Produções

Criar nova produção

Existem 109 produções

<< Anterior 1 2 3 4 5 6 Próximo >>

Título	BTexto Original - Natural	BTexto Natural - Forte	Remover	Alinhamentos
Palometas atacam veranistas na Fronteira Oeste	Simplificação Natural	Simplificação Forte		
Projeto propõe distribuição de uniforme à rede pública	Simplificação Natural	Simplificação Forte		
EUA enfrentam nova denúncia de massacre	Simplificação Natural	Simplificação Forte		
Estado revê estratégias contra ladrões de carros	Simplificação Natural	Simplificação Forte		

Figura 8 - Lista de Produções no Editor de Anotação de Simplificação [CAS08]

Além da tela principal, o sistema disponibiliza um conjunto de telas responsável pelo fluxo de criação, edição, anotação e simplificação (natural e forte) dos textos de entrada. Uma vez criado o texto, esse pode ser editado antes do sistema se redirecionar para a interface de Anotação de Simplificação (Figura 9).

:: Título Natural | **Léxico** | **Léxico** | **Sintático**

Palometas atacam veranistas na Fronteira Oeste.

:: Texto Natural | **Léxico**

O ano era 1978. As salas de cinema de todo o mundo exibiam uma produção do diretor Joe Dante. Na produção do diretor Joe Dante, um cardume de piranhas escapava de um laboratório militar. O cardume de piranhas atacava participantes de um festival aquático.

Quase 30 anos depois, banhistas assustados estão se afastando do principal balneário de Uruguaiana. O balneário fica na Fronteira Oeste. Palometas(espécie de piranha) mordem mais de 20 pessoas. As palometas vivem nas águas da barragem Sanchuri. A barragem Sanchuri fica na margem da Br-472. A barragem Sanchuri fica a 40 quilômetros da cidade.

- Os ataques se tornaram mais frequentes. Por isso, aconselhamos mais cuidado- diz o subprefeito Nei Pinto. As mordidas em pés e canelas de banhistas não são novidade. O

- Um amigo meu ficou seriamente ferido. Um adolescente Lindomar Menezes, 15 anos.

Ele e amigos, como Giovane Silva Ferreira, 13 anos, passam as tardes pescando o peixe. Depois, o peixe levado para uma associação de artesãos. A associação faz o curtimento da pele do animal.

Com o fim da piracema, será realizado no domingo o

:: Título Simplificado

Palometas atacam veranistas na Fronteira Oeste

:: Texto Simplificado

O ano era 1978. As salas de cinema de todo o mundo exibiam um filme do diretor Joe Dante. no filme, um cardume de piranhas escapava de um laboratório militar. O cardume de piranhas atacava participantes de um festival aquático.

Quase 30 anos depois, banhistas assustados estão se afastando do principal balneário de Uruguaiana. O balneário fica na Fronteira Oeste. Palometas(espécie de piranha) mordem mais de 20 pessoas. As palometas vivem nas águas da barragem Sanchuri. A barragem Sanchuri fica na margem da Br-472. A barragem Sanchuri fica a 40 quilômetros da cidade.

- Os ataques se tornaram mais frequentes. Por isso, aconselhamos mais cuidado- diz o subprefeito Nei Pinto. As mordidas em pés e canelas de banhistas não são novidade. O

Sentença: 3 | **:: Operações**

Fazer reescrita simples
Dividido em 2 sentenças

Remover Operação | **Detalhar Operação**

Figura 9 - Interface de anotação de simplificação [CAS08]

A interface de Anotação de Simplificação 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 destaque no texto e sugere a substituição de palavras consideradas complexas com base em um léxico de palavras simples e marcadores pouco usuais ou ambíguos. No canto inferior direito da tela, 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 ressalta os trechos que prejudicam a compreensão de um texto e que precisam ser simplificados. A Figura 10 apresenta a tela de simplificação forte, em que o usuário é apoiado pelo processador sintático do editor. Neste exemplo, está selecionada a palavra “e” (destacada), a qual está classificada como indicador de orações coordenadas.

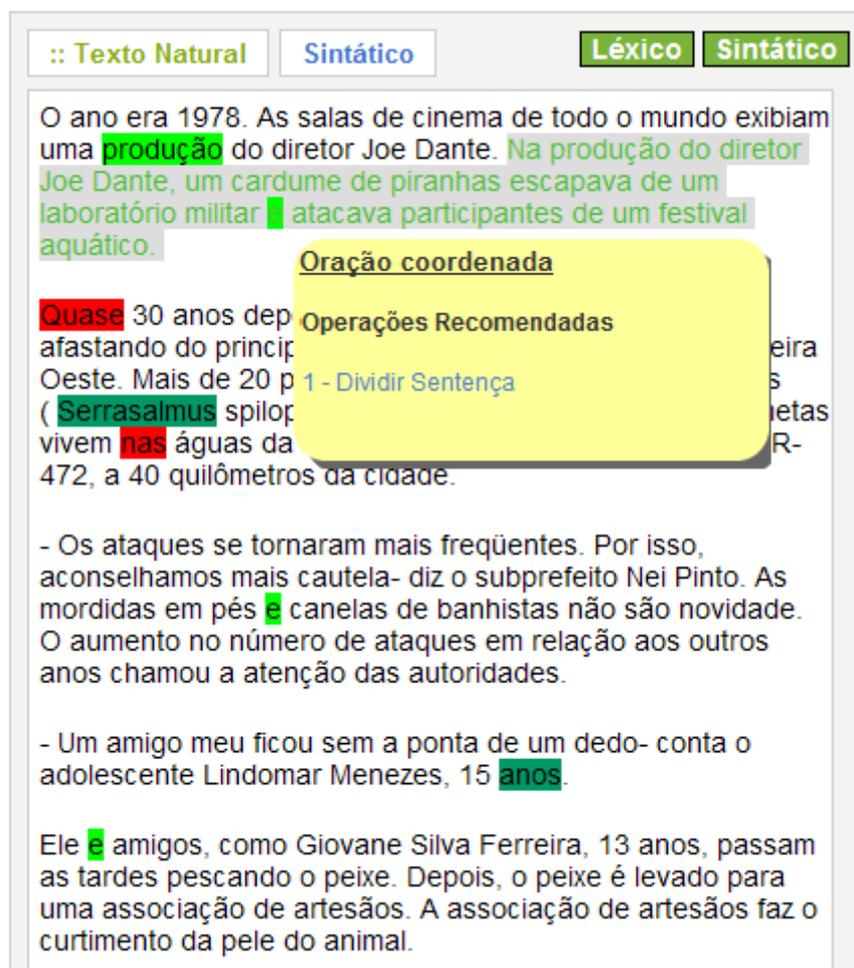


Figura 10 - Interface de Simplificação Forte [CAS08]

5.1.2 Operações e Alinhamento

O corpus paralelo de simplificação é gerado relacionando a versão original e as versões simplificadas do texto. A partir disso, duas abordagens são exploradas no contexto do projeto. Primeiramente, é feita uma abordagem simbólica que usa regras definidas manualmente, com base na análise de textos simplificados em português, na forma de sugestões dirigidas ao usuário que faz a simplificação. Assim, com o trabalho de um linguista, é criado um corpus paralelo anotado que serve como base para uma abordagem estatística que visa à indução automática das regras de transformação [PER08]. Para explorar essa abordagem simbólica, segundo Caseli *et al.* [CAS08], um manual de simplificação sintática foi elaborado, com base na Plain Language³, em trabalhos de geração e simplificação automática para a língua inglesa [SID06], na análise de textos simples disponíveis na *Web* brasileira (por exemplo, a cartilha Brasileiras e Brasileiros no Exterior⁴ e bulas de alguns remédios do Bulário da ANVISA⁵) e nos fenômenos da gramática do português que tornam um texto complexo.

Dessa forma, são definidas as seguintes operações para simplificação de sentenças dentro de um texto:

- Não Simplifica: não há modificação na sentença;
- Fazer reescrita forte: há modificações complexas na ordem ou no modo de apresentar os fatos na sentença;
- Fazer reescrita simples: há apenas substituições de palavras ou expressões;
- Reescrever em SVO: há modificação na estrutura da sentença de forma que essa obedeça à ordem Sujeito Verbo Objeto (usadas para construções apositivas e relativas bem como para orações coordenadas e subordinadas);
- Mudança de voz: há modificação da voz passiva para voz ativa na sentença;
- Inverter ordem da sentença: há modificação na posição de partes da sentença, ou seja, dois seguimentos distintos da sentença trocam de ordem;
- Dividir sentença: há, a partir da sentença original, a ocorrência de mais de uma sentença na versão simplificada;
- Unir sentença: há uma união da sentença selecionada com a sentença anterior do texto;

³Disponível em: <http://www.plainlanguage.gov>

⁴Disponível em: <http://download.uol.com.br/ultnot/cartilha.pdf>

⁵Disponível em: <http://bulaio.bvs.br/>

- Remove sentença: a sentença original não ocorre no texto simplificado;
- Remove parte da sentença: há remoção de parte da sentença na versão simplificada;
- Substituição léxica: há substituição de uma palavra ou expressão específica no texto com base na lista de palavras consideradas complexas.

Para cada operação, o Editor de Anotação de Simplificação define um parâmetro que determina o comportamento da operação no que se refere ao número de ocorrências da sentença no texto durante o processo de alinhamento: a cardinalidade. Essa propriedade é que define quantas sentenças são produzidas no texto reescrito a partir da sentença anotada. A tabela 6 apresenta a lista de operações e suas respectivas cardinalidades.

Tabela 6 - Cardinalidade das operações de simplificação [PER08]

Operação	Cardinalidade
Não Simplifica	1
Reescrita Forte	1
Reescrita Simples	1
Reescrita SVO	1
Mudança de voz	1
Inversão na ordem da sentença	1
Divisão de sentença	N
União da sentença	-1
Remoção da sentença	0
Remoção de parte da sentença	1
Substituição léxica	1

5.1.3 Arquitetura

O Editor de Anotação de Simplificação é um sistema disponibilizado em ambiente WEB baseado no *framework Ruby on Rails*, tendo como repositório de dados o sistema de gerenciamento de banco de dados MySQL. Segundo Thomas [THO07], o *Ruby on Rails* é um *framework* gratuito que tem a intenção de aumentar a produtividade no desenvolvimento de sites cuja programação está centrada no banco de dados, uma vez que é possível criar aplicações com base em estruturas pré-definidas. Esses aplicativos são escritos em *Ruby*, uma linguagem para criação de scripts orientada a objetos. Isso possibilita o mapeamento das tabelas da base de dados em forma de classes. Além

disso, as aplicações criadas com esse framework são desenvolvidas com base no padrão de projeto Model View Controller (MVC). A camada *model* é responsável por manter o estado do objeto (em relação ao banco de dados), enquanto que a camada *view* fica responsável pela interface com o usuário baseado na camada *model*. Contudo, é a camada *controler* que integra as demais camadas, tratando eventos externos (da interface) e solicitando a atualização dos objetos. Nesse contexto, o modelo de dados do Editor de Anotação de Simplificação pode ser representado conforme a figura 11.

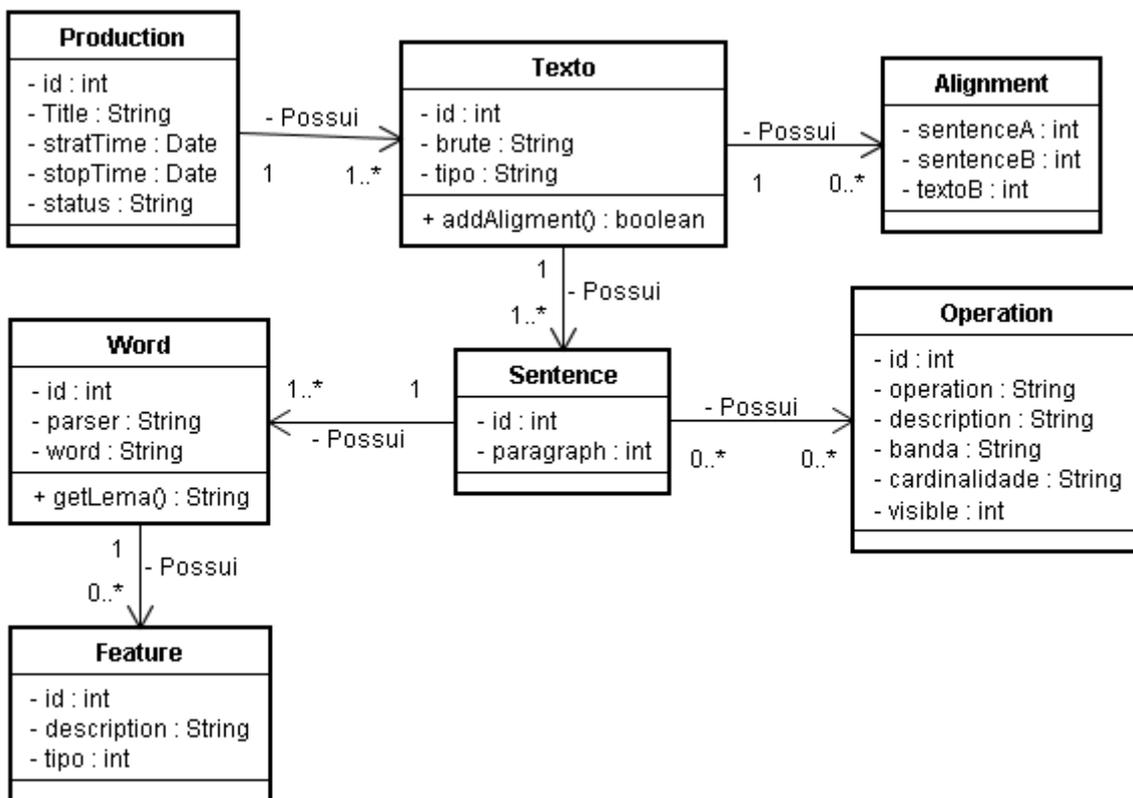


Figura 11 - Estrutura de dados do Editor de Anotação de Simplificação [PER08]

Mais especificamente, a classe *Production* representa o registro de uma produção a ser simplificada, ou seja, é o que engloba as diferentes versões de texto criadas pelo editor no processo de simplificação. Nessa classe encontram-se atributos referentes à identificação da produção (*id*), o título da produção (*Title*), o intervalo de tempo em que a produção foi simplificada (*startDate* e *stopDate*), a validade da produção no sistema (*status*) e o tipo da produção (*tipo*), que se refere a sua classificação em forte, natural, ou completa, classificando uma produção quanto ao agrupamento apenas um texto simplificado ou uma coleção desses.

Os textos vinculados a uma produção são representados pela classe *Texto*. Os atributos dessa classe representam sua identificação (*id*), o conteúdo do texto em sua forma bruta (*brute*), ou seja, sem ser processado pelo sistema e o tipo deste texto (*tipo*), podendo variar entre Original, Natural ou Forte. É a partir dos textos que se organizam o alinhamento das sentenças do texto, bem como as sentenças do texto propriamente ditas.

A classe *Alignment*, por sua vez, responsabiliza-se pelo armazenamento da relação entre as sentenças de cada versão de texto de uma produção. É a partir de um texto que as sentenças são alinhadas, de forma que esta classe apresenta atributos que representam a sentença origem à qual também fica relacionado o alinhamento (*sentenceA*); a sentença destino, na qual a sentença origem foi alinhada (*sentenceB*); e o texto a que pertence a sentença destino (*textoB*). Sendo assim, essa estrutura faz referência a sentenças e seus textos, e com quais sentenças essas foram alinhadas.

As sentenças, como parte de um texto, são estruturadas na classe *Sentence*. A classe armazena o resultado de uma segmentação por parágrafo e sentença, representados com os atributos referentes a sua identificação (*id*) e ao número de sequência do parágrafo do texto ao qual a sentença pertence (*paragraph*). É a partir da sentença que se estruturam as palavras do texto e é na sentença que estão relacionadas as operações envolvidas no processo de simplificação.

O processo de simplificação ocorre sobre cada sentença onde ficam armazenadas as operações realizadas. Sendo assim, a classe *Operation* representa a estrutura das operações possíveis no sistema. Os atributos da classe definem a identificação da operação (*id*), o nome da operação (*operation*), uma descrição referente à operação (*description*), uma classificação quanto ao tipo de operação (*banda*), informação sobre o comportamento referente ao número de ocorrências no alinhamento (*cardinalidade*) e a indicação de disponibilidade da operação (*visible*).

Assim como o texto, a sentença também é segmentada e cada palavra é armazenada a partir da classe *Word*. Os atributos da classe definem a identificação da sentença (*id*), a palavra em si (*word*) e as informações morfossintáticas dela (*parser*). Além disso, a classe *Feature* funciona como estrutura auxiliar para registrar informações sobre as palavras com atributos referentes à identificação (*id*), uma descrição referente às características de cada palavra (*description*) – sendo estas extraídas das informações morfossintáticas obtidas pelo *parser* Palavras e de uma lista de marcadores discursivos

extraídos do software DiZer⁶ – e a informação sobre o tipo de característica (*tipo*), podendo esta ser léxica ou sintática.

O *parser* Palavras, segundo Bick [BIC00], é uma ferramenta para realização de etiquetagem morfossintática e semântica em textos da língua portuguesa. É a partir dele que é feito o processo de segmentação e etiquetagem de cada palavra do texto. O formato de saída do seu processamento segue o seguinte padrão:

Token [lema] Informação morfossintática

Sendo que *Token* é a palavra que foi analisada pelo sistema e essa é seguida por sua forma lematizada e sua informação morfossintática. O *parser* apresenta também formatos específicos para a pontuação e para indicar o fim de uma sentença. São eles respectivamente: \$pontuação e <\s>

5.2 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 XCES para o corpus 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. A Tabela 7 apresenta a listagem dos tipos de arquivos gerados pelo portal no formato XCES.

⁶ Disponível em: <http://www.icmc.usp.br/~tasparado/DiZer.htm>

Tabela 7 - Tipos de arquivos de anotação XCES [FRA08]

	Arquivo	Conteúdo do arquivo	Padrão para o nome
ESSENCIAIS	1. Texto cru	Texto cru, ou seja, sem nenhuma etiqueta.	<NomeArquivo>.txt
	2. Cabeçalho	Informações a respeito do projeto e do texto propriamente dito, além da descrição de todas as anotações e os arquivos nos quais elas estão armazenadas.	<NomeArquivo>-xces.xml
	3. Divisão lógica	Partes que compõem o texto, até o nível de parágrafo. Também é denominada de divisão ou marcação estrutural.	<NomeArquivo>-logical.xml
	4. Divisão de sentenças	Sentenças que compõem o texto.	<NomeArquivo>-s.xml
ADICIONAIS	5. <i>Tokens</i>	<i>Tokens</i> (palavras) que compõem o texto.	<NomeArquivo>-token.xml
	6. Características dos <i>tokens</i>	Características (<i>part-of-speech</i> , lema etc.) dos <i>tokens</i> que compõem o texto. As características são ligadas aos <i>tokens</i> por meio dos identificadores definidos no arquivo 5.	<NomeArquivo>-pos.xml
SIMPLIFICAÇÃO	7. Alinhamento	Alinhamento entre as sentenças de um texto e de sua simplificação. Os alinhamentos são especificados com base nos identificadores de sentenças definidos no arquivo 4.	<NomeArquivo>-align.xml
	8. Operações de simplificação	Operações utilizadas na simplificação de um texto. As operações são especificadas para cada sentença do arquivo sendo simplificado e, por isso, fazem referência aos identificadores de sentenças definidos no arquivo 4.	<NomeArquivo>-simplify.xml

O processo de geração dos arquivos no formato XCES ocorre com base nos dados extraídos da base de dados do Editor de Anotações de Simplificação, pois é a partir dele que os textos são registrados no que se refere a anotações, à segmentação e a alinhamento das sentenças e dos textos. Dessa forma, segundo Fradico e Aluísio [FRA08], os arquivos do grupo Essenciais são gerados para todos os três tipos de texto (Original, Simplificação Natural e Simplificação Forte), enquanto os arquivos dos grupos Adicionais e Simplificação são gerados apenas para os dois primeiros tipos. Essa diferença pode ser observada na interface do sistema conforme a figura 12.

PorSimples
Simplificando o Português

Portal de Corpora Paralelos de Simplificação

Buscar

- Textos
- Dicionários
- Marcadores discursivos
- Operações de simplificação
- Construções sintáticas
- Anotações XCES**

Anotações XCES

Selecione arquivos

Original	Natural	Forte
<input type="checkbox"/> Todos	<input type="checkbox"/> Todos	<input type="checkbox"/> Todos
<input type="checkbox"/> Texto	<input type="checkbox"/> Texto	<input type="checkbox"/> Texto
<input type="checkbox"/> Cabeçalho	<input type="checkbox"/> Cabeçalho	<input type="checkbox"/> Cabeçalho
<input type="checkbox"/> Divisão Lógica	<input type="checkbox"/> Divisão Lógica	<input type="checkbox"/> Divisão Lógica
<input type="checkbox"/> Sentenças	<input type="checkbox"/> Sentenças	<input type="checkbox"/> Sentenças
<input type="checkbox"/> Tokens	<input type="checkbox"/> Tokens	
<input type="checkbox"/> Características dos Tokens	<input type="checkbox"/> Características dos Tokens	
<input type="checkbox"/> Alinhamento	<input type="checkbox"/> Alinhamento	
<input type="checkbox"/> Simplificação	<input type="checkbox"/> Simplificação	

Baixar

Figura 12 - Arquivos em anotação XCES do corpus paralelo [FRA08]

6 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 LIBRAS. 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.

Mais especificamente, 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 linguísticos 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 aqui é que esse processo ocorrerá sobre a versão de Simplificação Forte, que não possui versões posteriores de texto. Essa escolha baseia-se na ideia de que é a versão que menos necessita de 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.

Nesse sentido, os tópicos a seguir descrevem – em maior detalhe – as modificações realizadas no Editor de Anotação de Simplificação de forma que suporte o processo de reescrita em LIBRAS.

6.1 Funcionalidades e Interface

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 13.

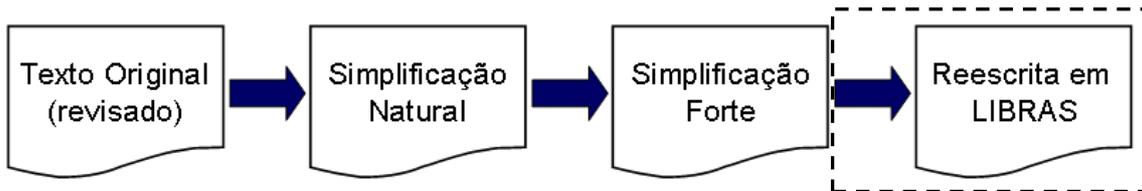


Figura 13 - Novo fluxo de criação de produções

Nesse fluxo, a interface permite a criação de uma produção completa considerando cada uma das versões do texto (Original, Simplificação Natural, Simplificação Forte e Reescrita em LIBRAS) sendo que os três primeiros estágios são obrigatórios e o último é opcional. Assim, a interface foi modificada para dar a opção ao usuário de criar uma nova versão do texto ou simplesmente finalizar a produção. A Figura 14 apresenta a interface que oferece a opção de gerar uma versão em LIBRAS.

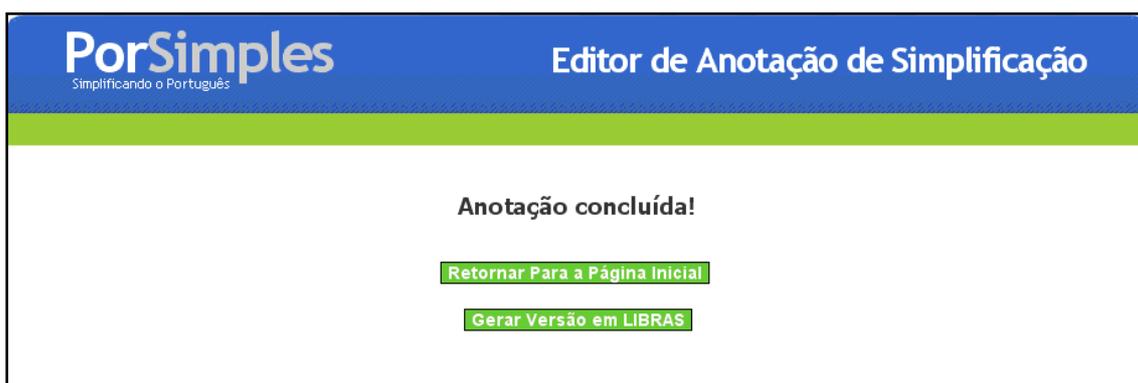


Figura 14 - Acesso a criação da versão em LIBRAS

No entanto, para que a funcionalidade do módulo de reescrita em LIBRAS não se restrinja apenas a novas produções completas, a interface oferece uma opção para criar versões de texto em LIBRAS para as produções completas já existentes no sistema. Mais

especificamente, na tela que lista as produções, foi inserida uma coluna que permite acessar o módulo de reescrita em LIBRAS para criar novas versões ou editá-las. A Figura 15 destaca (com uma borda) a coluna para novas versões em LIBRAS, bem como as opções de criação e edição dessas versões.

PorSimples
Simplificando o Português

Editor de Anotação de Simplificação

Produções do corpus Zero Hora

Nova produção Completa Nova produção Natural Nova produção Forte

Existem 113 produções

<< Anterior 1 2 3 4 5 6 Próximo >>

Título	BITexto Original - Natural	BITexto Natural - Forte	BITexto LIBRAS	Remover	Alinhamentos
Palometas atacam veranistas na Fronteira Oeste	Simplificação Natural	Simplificação Forte	Editar Versão LIBRAS		
Projeto propõe distribuição de uniforme à rede pública	Simplificação Natural	Simplificação Forte	Editar Versão LIBRAS		
EUA enfrentam nova denúncia de massacre	Simplificação Natural	Simplificação Forte	Editar Versão LIBRAS		
Estado revê estratégias contra ladrões de carros	Simplificação Natural	Simplificação Forte	Editar Versão LIBRAS		
Um jogo sem escalção definida	Simplificação Natural	Simplificação Forte	Nova Versão LIBRAS		

Figura 15 - Opção de Reescrita em LIBRAS para produções já existentes.

Uma vez acessado, independente da opção de editar ou criar uma versão em LIBRAS, 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 16 exemplifica o processo de reescrita do texto em LIBRAS. Mais 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.

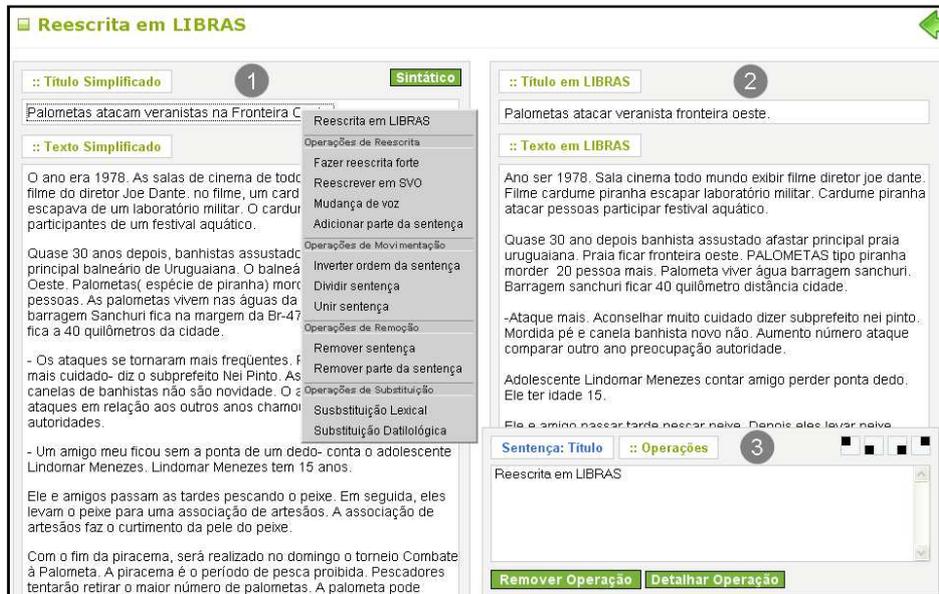


Figura 16 - Interface para reescrita em LIBRAS

6.2 Operações

Diferente do processo de simplificação, a reescrita em uma interlíngua da LIBRAS aproxima-se de uma tradução entre a Língua Portuguesa e a LIBRAS. Nesse sentido, 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.

6.2.1 Reescrita em LIBRAS

Diferente das demais operações, além de registrar uma anotação na sentença que foi marcada, a operação de reescrita em LIBRAS sugere 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 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. A tabela 8 exemplifica a ocorrência dessa operação na reescrita da versão em LIBRAS.

Tabela 8 - Exemplo de uso da operação de Reescrita em LIBRAS

Simplificação Forte	O ano era 1978.
Interlíngua LIBRAS	<i>Ano ser 1978.</i>

6.2.2 Reescrita Forte

De forma semelhante ao processo de simplificação, nessa operação há 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. A tabela 9 exemplifica a ocorrência dessa operação na reescrita da versão em LIBRAS.

Tabela 9 - Exemplo de uso da operação de Reescrita Forte

Simplificação Forte	Além disso, a proposta reanima o debate sobre a obrigação de uma roupa igual para todos os alunos.
Interlíngua LIBRAS	<i>Também de novo discutir roupa igual todos alunos obrigatório dinheiro público tem? Falta.</i>

6.2.3 Mudança de voz

A estrutura de frase da LIBRAS não permite ocorrência de voz passiva. Nesse sentido, essa operação tem a finalidade de garantir o registro de mudança de voz (sempre de passiva para ativa) quando ocorrer. A tabela 10 exemplifica a ocorrência dessa operação na reescrita da versão em LIBRAS.

Tabela 10 - Exemplo de uso da operação de Mudança de voz

Simplificação Forte	Mais de 20 pessoas foram mordidas por palometas (espécie de piranha).
Interlíngua LIBRAS	<i>PALOMETAS tipo piranha morder 20 pessoa mais.</i>

6.2.4 Inversão na ordem da sentença

Em alguns casos, a sentença em Língua Portuguesa pode ser melhor representada em LIBRAS quando a ordem de alguns elementos da sentença é modificada. A tabela 11 exemplifica a ocorrência desse tipo de operação de inversão na reescrita da versão em LIBRAS.

Tabela 11 - Exemplo de inversão de sentença

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

6.2.5 Divisão de sentenças

Dependendo do tipo de relação entre as frases da sentença, pode ser necessário dividir a sentença para melhor expressar a informação original. Com a divisão das sentenças, a relação que dificulta a reescrita em LIBRAS se desfaz e o texto mantém o sentido. A tabela 12 exemplifica a ocorrência dessa operação na reescrita da versão em LIBRAS.

Tabela 12 - Exemplo de uso da operação 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>

6.2.6 União de sentenças

Apesar de se tratar de uma língua complexa e completa, a informação em LIBRAS é, normalmente, expressa de forma direta (evitando redundâncias, por exemplo). Dependendo da forma como uma sequência de sentenças está organizada, pode ser necessária a união de sentenças como forma de torná-las mais diretas. A tabela 13 exemplifica a ocorrência dessa operação na reescrita da versão em LIBRAS.

Tabela 13 - Exemplo de uso da operação de União de sentenças

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 ter 33 anos.</i>

6.2.7 Remoção da sentença e Remoção de parte da sentença

Como já citado, para expressar informação de forma direta, ocorrências de redundância tendem a ser removidas do texto quando reescritas em LIBRAS. Essa remoção pode ser de uma sentença inteira ou de apenas parte dela, sendo assim necessária a opção de escolha entre essas duas operações. A tabela 14 exemplifica a ocorrência da operação de Remoção de parte da sentença na reescrita da versão em LIBRAS.

Tabela 14 - Exemplo de uso da operação de Remoção de parte da sentença

Simplificação Forte	O secretário calculou que a clonagem pode ser o destino de até 40% dos veículos roubados.
Interlíngua LIBRAS	<i>Secretário explicou 40% carros roubados é clonagem.</i>

6.2.8 Substituição Lexical

Nem todas as palavras da língua portuguesa podem ser usadas em LIBRAS para expressar a mesma ideia. Logo, no processo de reescrita, essa operação é necessária, pois muitas palavras acabam sendo substituídas para adequar o sentido da sentença em LIBRAS. A tabela 15 exemplifica a ocorrência dessa operação na reescrita da versão em LIBRAS, com o respectivo detalhamento inserido pelo usuário.

Tabela 15 - Exemplo de uso da operação de Substituição Lexical

Simplificação Forte	O Aeroporto Internacional Salgado Filho fica na Capital.
Interlíngua LIBRAS	<i>Aeroporto internacional SALGADO FILHO lugar capital.</i>
Detalhamento	O verbo ficar foi substituído pela palavra lugar.

6.2.9 Substituição Datilológica

Um recurso para superar o problema da falta de um sinal para um vocábulo é a datilologia (uso do alfabeto manual). Trata-se de uma soletração manual (através de configurações de mão que representam as letras do português) para representar termos técnicos ou nomes de pessoas que não existem em forma de sinais. A tabela 16 exemplifica a ocorrência dessa operação na reescrita da versão em LIBRAS, com o respectivo detalhamento inserido pelo usuário.

Tabela 16 - Exemplo de uso da operação de Substituição Datilológica

Simplificação Forte	O Aeroporto Internacional Salgado Filho fica na Capital.
Interlíngua LIBRAS	<i>Aeroporto internacional SALGADO FILHO lugar capital.</i>
Detalhamento	O nome do aeroporto é representado com o alfabeto manual (o uso do alfabeto manual é destacado através da reescrita da palavra em letras maiúsculas).

6.3 Cardinalidade

O conceito de cardinalidade de operação (parâmetro que determina o comportamento da operação no que se refere ao número de ocorrências da sentença no texto durante o processo de alinhamento) não é alterado no funcionamento do módulo. Uma vez finalizada a reescrita do texto, o sistema executa 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. A tabela 17 apresenta a lista de operações e suas respectivas cardinalidades.

Tabela 17 - Cardinalidade das operações de simplificação

Operação	Cardinalidade
Reescrita em LIBRAS	1
Reescrita Forte	1
Mudança de voz	1
Inversão na ordem da sentença	1
Divisão de sentença	N
União da sentença	-1
Remoção da sentença	0
Remoção de parte da sentença	1
Substituição lexical	-
Substituição datilológica	-

6.4 Arquitetura

O Módulo de Reescrita em LIBRAS é integrado à arquitetura do Editor de Anotação de Simplificação, ou seja, as estruturas usadas para organização e registro são as mesmas. Entretanto, foram necessárias algumas modificações em estruturas específicas para atender às funcionalidades desse módulo.

A primeira modificação teve o objetivo de dar suporte à reescrita (e, conseqüentemente, à simplificação) do título dos textos. Nesse sentido, o atributo *title* da classe *production* foi replicado para a classe *texto*, permitindo, assim, que cada texto possa ter uma versão própria do título (Original, Natural, Forte e LIBRAS). Além disso, dentro dos objetivos do Editor de Anotação de Simplificação, o título é tratado como uma sentença específica do texto, ou seja, é processado pelo *parser* Palavras e armazenado

na estrutura *sentence* indicando o atributo *paragraph* como zero para representar uma sentença título.

No que se refere às operações, a principal alteração ocorreu no âmbito da sua visibilidade. Mais especificamente, na estrutura original, o atributo *visible* da classe *operation* define se uma operação pode ser visualizada no menu. Todavia, as operações definidas para o Módulo de Reescrita em LIBRAS são compartilhadas com o restante das funcionalidades do Editor de Anotação de Simplificação. Sendo assim, a comparação entre os estados possíveis para o atributo, considerando a funcionalidade original e as novas necessidade do editor, são representados na Tabela 18, enquanto que a Tabela 19 representa como ficou configurada a visibilidade de cada operação oferecida em todo o sistema (considerando operações de ambas as partes).

Tabela 18 - Definição de estado do atributo *visible*

Editor de Anotação de Simplificação		Módulo de Reescrita em LIBRAS	
Valor	Função	Valor	Função
S	Define que a operação faz parte do menu.	S	Define que a operação faz parte do menu apenas na interface de simplificação.
		N	Define que a operação faz parte do menu em qualquer interface.
N	Define que a operação não faz parte do menu.	C	Define que a operação faz parte do menu em qualquer interface.
		L	Define que a operação faz parte do menu apenas na interface de reescrita em LIBRAS.

Tabela 19 - Relação de visibilidade para cada operação

Operação	Visible
Não simplifica	S
Reescrita em LIBRAS	L
Reescrita Forte	C
Mudança de voz	C
Inversão na ordem da sentença	C
Divisão de sentença	C
União da sentença	C
Remoção da sentença	C
Remoção de parte da sentença	C
Substituição léxica	L
Substituição Datilológica	L
Reescrever em SVO	S
Fazer reescrita simples	S

7 A EXPERIÊNCIA DE USO

A fim de explorar o uso do anotador do Módulo de Reescrita em LIBRAS, no contexto do Editor de Anotações de Simplificação, para produção de textos paralelos em Língua Portuguesa e em LIBRAS, foi realizado uma experiência de uso do mesmo. Complementarmente, a experiência também visa averiguar o uso das operações disponibilizadas nesse módulo na reescrita do texto em LIBRAS, no sentido de melhor adequá-las ao contexto em questão.

Esta experiência de uso foi realizada por uma especialista em LIBRAS com nível superior (Bacharelado em Artes Plásticas), tendo ampla formação como intérprete de LIBRAS. Ela possui 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.

7.1 Atividade realizada

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. Além disso, com a disponibilidade de uma opção de detalhes para uma operação, o usuário ficava livre para registrar comentários ou observações sobre o uso de cada operação.

Os requisitos mínimos exigidos ao especialista, para dar início aos testes foram:

- computador com acesso à Internet;
- navegador Firefox 2.0 ou superior,
- rápido treinamento sobre o uso da ferramenta.

Os textos foram inicialmente reescritos fora da ferramenta (em um editor de textos) e, posteriormente, inseridos no sistema. Esse processo contribuiu para definição de uma forma adequada de fazer uma tradução da Língua Portuguesa para a interlíngua da

LIBRAS, considerando que o trabalho foi realizado por um profissional intérprete e não tradutor. Para compreender a implicação desse detalhe, é importante considerar que não há diferença entre o conteúdo que é produzido por um tradutor e um intérprete. Ambos produzem uma tradução, mas deve haver diferenças na forma como esse conteúdo é traduzido. Mais especificamente, um tradutor costuma pesquisar o termo mais adequado, fazendo uso de dicionários e modificando inúmeras vezes o texto, enquanto que um intérprete concentra-se nos aspectos essenciais da mensagem e necessita de opções de tradução extremamente rápidas e habilidades cognitivas simultâneas. Nesse sentido, para manter uma maior proximidade com o texto original, foi necessário adaptar a aptidão de um intérprete à de um tradutor na produção dos textos na interlíngua da LIBRAS.

A partir da experiência de produção dos textos por parte do especialista, dois tipos de características foram observadas. Primeiramente, buscou-se alternativas para melhorar o uso do sistema em um contexto geral como problemas no registro do conteúdo e das anotações. A partir das anotações registradas, procurou-se observar os comentários e as observações sobre o uso das operações no sentido de melhor adequá-las a uma determinada ocorrência de reescrita.

7.2 Problemas encontrados

O primeiro problema encontrado refere-se ao uso das operações de substituição (operações de substituição lexical e substituição datilológica). Nesses casos, o sistema prevê que sejam registradas, em forma de anotação, tanto a nova palavra quanto a usada na versão original do texto. Entretanto, no fluxo de simplificação, a operação de substituição lexical não é sugerida pelo menu, mas sim quando o sistema faz uma verificação léxica do texto procurando por palavras consideradas complexas, oferecendo a opção de substituição.

Constatou-se que essa verificação léxica não tem sentido para o Módulo de Reescrita em LIBRAS, uma vez que a língua de sinais em uso pode apresentar correspondência para palavras complexas (como fonoaudiologia, por exemplo) e não para palavras simples (como banhistas, por exemplo). Isso levou ao uso dessa operação por meio do menu exatamente como as demais. Entretanto, o sistema não estava considerando o registro de quais palavras estavam sendo trocadas, deixando a funcionalidade de adicionar detalhes como única opção para registrar essa informação.

De forma semelhante, a substituição datilológica precisa registrar qual palavra está sendo reescrita de forma a ressaltar essa informação nas operações realizadas sobre a sentença. Nesse sentido, a solução adotada foi a criação de uma interface que permita ao usuário informar quais são as palavras envolvidas no processo de substituição. A partir disso, a decisão pelo uso da substituição passou a registrar detalhes dessa operação. A Figura 17 apresenta a interface disponibilizada para o uso de operações de substituição.

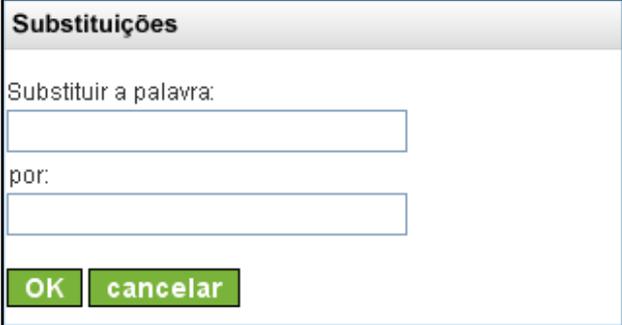
A interface é uma janela de diálogo com o título "Substituições". Ela contém dois campos de texto: o primeiro é rotulado "Substituir a palavra:" e o segundo é rotulado "por:". Abaixo dos campos, há dois botões: "OK" e "cancelar".

Figura 17 - Interface para anotação de substituição

O segundo problema diz respeito a uma configuração errada de cardinalidade em uma das operações. Isso não reflete diretamente no uso do sistema, mas no processo de alinhamento das sentenças dos textos. Mais especificamente, a operação relacionada à remoção de parte da sentença estava configurada com cardinalidade zero, ao invés de um. Assim, o alinhamento considera que a sentença não existe na versão em LIBRAS e acaba alinhando o restante do texto de forma errada. Nesse caso a solução do problema se deu apenas com a correção da cardinalidade da operação, não necessitando modificar o processo de alinhamento.

7.3 Uso das operações

No que se refere ao uso das operações, diferentes aspectos foram observados. Primeiramente, uma parte significativa das ocorrências da operação de união de foi registrada sobre sentenças que já tratavam do resultado de uma divisão no processo de simplificação forte. O que pode indicar que a opção de limitar a reescrita em LIBRAS

apenas a versões de simplificação forte seja prejudicial à produção dos textos. A Tabela 20 exemplifica esse tipo de ocorrência.

Tabela 20 - Exemplo de união de sentenças já divididas

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 que sentenças já divididas voltam a sua construção original, há ocorrência de sentenças que foram divididas na reescrita em LIBRAS como uma forma de expressar melhor a informação da sentença original (Tabela 21). Entretanto, levando em consideração que se trata de uma ferramenta para simplificação de textos, adjungindo a reescrita em LIBRAS e a simplificação textual.

Tabela 21 - Exemplo de união de sentenças

Simplificação Forte	A neblina causou o cancelamento de 12 vôos. A neblina causou o atraso de 26 pousos e decolagens.
Interlíngua LIBRAS	Neblina culpa cancelados 12 vôos também atrasados 26 pousos e decolagens.

O segundo aspecto observado refere-se ao uso adequado das operações. Nesse ponto, a atenção foi direcionada aos tipos de ocorrência das operações mais utilizadas na reescrita. Nesse contexto, a operação de substituição lexical foi utilizada de três 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 22).

Tabela 22 - 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>pessoas 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 ideia do texto original. Juntamente com outra ocorrência para esse tipo de substituição relacionado ao uso de pronomes para referenciar nomes já citados no texto. Ambos os casos são exemplificados na Tabela 23.

Tabela 23 - Exemplo de substituição lexical

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

Para os dois primeiros casos é interessante diferenciar o tipo de ocorrência em forma de duas novas operações. A primeira serve para identificar uma substituição por carência de um sinal com relação apropriada 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.

Outra operação de frequente utilização refere-se à remoção de parte da sentença. Essa operação 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 24). Para garantir a qualidade das anotações, é interessante a criação de uma operação específica para registrar a remoção de uma palavra específica (Remover palavra), permitido, inclusive, o seu registro.

Tabela 24 - 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 carros carro também golpe seguro.</i>
Detalhamento	Foi retirada a palavra "chamado" por não existir esta expressão em LIBRAS.

Para a operação de inversão da ordem da sentença, as ocorrências se caracterizaram em dois comportamentos distintos. Na primeira, ocorre uma modificação na ordem básica da sentença com o objetivo de posicionar a ideia principal da sentença em primeiro lugar. Em LIBRAS, colocar em evidência um determinado trecho da sentença é denominado Topicalização. Sendo assim, como forma de diferenciar essa inversão, é necessária a criação de uma operação específica para reescrita em LIBRAS: reescrever topicalização. A Tabela 25 apresenta um caso para uso dessa operação.

Tabela 25 - 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. Mais especificamente, 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 26 e 27 exemplificam o segundo caso de inversão.

Tabela 26 - Exemplo de inversão para discurso indireto

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

Tabela 27 - 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>

7.4 Depoimento do especialista

De acordo com o especialista:

O primeiro contato que tive com a proposta de implementação do anotador de textos foi surpreendente. Me pareceu muito simples e prático do ponto de vista funcional, ou seja, mesmo tendo conhecimentos básicos sobre informática, não tive maiores dificuldades de manipular com as ferramentas, e o sistema é de fácil compreensão.

Quanto a aplicabilidade, no sentido de contemplar as necessidades da comunidade surda em relação à língua portuguesa escrita, ou mesmo para profissionais de várias áreas, surdos ou ouvintes, que necessitem da comunicação escrita de forma que seja funcional, ou seja, que exista a possibilidade deste intertexto, português escrito mas com estrutura de Libras, a proposta é absolutamente adequada e eficiente.

Professores de escolas de surdos, instrutores de Libras, intérpretes de Libras, alunos surdos dos cursos de licenciatura Letras/Libras e intérpretes alunos do bacharelado Letras/Libras, são alguns dos profissionais para os quais esta proposta trará facilidades. Por exemplo, os textos na escola poderão ser trabalhados a partir do português escrito, em contraste à reescrita em Libras, oportunizando ao aluno uma metodologia de ensino de segunda língua a partir da comparação com a estrutura da primeira. Este seria um exemplo entre tantos que poderíamos citar enfatizando a aplicabilidade, funcionalidade e adequação do anotador de textos, e portanto sugerimos sua divulgação nos espaços onde possa servir para a melhoria da qualidade de acesso ao conhecimento pela comunidade surda.

8 CONSIDERAÇÕES FINAIS

Um bom trabalho não se resume apenas aos conceitos já dominados pelo autor. É preciso buscar na comunidade científica os pilares para a construção de novos conhecimentos. Baseado nisso, ressalta-se a importância da pesquisa para o satisfatório andamento de um trabalho, tendo as referências teóricas apresentadas o papel de embasá-lo diante do contexto no qual ele está inserido.

A construção do Módulo de Reescrita em LIBRAS é o resultado de uma parceria com o grupo de pesquisa NILC, através da qual se pôde fazer uso de uma arquitetura já definida para o desenvolvimento de uma abordagem diferenciada para a produção de textos em LIBRAS. Propôs-se, assim, uma ferramenta web para o apoio ao processo de produção de um corpus paralelo de textos alinhados na versão da Língua Portuguesa e na versão da LIBRAS.

Os estudos iniciais realizados sobre os aspectos gramaticais da LIBRAS e sobre como ocorre a educação do surdo dentro de um contexto bilíngue refletem não só no conhecimento dessa língua ou nas diferenças entre sua estrutura e a estrutura oral da Língua Portuguesa, mas também na diferença cultural e de aprendizagem que envolve o surdo enquanto este adquire sua segunda língua e produz textos. A partir disso, é possível entender o que faz da LIBRAS – como também das outras línguas de sinais – uma forma de comunicação tão semelhante e, ao mesmo tempo, tão independente da Língua Portuguesa.

Uma vez que a base da arquitetura do trabalho trata de simplificação textual, é feita referência a conceitos ligados a essa linha de pesquisa com a intenção de compreender as necessidades provenientes desse tipo de sistema. Assim, o estudo sobre simplificação textual proporcionou os conceitos iniciais relacionados ao que se detém o objetivo desse trabalho.

Aliado a esse fato, o estudo dos trabalhos relacionados com a produção de um corpus em língua de sinais e com o estudo das relações entre textos originais e simplificados remetem também ao entendimento das necessidades envolvidas no trabalho desenvolvido. Nesse sentido, a importância de se organizar adequadamente um corpus e as possíveis formas de analisá-lo foram as contribuições dos trabalhos estudados nessa pesquisa.

Complementarmente aos estudos bibliográficos, houve um estudo sobre as características do sistema ao qual o módulo desenvolvido foi anexado: o Editor de

Anotação de Simplificação. A partir da análise de sua arquitetura e funcionalidades, foi elaborada a definição das modificações necessárias para a implementação do Módulo de Reescrita em LIBRAS, levando em consideração os aspectos da língua de sinais envolvida e as alterações no uso das operações propostas no sistema. A partir disso, o módulo desenvolvido foi utilizado por um especialista em LIBRAS, de forma a verificar a aplicabilidade do mesmo e a possibilidade de adaptá-lo a novas operações de reescrita para LIBRAS.

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 bilíngue 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 a informação por parte do surdo no sentido de oferecer a possibilidade de uma melhor inteligibilidade dos textos durante o aprendizado de sua segunda língua.

8.1 Trabalhos Futuros

No que se refere a trabalhos futuros, após a execução deste trabalho, vislumbra-se algumas possibilidades para complementação e continuação dessa pesquisa no que diz respeito à manutenção das funcionalidades do sistema.

- Produção do Corpus Paralelo: Em virtude de verificar a validade de uma abordagem de simplificação textual para produção de textos em LIBRAS, a produção do corpus paralelo não foi a prioridade desse trabalho. Entretanto, é possível estender o uso da ferramenta a um conjunto maior de especialistas em LIBRAS com o objetivo de alimentar o corpus proposto.
- Inclusão de novas operações: A partir da experiência de uso realizada, novas possíveis operações foram observadas. Para melhorar a qualidade das anotações, recomenda-se a inclusão dessas novas operações assim como a constante análise do corpus para descoberta de novas possibilidades.

- Oferecer mais possibilidades de escrita: A notação de uma interlíngua adotada para a produção advém do objetivo de se usar o léxico da linguagem oral para melhor possibilidade de análise do corpus. Entretanto, é importante destacar que o aprendizado e o acesso à Língua Portuguesa pelo surdo, não exclui a necessidade de seu conhecimento sobre a língua de sinais. Sendo assim, uma alteração no sistema que permita uma escrita em uma notação capaz de melhor representar as características da língua de sinais permitiria um aprofundamento nos estudos linguísticos relacionados a ela.
- Avaliação da Inteligibilidade do Textos: Seguindo a proposta de garantir acesso a informação e inclusão digital, 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.
- Ferramenta de Tradução Automática: O uso de tradutores automáticos da Língua Portuguesa para LIBRAS proporciona uma integração linguística entre surdos e ouvintes e permite que a pessoa surda tenha pleno acesso a meios de comunicação e entretenimento tais como: jornais, revistas, livros, televisão, teatro e cinemas, entre outros [ACE07]. Conforme citado na introdução desse trabalho, existem abordagens nessa linha de pesquisa que podem se beneficiar do corpus alinhado produzido pela ferramenta.

REFERÊNCIAS BIBLIOGRÁFICAS

- [ACE07] ACESSOBRASIL. “Acessibilidade Brasil.” Capturado em: <http://www.acessobrasil.org.br>, Maio 2007.
- [ALU08] ALUÍSIO, S. M. et al. “Towards Brazilian Portuguese Automatic Text Simplification Systems”. Em: 8th ACM Symposium on Document Engineering, 2008, pp. 240-248.
- [BIC00] BICK, E. “The Parsing System “Palavras”: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework”. Aarhus: Aarhus University Press, 2000, 411p.
- [BRI96] BRITO, L. F. “Apostila sobre a estrutura da língua brasileira de sinais”. Capturada em: http://www.ines.org.br/ines_livros/35/35_Principal.htm, Junho 2007.
- [BRO06] BROCHADO, S. M. D. “A apropriação da escrita por crianças surdas”. Em: QUADROS, R. M. (Org.). “Estudos Surdos I”. Petrópolis: Arara Azul, 2006, pp. 284-322.
- [BUN06] BUNGEROTH, J.; STEIN, D.; DREUW, P.; ZAHEDI, M.; NEY, H. “A German sign language corpus of the domain weather report”. Em: 5th International Conference on Language Resources and Evaluation, 2006, pp. 2000-2003.
- [CAM01] CAMPOS, M. B. “Ambiente Telemático de Interação e Comunicação para Suporte à Educação Bilíngüe de Surdos”. Tese de Doutorado em Informática na Educação, Programa de Pós-Graduação em Informática na Educação, UFRGS, Porto Alegre, 2001, 204p.
- [CAM07] CAMPOS, M. B.; SILVEIRA, M. S. “Promoção da Cidadania da Comunidade Surda: o Uso das TICs na Apropriação de sua Língua Materna”. Em: 34th Seminário Integrado de Software e Hardware, 2007, pp. 2232-2246.
- [CAS08] CASELI, H. M.; PEREIRA, T. F.; ALUÍSIO, S. M. “Editor de Anotação de Simplificação: Manual do Usuário”. Relatório Técnico, NILC – ICMC, USP, São Carlos, 2008, 17p.
- [CHA07] CHAN-VIANNA, A. C. “Formulação de inferências e propriedades da interlíngua de surdos na aquisição de português (escrito)”. Em: SALLES, H. M. M. (Org.). “Bilingüismo dos Surdos. Questões Lingüísticas e Educacionais”. Brasília: Cãnone, 2007, pp. 97-118.
- [CHA96] CHANDRASEKAR, R.; DORAN, C.; SRINIVRA, B. “Motivations and Methods for Text Simplification”. Em: 16th International Conference on Computational Linguistics, 1996, pp.1041-1044.
- [CHA97] CHANDRASEKAR, R.; SRINIVAS, B. “Automatic induction of rules for text simplification”. *Knowledge Based Systems*, vol.10, Out 1997, pp.183-190.

- [EST08] ESTELITA, M. B. "ELiS - Escrita das Línguas de Sinais: Proposta teórica e verificação prática". Tese de Doutorado, Programa de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, Florianópolis, 2008, 197p.
- [FEL97] FELIPE, T. A. "Introdução à Gramática da LIBRAS". *Educação Especial - Língua Brasileira de Sinais*, vol. III - Série Atualidades Pedagógicas 4, MEC/SEESP, 1997, pp. 81-123.
- [FEN02] FENEIS. "Lei n.º 10.436, de 24 de abril de 2002. Dispõe sobre a Língua Brasileira de Sinais - Libras e dá outras providências, Brasília, DF, 24 abr. 2002". Capturado em: http://www.feneis.com.br/page/legislacao_resultado.asp?1, Maio 2007.
- [FIN07] FINAU, R. A. "O processo de formação de interlíngua na aquisição de língua portuguesa por surdos e as categorias tempo e aspecto". Em: SALLES, H. (Org.). "Bilingüismo dos Surdos: questões Lingüísticas e Educacionais". Brasília: Cânone, 2007, pp. 161-191.
- [FRA08] FRADICO, J. C.; ALUÍSIO, S. M. "O Portal de Corpora Paralelos de Simplificação do PorSimples: visualização das buscas e a anotação XCES". Relatório Técnico, NILC – ICMC, USP, São Carlos, 2008, 27p.
- [MAX06] MAX, A. "Writing for Language-impaired Readers". Em: 7th International Conference on Intelligent Text Processing and Computational Linguistics, 2006, pp. 567-570.
- [PEI06] PEIXOTO, R. C. "Algumas considerações sobre a interface entre a Língua Brasileira de Sinais (LIBRAS) e a língua portuguesa na construção inicial da escrita pela criança surda". Campinas: Cadernos CEDES, 2006.
- [PER08] PEREIRA, T. F.; ALUISIO, S. M. "Editor de Anotação de Simplificação: Construção". Relatório Técnico, NILC – ICMC, USP, São Carlos, 2008, 27p.
- [PET07] PETERSEN, S. E.; OSTENDORF, M. "Text Simplification for Language Learners: A Corpus Analysis". Em: Speech and Language Technology for Education Workshop, 2007, 4p.
- [QUA97] QUADROS, R. M. "Educação de surdos: a aquisição da linguagem". Porto Alegre: Artes Médicas, 1997, 126p.
- [QUA04] QUADROS, R. M.; KARNOPP, L. B. "Língua de Sinais Brasileira: Estudos Lingüísticos". Porto Alegre: Artmed, 2004, 222 p.
- [QUA06] QUADROS, R. M.; SCHMIEDT, M. L. P. "Idéias para ensinar português para alunos surdos". Brasília: MEC/SEESP, 2006, 120p.
- [SBC06] SBC. "Grandes Desafios da Pesquisa em Computação no Brasil: 2006- 2016". São Paulo: SBC, 2006, 25p.
- [SID02] SIDDHARTHAN, A. "An Architecture for a Text Simplification System". Em: Language Engineering Conference, 2002, pp. 64-71.

- [SID06] SIDDHARTHAN, A. "Syntactic Simplification and Text Cohesion". Research on Language and Computation, vol. 4, Junho 2006, pp. 77-109.
- [SIG08] SIGNWRITING. "Sutton's Sign Writing Site". Capturado em: <http://www.signwriting.org>, Junho 2008.
- [SOU06] SOUZA, V. C.; VIEIRA, R. "Uma Proposta para Tradução Automática entre Libras e Português no SignWebMessage". Em: 4th Workshop in Information and Human Language Technology, 2006, 10p.
- [SPE08] SPECIA, L.; ALUISIO, S. M.; PARDO, T. A. S. "Manual de Simplificação Sintática para o Português". Relatório Técnico, NILC – ICMC, USP, São Carlos, 2008, 31p.
- [TAV05] TAVARES, O. L.; CORADINE, L. C.; BREDA, W. L. "Falibras-MT Autoria de tradutores automáticos de textos do português para LIBRAS, na forma gestual animada: Uma abordagem com memória de tradução". Em: 25th Congresso da Sociedade Brasileira de Computação, 2005, pp. 2099-2107.
- [THO07] THOMAS, D. et al. "Desenvolvimento Web ágil com Rails". Dallas: Pragmatic Bookshelf, 2007, 719p.

APÊNDICE – EXEMPLO DO FORMATO XCES GERADO NO SISTEMA

Texto bruto da simplificação forte.

Projeto propõe distribuição de uniforme à rede pública

O boné é um dos itens da roupa imaginada pela proposta.

Um projeto em discussão no Congresso criou uma polêmica em escolas, prefeituras e secretarias de educação. O projeto prevê a distribuição gratuita de uniforme para alunos da rede pública.

A proposta cria uma discussão sobre a inclusão do boné no uniforme obrigatório. Além disso, a proposta reanima o debate sobre a obrigação de uma roupa igual para todos os alunos. A proposta também renova a preocupação com a falta de recursos públicos. O projeto ainda precisa passar pela Câmara dos Deputados e pela Presidência da República antes de entrar em vigor.

O boné problemático entrou no projeto graças a uma emenda aprovada na Comissão de Educação do Senado em outubro. A emenda é do senador Cícero Lucena. Cícero Lucena é do PSDB-PB. Foi o senador Flávio Arns que sugeriu a inclusão do boné no uniforme de alunos dos ensinos Fundamental e Médio nas escolas municipais, estaduais e federais. Flávio Arns é do PT-PR. Ele defende a medida como forma de proteger crianças e adolescentes dos males provocados pelo excesso de exposição aos raios solares. Se a idéia for aprovada, os alunos receberão dois conjuntos anuais, completados com calçado, meias, calça e camiseta.

- A gurizada precisa se habituar a se proteger do sol. O boné será comprado em quantidade. Com isso, o boné custará R\$ 1 ou R\$ 2. É muito mais barato do que adotar outra medida para evitar os males causados pelo sol - avalia Arns.

O desentendimento ocorre porque educadores como Helvécio Aguiar afirmam que o boné não é necessário no uniforme obrigatório. Helvécio Aguiar é da Faculdade de Educação da Universidade Federal do Rio Grande do Sul. Professores temem também que a distribuição de boné torne difícil impor nas classes uma regra da boa maneira de se comportar: retirar chapéus e bonés da cabeça em ambientes fechados.

A preocupação da Federação das Associações de Municípios do RS (Famurs) e da Secretaria Estadual de Educação (SEC) é outra. A preocupação é com o dinheiro para pagar os gastos de milhões de reais com os bonés. O projeto prevê a criação do Programa Nacional de Uniforme Escolar para que o governo federal ajude Estados e municípios. Mas o projeto mantém as dúvidas se o dinheiro será suficiente.

- Não podemos acrescentar nenhuma despesa no nosso orçamento. Já não temos dinheiro suficiente para manter as escolas. Além disso, precisamos valorizar os professores - justifica a diretora do Departamento Pedagógico da SEC. A diretora é Sonia Balzano.

Se não fosse o problema de dinheiro, o projeto teria o apoio total de Balzano e do presidente da Famurs. Eles repetem o argumento que justifica a proposta: se o uniforme for oferecido, as escolas públicas poderão torná-lo obrigatório. Com isso, a roupa deixaria de ser um indicador

de diferenças sociais nas escolas e não criaria vergonha nos alunos mais pobres. A principal resistência vem dos próprios alunos: alguns criticam o uniforme. Alguns acreditam que o uniforme ameaça a individualidade de cada um.

Cabeçalho da versão de Simplificação Forte

```

UTF-8"?>
<cesHeaderxmlns="http://www.xces.org/schema/2003"xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.xces.org/schema/2003" version="1.0.4">
<fileDesc>
<titleStmt>
<title>O11_01112007</title>
<respStmt>
<respType>Criação do Header</respType>
<respName type="person">Felipe Vianna Perez</respName>
</respStmt>
</titleStmt>
<extent>
<wordCount>547</wordCount>
</extent>
<sourceDesc>
<biblStruct>
<monogr>
<title>Projeto propedistribuio de uniforme rede pblica</title>
<author>Mônica Martins</author>
<respStmt>
<respType>crédito</respType>
<respNametype="institution">DA REDAÇÃO</respName>
</respStmt>
<imprint>
<pubPlace>Jornal Zero Hora</pubPlace>
<publisher type="org">Grupo RBS</publisher>
<pubDate>01112007</pubDate>
<pubAddress>Av. Ipiranga, 1075 - Porto Alegre, RS</pubAddress>
</imprint>
</monogr>
</biblStruct>
</sourceDesc>
</fileDesc>
<encodingDesc>
<projectDesc>O projeto PorSimples (Simplificação Textual do Português para Inclusão e
Acessibilidade Digital) propõe o desenvolvimento de uma tecnologia para facilitar o acesso à
informação dos analfabetos funcionais (AF) e, potencialmente, de pessoas com outras
deficiências cognitivas, como afasia e dislexia. Ela será disponibilizada em dois sistemas
destinados a públicos alvos diferentes: um sistema de autoria para ajudar autores a produzirem
textos simplificados destinados aos AFs, textos estes que serão validados pelos autores, e um
sistema facilitador para ajudar AFs a lerem um dado conteúdo da Web. Este último inclui
tarefas de sumarização textual, simplificação, e apresentação do texto salientando as relações

```

entre as idéias do texto.

Quanto aos cenários de uso, esta tecnologia proposta inclui a tarefa de simplificar textos principalmente para assistir AFs na leitura de textos eletrônicos produzidos, por exemplo, pelo governo ou por jornais de grande circulação, promovendo assim a inclusão e acessibilidade digital.

O foco é em AFs pois, segundo a Síntese dos Indicadores Sociais de 2006 do IBGE, o número de pessoas nestas condições alcançava 23,5%, em 2005. Além disso, pode apoiar crianças em fase de aprendizado de leitura ou ainda adultos em fase de alfabetização. A língua dos textos é o português, para a qual, no melhor do nosso conhecimento, não há sistemas de simplificação textual disponíveis.

Esse projeto teve início em Novembro de 2007 e tem duração prevista de 2 anos. Conta com o apoio da FAPESP (Fundação de Amparo à Pesquisa de São Paulo) e da MSR (Microsoft Research). </projectDesc>

<samplingDecl>O corpus paralelo de simplificações do PorSimples é formado por textos dos anos 2006 e 2007 do jornal Zero Hora que possuem versões correspondentes simplificadas também por jornalistas da Seção Para Seu Filho Ler. Ele é formado por 100 pares de textos (original-natural) e mais 100 pares (natural - simplificação forte). A simplificação foi criada manualmente por Mônica Martins, apoiada pelo Editor de Anotação de Simplificação do PorSimples.</samplingDecl>

</encodingDesc>

<profileDesc>

<annotations>

<annotation type="content" ann.loc="O11_01112007_strong.txt">Document content</annotation>

<annotation type="logical" ann.loc="O11_01112007_strong-logical.xml">Logical markup</annotation>

<annotation type="s" ann.loc="O11_01112007_strong-s.xml">Sentence boundaries</annotation>

<annotation type="tokens" ann.loc="O11_01112007_strong-token.xml">Tokens</annotation>

<annotation type="pos" ann.loc="O11_01112007_strong-pos.xml">Pos</annotation>

<annotation type="alignment" ann.loc="O11_01112007_strong-align.xml">Alignment between original and simplified sentences</annotation>

<annotation type="simplification" ann.loc="O11_01112007_strong-simplify.xml">Simplification operations applied to the original sentences</annotation>

</annotations>

</profileDesc>

</cesHeader>

Divisão lógica da versão de Simplificação Forte

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<cesAna version="1.0.4" xmlns="http://www.xces.org/schema/2003">
```

```
<struct type="cesDoc" from="0" to="3065">
```

```
<feat name="version" value="1.0.4"/>
```

```
<feat name="id" value="O11_01112007"/>
```

```
<feat name="xmlns:xsi" value="http://www.w3.org/2001/XMLSchema-instance"/>
```

```
<feat name="xmlns:xlink" value="http://www.w3.org/1999/xlink"/>
```

```
<feat name="xmlns" value="http://www.xces.org/schema/2003"/>
```

```

</struct>
<struct type="text" from="0" to="3064"/>
<struct type="body" from="1" to="3063"/>
<struct type="div" from="2" to="3062">
  <feat name="type" value="materia"/>
</struct>
<struct type="p" from="0" to="56">
  <feat name="id" value="p1"/>
</struct>
<struct type="p" from="57" to="247">
  <feat name="id" value="p2"/>
</struct>
<struct type="p" from="248" to="619">
  <feat name="id" value="p3"/>
</struct>
<struct type="p" from="620" to="1242">
  <feat name="id" value="p4"/>
</struct>
<struct type="p" from="1243" to="1480">
  <feat name="id" value="p5"/>
</struct>
<struct type="p" from="1481" to="1881">
  <feat name="id" value="p6"/>
</struct>
<struct type="p" from="1882" to="2284">
  <feat name="id" value="p7"/>
</struct>
<struct type="p" from="2285" to="2544">
  <feat name="id" value="p8"/>
</struct>
<struct type="p" from="2545" to="3065">
  <feat name="id" value="p9"/>
</struct>
</cesAna>

```

Divisão lógica da versão de Simplificação Forte

```

<?xml version="1.0" encoding="UTF-8"?>
<cesAna version="1.0.4" xmlns="http://www.xces.org/schema/2003">
<struct type="cesDoc" from="0" to="3065">
  <feat name="version" value="1.0.4"/>
  <feat name="id" value="O11_01112007"/>
  <feat name="xmlns:xsi" value="http://www.w3.org/2001/XMLSchema-instance"/>
  <feat name="xmlns:xlink" value="http://www.w3.org/1999/xlink"/>
  <feat name="xmlns" value="http://www.xces.org/schema/2003"/>
</struct>
<struct type="text" from="0" to="3064"/>
<struct type="body" from="1" to="3063"/>
<struct type="div" from="2" to="3062">

```

```

    <feat name="type" value="materia"/>
</struct>
<struct type="p" from="0" to="56">
    <feat name="id" value="p1"/>
</struct>
<struct type="p" from="57" to="247">
    <feat name="id" value="p2"/>
</struct>
<struct type="p" from="248" to="619">
    <feat name="id" value="p3"/>
</struct>
<struct type="p" from="620" to="1242">
    <feat name="id" value="p4"/>
</struct>
<struct type="p" from="1243" to="1480">
    <feat name="id" value="p5"/>
</struct>
<struct type="p" from="1481" to="1881">
    <feat name="id" value="p6"/>
</struct>
<struct type="p" from="1882" to="2284">
    <feat name="id" value="p7"/>
</struct>
<struct type="p" from="2285" to="2544">
    <feat name="id" value="p8"/>
</struct>
<struct type="p" from="2545" to="3065">
    <feat name="id" value="p9"/>
</struct>
</cesAna>

```

Divisão de sentenças da versão de Simplificação Forte

```

<?xml version="1.0" encoding="UTF-8"?>
<cesAna version="1.0.4" xmlns="http://www.xces.org/schema/2003">
<struct type="s" from="0" to="56">
    <feat name="id" value="p1s1"/>
</struct>
<struct type="s" from="57" to="165">
    <feat name="id" value="p2s1"/>
</struct>
<struct type="s" from="166" to="247">
    <feat name="id" value="p2s2"/>
</struct>
<struct type="s" from="248" to="328">
    <feat name="id" value="p3s1"/>
</struct>
<struct type="s" from="329" to="429">
    <feat name="id" value="p3s2"/>

```

```
</struct>
<struct type="s" from="430" to="503">
  <feat name="id" value="p3s3"/>
</struct>
<struct type="s" from="504" to="619">
  <feat name="id" value="p3s4"/>
</struct>
<struct type="s" from="620" to="733">
  <feat name="id" value="p4s1"/>
</struct>
<struct type="s" from="734" to="771">
  <feat name="id" value="p4s2"/>
</struct>
<struct type="s" from="772" to="800">
  <feat name="id" value="p4s3"/>
</struct>
<struct type="s" from="801" to="959">
  <feat name="id" value="p4s4"/>
</struct>
<struct type="s" from="960" to="984">
  <feat name="id" value="p4s5"/>
</struct>
<struct type="s" from="985" to="1119">
  <feat name="id" value="p4s6"/>
</struct>
<struct type="s" from="1120" to="1242">
  <feat name="id" value="p4s7"/>
</struct>
<struct type="s" from="1243" to="1299">
  <feat name="id" value="p5s1"/>
</struct>
<struct type="s" from="1300" to="1336">
  <feat name="id" value="p5s2"/>
</struct>
<struct type="s" from="1337" to="1377">
  <feat name="id" value="p5s3"/>
</struct>
<struct type="s" from="1378" to="1480">
  <feat name="id" value="p5s4"/>
</struct>
<struct type="s" from="1481" to="1606">
  <feat name="id" value="p6s1"/>
</struct>
<struct type="s" from="1607" to="1696">
  <feat name="id" value="p6s2"/>
</struct>
<struct type="s" from="1697" to="1823">
  <feat name="id" value="p6s3"/>
</struct>
```

```
<struct type="s" from="1824" to="1881">
  <feat name="id" value="p6s4"/>
</struct>
<struct type="s" from="1882" to="2010">
  <feat name="id" value="p7s1"/>
</struct>
<struct type="s" from="2011" to="2097">
  <feat name="id" value="p7s2"/>
</struct>
<struct type="s" from="2098" to="2220">
  <feat name="id" value="p7s3"/>
</struct>
<struct type="s" from="2221" to="2284">
  <feat name="id" value="p7s4"/>
</struct>
<struct type="s" from="2285" to="2348">
  <feat name="id" value="p8s1"/>
</struct>
<struct type="s" from="2349" to="2406">
  <feat name="id" value="p8s2"/>
</struct>
<struct type="s" from="2407" to="2515">
  <feat name="id" value="p8s3"/>
</struct>
<struct type="s" from="2516" to="2544">
  <feat name="id" value="p8s4"/>
</struct>
<struct type="s" from="2545" to="2651">
  <feat name="id" value="p9s1"/>
</struct>
<struct type="s" from="2652" to="2703">
  <feat name="id" value="p9s2"/>
</struct>
<struct type="s" from="2704" to="2786">
  <feat name="id" value="p9s3"/>
</struct>
<struct type="s" from="2787" to="2916">
  <feat name="id" value="p9s4"/>
</struct>
<struct type="s" from="2917" to="2966">
  <feat name="id" value="p9s5"/>
</struct>
<struct type="s" from="2967" to="2995">
  <feat name="id" value="p9s6"/>
</struct>
<struct type="s" from="2996" to="3065">
  <feat name="id" value="p9s7"/>
</struct>
</cesAna>
```

Divisão de sentenças da versão de Simplificação Forte

```

<?xml version="1.0" standalone="yes"?>
<cesAna version="1.0.4" xmlns="http://www.xces.org/schema/2003">
<struct type='token' from='0' to='1'>
  <feat name='id' value='t1'/>
  <feat name='base' value='O'/>
</struct>
<struct type='token' from='2' to='6'>
  <feat name='id' value='t2'/>
  <feat name='base' value='bon'/>
</struct>
<struct type='token' from='7' to='8'>
  <feat name='id' value='t3'/>
  <feat name='base' value=''/>
</struct>
<struct type='token' from='9' to='11'>
  <feat name='id' value='t4'/>
  <feat name='base' value='um'/>
</struct>
<struct type='token' from='12' to='15'>
  <feat name='id' value='t5'/>
  <feat name='base' value='de'/>
</struct>
<struct type='token' from='12' to='15'>
  <feat name='id' value='t6'/>
  <feat name='base' value='o'/>
</struct>
<struct type='token' from='16' to='21'>
  <feat name='id' value='t7'/>
  <feat name='base' value='itens'/>
</struct>
<struct type='token' from='22' to='24'>
  <feat name='id' value='t8'/>
  <feat name='base' value='de'/>
</struct>
<struct type='token' from='22' to='24'>
  <feat name='id' value='t9'/>
  <feat name='base' value='o'/>
</struct>
<struct type='token' from='25' to='30'>
  <feat name='id' value='t10'/>
  <feat name='base' value='roupa'/>
</struct>
<struct type='token' from='31' to='40'>
  <feat name='id' value='t11'/>
  <feat name='base' value='imaginada'/>
</struct>
<struct type='token' from='41' to='45'>

```

```

    <feat name='id' value='t12'/>
    <feat name='base' value='por'/>
</struct>
<struct type='token' from='41' to='45'>
    <feat name='id' value='t13'/>
    <feat name='base' value='o'/>
</struct>
<struct type='token' from='46' to='54'>
    <feat name='id' value='t14'/>
    <feat name='base' value='proposta'/>
</struct>
<...>

<struct type='token' from='2545' to='2547'>
    <feat name='id' value='t449'/>
    <feat name='base' value='Se'/>
</struct>
<struct type='token' from='2548' to='2551'>
    <feat name='id' value='t450'/>
    <feat name='base' value='no'/>
</struct>
<struct type='token' from='2552' to='2557'>
    <feat name='id' value='t451'/>
    <feat name='base' value='fosse'/>
</struct>
<struct type='token' from='2558' to='2559'>
    <feat name='id' value='t452'/>
    <feat name='base' value='o'/>
</struct>
<struct type='token' from='2560' to='2568'>
    <feat name='id' value='t453'/>
    <feat name='base' value='problema'/>
</struct>
<struct type='token' from='2569' to='2571'>
    <feat name='id' value='t454'/>
    <feat name='base' value='de'/>
</struct>
<struct type='token' from='2572' to='2580'>
    <feat name='id' value='t455'/>
    <feat name='base' value='dinhaeiro'/>
</struct>
<struct type='token' from='2581' to='2582'>
    <feat name='id' value='t456'/>
    <feat name='base' value=','/>
</struct>
<struct type='token' from='2583' to='2584'>
    <feat name='id' value='t457'/>
    <feat name='base' value='o'/>

```

```

</struct>
<struct type='token' from='2585' to='2592'>
  <feat name='id' value='t458'/>
  <feat name='base' value='projeto'/>
</struct>
<struct type='token' from='2593' to='2598'>
  <feat name='id' value='t459'/>
  <feat name='base' value='teria'/>
</struct>
<struct type='token' from='2599' to='2600'>
  <feat name='id' value='t460'/>
  <feat name='base' value='o'/>
</struct>
<struct type='token' from='2601' to='2606'>
  <feat name='id' value='t461'/>
  <feat name='base' value='apoiio'/>
</struct>
<struct type='token' from='2607' to='2612'>
  <feat name='id' value='t462'/>
  <feat name='base' value='total'/>
</struct>
<struct type='token' from='2613' to='2615'>
  <feat name='id' value='t463'/>
  <feat name='base' value='de'/>
</struct>
<struct type='token' from='2616' to='2623'>
  <feat name='id' value='t464'/>
  <feat name='base' value='Balzano'/>
</struct>
<struct type='token' from='2624' to='2625'>
  <feat name='id' value='t465'/>
  <feat name='base' value='e'/>
</struct>
<struct type='token' from='2626' to='2628'>
  <feat name='id' value='t466'/>
  <feat name='base' value='de'/>
</struct>
<struct type='token' from='2626' to='2628'>
  <feat name='id' value='t467'/>
  <feat name='base' value='o'/>
</struct>
<struct type='token' from='2629' to='2639'>
  <feat name='id' value='t468'/>
  <feat name='base' value='presidente'/>
</struct>
<struct type='token' from='2640' to='2642'>
  <feat name='id' value='t469'/>
  <feat name='base' value='de'/>
</struct>

```

```

<struct type='token' from='2640' to='2642'>
  <feat name='id' value='t470'/>
  <feat name='base' value='o'/>
</struct>
<struct type='token' from='2643' to='2649'>
  <feat name='id' value='t471'/>
  <feat name='base' value='Famurs'/>
</struct>
<struct type='token' from='2650' to='2651'>
  <feat name='id' value='t472'/>
  <feat name='base' value='.'/>
</struct>

<...>

<struct type='token' from='2917' to='2918'>
  <feat name='id' value='t522'/>
  <feat name='base' value='A'/>
</struct>
<struct type='token' from='2919' to='2928'>
  <feat name='id' value='t523'/>
  <feat name='base' value='principal'/>
</struct>
<struct type='token' from='2929' to='2940'>
  <feat name='id' value='t524'/>
  <feat name='base' value='resistncia'/>
</struct>
<struct type='token' from='2941' to='2944'>
  <feat name='id' value='t525'/>
  <feat name='base' value='vem'/>
</struct>
<struct type='token' from='2945' to='2948'>
  <feat name='id' value='t526'/>
  <feat name='base' value='de'/>
</struct>
<struct type='token' from='2945' to='2948'>
  <feat name='id' value='t527'/>
  <feat name='base' value='o'/>
</struct>
<struct type='token' from='2949' to='2957'>
  <feat name='id' value='t528'/>
  <feat name='base' value='prprios'/>
</struct>
<struct type='token' from='2958' to='2964'>
  <feat name='id' value='t529'/>
  <feat name='base' value='alunos'/>
</struct>
<struct type='token' from='2965' to='2966'>
  <feat name='id' value='t530'/>

```

```

    <feat name='base' value='.'/>
</struct>
<struct type='token' from='2967' to='2973'>
    <feat name='id' value='t531'/>
    <feat name='base' value='Alguns'/>
</struct>
<struct type='token' from='2974' to='2982'>
    <feat name='id' value='t532'/>
    <feat name='base' value='criticam'/>
</struct>
<struct type='token' from='2983' to='2984'>
    <feat name='id' value='t533'/>
    <feat name='base' value='o'/>
</struct>
<struct type='token' from='2985' to='2993'>
    <feat name='id' value='t534'/>
    <feat name='base' value='uniforme'/>
</struct>
<struct type='token' from='2994' to='2995'>
    <feat name='id' value='t535'/>
    <feat name='base' value='.'/>
</struct>
<struct type='token' from='2996' to='3002'>
    <feat name='id' value='t536'/>
    <feat name='base' value='Alguns'/>
</struct>
<struct type='token' from='3003' to='3012'>
    <feat name='id' value='t537'/>
    <feat name='base' value='acreditam'/>
</struct>
<struct type='token' from='3013' to='3016'>
    <feat name='id' value='t538'/>
    <feat name='base' value='que'/>
</struct>
<struct type='token' from='3017' to='3018'>
    <feat name='id' value='t539'/>
    <feat name='base' value='o'/>
</struct>
<struct type='token' from='3019' to='3027'>
    <feat name='id' value='t540'/>
    <feat name='base' value='uniforme'/>
</struct>
<struct type='token' from='3028' to='3034'>
    <feat name='id' value='t541'/>
    <feat name='base' value='ameaa'/>
</struct>
<struct type='token' from='3035' to='3036'>
    <feat name='id' value='t542'/>
    <feat name='base' value='a'/>

```

```

</struct>
<struct type='token' from='3037' to='3052'>
  <feat name='id' value='t543' />
  <feat name='base' value='individualidade' />
</struct>
<struct type='token' from='3053' to='3055'>
  <feat name='id' value='t544' />
  <feat name='base' value='de' />
</struct>
<struct type='token' from='3056' to='3063'>
  <feat name='id' value='t545' />
  <feat name='base' value='cada=um' />
</struct>
<struct type='token' from='3064' to='3065'>
  <feat name='id' value='t546' />
  <feat name='base' value='.' />
</struct>
</cesAna>

```

Operações Realizadas para Reescrita em LIBRAS

```

<?xml version="1.0" standalone="yes"?>
<cesAna version="1.0.4" xmlns="http://www.xces.org/schema/2003">
  <struct type="opr">
    <feat name="id" value="opr1" />
    <feat name="type" value="LIBRAS_rewrite" />
    <feat name="sentenceref" value="p1s1" />
  </struct>
  <struct type="opr">
    <feat name="id" value="opr2" />
    <feat name="type" value="LIBRAS_rewrite" />
    <feat name="sentenceref" value="p2s1" />
  </struct>
  <struct type="opr">
    <feat name="id" value="opr3" />
    <feat name="type" value="LIBRAS_rewrite" />
    <feat name="sentenceref" value="p2s2" />
  </struct>
  <struct type="opr">
    <feat name="id" value="opr4" />
    <feat name="type" value="LIBRAS_rewrite" />
    <feat name="sentenceref" value="p3s1" />
  </struct>
  <struct type="opr">
    <feat name="id" value="opr5" />
    <feat name="type" value="LIBRAS_rewrite" />
    <feat name="sentenceref" value="p3s2" />
  </struct>
  <struct type="opr">

```

```

    <feat name="id" value="opr6"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p3s3"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr7"/>
    <feat name="type" value="join_sentence"/>
    <feat name="sentenceref" value="p3s3"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr8"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p3s4"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr9"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p4s1"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr10"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p4s2"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr11"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p4s4"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr12"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p4s5"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr13"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p4s6"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr14"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p4s7"/>
</struct>
<struct type="opr">
    <feat name="id" value="opr15"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p5s1"/>
</struct>

```

```

<struct type="opr">
  <feat name="id" value="opr16"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p5s2"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr17"/>
  <feat name="type" value="join_sentence"/>
  <feat name="sentenceref" value="p5s3"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr18"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p5s3"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr19"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p5s4"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr20"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p6s1"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr21"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p6s2"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr22"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p6s3"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr23"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p6s4"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr24"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p7s1"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr25"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p7s2"/>

```

```

</struct>
<struct type="opr">
  <feat name="id" value="opr26"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p7s3"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr27"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p7s4"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr28"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p8s1"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr29"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p8s2"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr30"/>
  <feat name="type" value="speech_mark"/>
  <feat name="sentenceref" value="p8s3"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr31"/>
  <feat name="type" value="split_sentence"/>
  <feat name="sentenceref" value="p8s3"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr32"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p8s3"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr33"/>
  <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p8s4"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr34"/>
  <feat name="type" value="join_sentence"/>
  <feat name="sentenceref" value="p8s4"/>
</struct>
<struct type="opr">
  <feat name="id" value="opr35"/>
  <feat name="type" value="LIBRAS_rewrite"/>

```

```

    <feat name="sentenceref" value="p9s1"/>
  </struct>
  <struct type="opr">
    <feat name="id" value="opr36"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p9s2"/>
  </struct>
  <struct type="opr">
    <feat name="id" value="opr37"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p9s3"/>
  </struct>
  <struct type="opr">
    <feat name="id" value="opr38"/>
    <feat name="type" value="join_sentence"/>
    <feat name="sentenceref" value="p9s3"/>
  </struct>
  <struct type="opr">
    <feat name="id" value="opr39"/>
    <feat name="type" value="LIBRAS_rewrite"/>
    <feat name="sentenceref" value="p9s4"/>
  </struct>
  <struct type="opr">
    <feat name="id" value="opr40"/>
    <feat name="type" value="split_sentence"/>
    <feat name="sentenceref" value="p9s4"/>
  </struct>
  <struct type="opr">
    <feat name="id" value="opr41"/>
    <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p9s5"/>
</struct>
  <struct type="opr">
    <feat name="id" value="opr42"/>
    <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p9s6"/>
</struct>
  <struct type="opr">
    <feat name="id" value="opr43"/>
    <feat name="type" value="LIBRAS_rewrite"/>
  <feat name="sentenceref" value="p9s7"/>
</struct>
  <struct type="opr">
    <feat name="id" value="opr44"/>
    <feat name="type" value="join_sentence"/>
    <feat name="sentenceref" value="p9s7"/>
  </struct>
</cesAna>

```

Alinhamento entre as sentenças

```

<?xml version="1.0" encoding="UTF-8"?>
<cesAlign version="1.0.4" type="sent" fromDoc="O11_01112007_strong-s.xml"
toDoc="O11_01112007_strong-s.xml" xmlns="http://www.xces.org/schema/2003"
xmlns:xlink="http://www.x3.org/1999/xlink">
<cesHeader version="1.0.4">
<fileDesc>
<titleStmt>
<title>Alignment of original and natural simplified versions of O11_01112007_strong.txt</title>
</titleStmt>
<sourceDesc>
<biblStruct>
<monogr>
<title>Projeto propedistribuio de uniformere de pblica</title>
<author>Mônica Martins</author>
</monogr>
</biblStruct>
</sourceDesc>
</fileDesc>
<profileDesc>
<translations>
<translation wsd="utf-8" trans.loc="O11_01112007_strong-s.xml"/>
<translation wsd="utf-8" trans.loc="O11_01112007_strong-s.xml"/>
</translations>
</profileDesc>
</cesHeader>
<linkList>
<linkGrp id="p1">
<link>
<align xlink:href="#p1s1"/>
<align xlink:href="#p1s1"/>
</link>
</linkGrp>
<linkGrp id="p2">
<link>
<align xlink:href="#p2s1"/>
<align xlink:href="#p2s1"/>
</link>
<link>
<align xlink:href="#p2s2"/>
<align xlink:href="#p2s2"/>
</link>
</linkGrp>
<linkGrp id="p3">
<link>
<align xlink:href="#p3s1"/>
<align xlink:href="#p3s1"/>
</link>
<link>

```

```

                <align xlink:href="#xpointer(id('p3s2')/range-to(id('p3s3')))" />
                <align xlink:href="#xpointer(id('p3s2')/range-to(id('p3s3')))" />
        </link>
        <link>
                <align xlink:href="#p3s4" />
                <align xlink:href="#p4s1" />
        </link>
</linkGrp>
<linkGrp id="p4">
        <link>
                <align xlink:href="#p4s1" />
                <align xlink:href="#p4s2" />
        </link>
        <link>
                <align xlink:href="#p4s2" />
                <align xlink:href="#p4s3" />
        </link>
<link>
        <align xlink:href="#p4s3" />
        <align xlink:href="#p4s4" />
</link>
<link>
                <align xlink:href="#p4s4" />
                <align xlink:href="#p4s5" />
</link>
<link>
        <align xlink:href="#p4s5" />
<align xlink:href="#p4s6" />
</link>
<link>
                <align xlink:href="#p4s6" />
                <align xlink:href="#p5s1" />
</link>
<link>
                <align xlink:href="#p4s7" />
                <align xlink:href="#p5s2" />
        </link>
</linkGrp>
<linkGrp id="p5">
        <link>
                <align xlink:href="#p5s1" />
                <align xlink:href="#p5s3" />
        </link>
        <link>
                <align xlink:href="#xpointer(id('p5s2')/range-to(id('p5s3')))" />
                <align xlink:href="#xpointer(id('p6s1')/range-to(id('p6s2')))" />
        </link>
        <link>
                <align xlink:href="#p5s4" />

```

```

    <align xlink:href="#p6s3"/>
  </link>
</linkGrp>
<linkGrp id="p6">
  <link>
    <align xlink:href="#p6s1"/>
    <align xlink:href="#p6s4"/>
    </link>
  <link>
    <align xlink:href="#p6s2"/>
    <align xlink:href="#p7s1"/>
  </link>
<link>
<align xlink:href="#p6s3"/>
<align xlink:href="#p7s2"/>
  </link>
<link>
    <align xlink:href="#p6s4"/>
    <align xlink:href="#p7s3"/>
  </link>
</linkGrp>
<linkGrp id="p7">
  <link>
    <align xlink:href="#p7s1"/>
    <align xlink:href="#p7s4"/>
  </link>
<link>
    <align xlink:href="#p7s2"/>
    <align xlink:href="#p8s1"/>
  </link>
<link>
    <align xlink:href="#p7s3"/>
    <align xlink:href="#p8s2"/>
  </link>
<link>
    <align xlink:href="#p7s4"/>
    <align xlink:href="#p8s3"/>
  </link>
</linkGrp>
<linkGrp id="p8">
<link>
    <align xlink:href="#p8s1"/>
    <align xlink:href="#p8s4"/>
  </link>
<link>
<align xlink:href="#p8s2"/>
<align xlink:href="#p9s1"/>
</link>
<link>

```

```

<align xlink:href="#xpointer(id('p8s3')/range-to(id('p8s4')))/>
<align xlink:href="#xpointer(id('p9s3')/range-to(id('p9s4')))/>
</link>
</linkGrp>
<linkGrp id="p9">
<link>
    <align xlink:href="#p9s1"/>
<align xlink:href="#p9s5"/>
</link>
</link>
    <align xlink:href="#xpointer(id('p9s2')/range-to(id('p9s3')))/>
    <align xlink:href="#p9s6"/>
</link>
</linkGrp>
</linkList>
</cesAlign>

```

Texto bruto da reescrita em LIBRAS

Projeto propõe distribuição de uniforme à rede pública

Proposta distribuir boné junto roupa.

Congresso discussão projeto muita polêmica escolas prefeituras secretarias educação. Projeto distribuir grátis roupa alunos rede pública.

Proposta discussão junto boné roupa obrigatório sim ou não? Também de novo discutir roupa igual todos alunos obrigatório dinheiro publico tem? Falta. Primeiro câmara deputados também presidente ler depois assina oficial.

Comissão educação senado colocou emenda boné obrigatório mês outubro passado. Emenda pessoa responsável senador CICERO LUCENA partido PSDB PB. Outro senador FLAVIO ARNS idéia boné junto roupa alunos ensino fundamental médio escolas municipais estaduais federais. Ele partido PT PR. Ele defesa idéia porque proteção crianças e adolescentes forte sol todos os dias. Se aprovação dar dois por ano calçado meias calça camiseta completos.

Aluno precisa adaptar proteção Sol. Bonés comprar muitos por isso pagar R\$ 1 ou 2 cada. Mais barato do que outra coisa evitar sol opinião ARNS.

Discussão tem porque professor nome HELVÉCIO AGUIAR opinião não precisa boné junto roupa obrigatório. Professor trabalha faculdade educação UFRGS. Professores medo dar boné depois difícil regras comportamento certo. Chapéu boné tirar lugar fechado.

FEDERAÇÃO ASSOCIAÇÃO MUNICIPIOS RS (FAMURS) também SECRETARIA ESTADUAL EDUCAÇÃO preocupação outra. É dinheiro pagar milhões reais comprar bonés. Projeto quer criação PROGRAMA NACIONAL UNIFORME ESCOLAR governo federal ajuda estados municípios. Mas projeto continua duvida dinheiro consegue?

Não pode despesa colocar orçamento. Dinheiro não tem escolas manter. Também precisa

valor professores. Opinião diretora DEPARTAMENTO PEDAGÓGICO SEE nome SONIA BALZANO.

Se problema dinheiro não tem apoio projeto BALZANO também presidente FAMURS. De novo argumento explica proposta se uniforme dar gratuito escolas públicas pode resolver roupa obrigatório. Roupa não mostra diferença social dentro escola. Alunos pobres não ter vergonha. Alunos principal resistência. Alguns criticar roupa igual porque ameaçar identidade cada um.