



**UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE TUCURUÍ
ENGENHARIA DE COMPUTAÇÃO**

JEREMIAS KALEBE CUNHA ESTUMANO

**LAWS-FERRAMENTA WEB PARA ANOTAÇÕES MANUAIS EM ARQUIVOS
AUDIOVISUAIS**

**Tucuruí-PA
Dezembro de 2021**



**UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE TUCURUÍ
ENGENHARIA DE COMPUTAÇÃO**

JEREMIAS KALEBE CUNHA ESTUMANO

**LAWS-FERRAMENTA WEB PARA ANOTAÇÕES MANUAIS EM ARQUIVOS
AUDIOVISUAIS**

Trabalho de Conclusão de Curso apresentado
para obtenção do grau de Bacharel em Engenharia
de Computação.

Orientador: Prof. Dr. Marcos Túlio Amaris González

**Tucuruí-PA
Dezembro de 2021**

Estumano, Jeremias

LAWS-Ferramenta web para anotações manuais em arquivos audiovisuais/ Jeremias Kalebe Cunha Estumano. – Tucuruí-PA, Dezembro de 2021.

47 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Marcos Túlio Amaris González

Monografia – UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE TUCURUÍ
, Dezembro de 2021.

1. Anotações. 2. Elan. 3. Linguistas. I. Título.

JEREMIAS KALEBE CUNHA ESTUMANO

**LAWS-FERRAMENTA WEB PARA ANOTAÇÕES MANUAIS EM
ARQUIVOS AUDIOVISUAIS**

Trabalho de Conclusão de Curso apresentado
para obtenção do grau de Bacharel em Engenharia
de Computação.

Data da Defesa: 12 de Janeiro de 2022

Conceito: EXCELENTE

Banca Examinadora

Prof. Dr. Marcos Túlio Amaris González

Faculdade de Engenharia de Computação -
UFPA (CAMTUC)
Orientador

Prof. Dr. Heleno Fulber

Faculdade de Engenharia de Computação -
UFPA (CAMTUC)
Membro da Banca

**Profa. Dra. Antônia Fernanda de Souza
Nogueira**

Faculdade de Letras - UFPA (ILC, Belém)
Membro da Banca

Tucuruí-PA
Dezembro de 2021

AGRADECIMENTOS

A Deus, por ter cuidado de mim e por todas as bênçãos concebidas na minha vida.

A toda minha família, pelo amor e carinho a mim dedicados, em especial à minha mãe Claudete, meus irmãos Samuel, Ester e Talita e meu pai Josoel.

Ao orientador deste trabalho, professor Dr. Marcos Amaris, pelos ensinamentos, boa vontade, apoio na condução deste trabalho e por toda ajuda que me permitiu ao final terminar e apresentar este trabalho de conclusão de curso.

A professora Dra. Fernanda Nogueira pelo apoio e por toda ajuda na condução deste trabalho.

Aos demais professores pelas correções e ensinamentos que me permitiram apresentar um bom desempenho no meu processo de formação profissional.

Aos meus colegas de turma pelos momentos de alegria e superação vividos juntos.

E a todos que de certa maneira contribuíram com a minha jornada acadêmica.

RESUMO

Linguistas passam horas fazendo anotações em mídias audiovisuais. E a facilidade de acesso e manuseio das ferramentas é muito importante no desempenho do trabalho dos mesmos. Existe a ferramenta ELAN, muito usada atualmente, para anotações. Entretanto, a sua interface robusta e o alto grau de complexidade torna um desafio para os iniciantes. Assim, desenvolvemos uma nova ferramenta, que seja web, simples e com interface de fácil visualização. Apresentamos o Laws, ferramenta web, para facilitar o acesso, manuseio e a documentação em linguística. O LAWS pode ser acessado nesse link: <https://marcosamaris.github.io/laws/>.

Palavras-chave: ELAN. LingView. Anotações. Ferramenta Web.

ABSTRACT

Linguists spend hours annotating audiovisual media. And the ease of access and handling of the tools is very important in the performance of their work. There is the ELAN tool, which is widely used today, for annotations. However, its robust interface and high degree of complexity makes it a challenge for beginners. Thus, a new tool was needed, which is web, simple and with an easy-to-view interface. We present Laws, a web tool, to facilitate access, handling and documentation in linguistics. LAWS can be accessed from this link: <https://marcosamaris.github.io/laws/>.

Keywords: ELAN. LingView. Annotations. Web Tool.

LISTA DE ILUSTRAÇÕES

Figura 1 – Interface do ELAN	20
Figura 2 – Interface do LingView	21
Figura 3 – Pipeline da ferramenta LingView	22
Figura 4 – Estrutura do JSON	23
Figura 5 – Componente de representação de uma sentença	24
Figura 6 – Interface do ETST	25
Figura 7 – Apresentação da língua Kadiwéu na ferramenta Tycho Brahe	26
Figura 8 – Pipeline do modelo Interativo do LAWS	27
Figura 9 – Diagrama de Casos de uso da ferramenta Laws	30
Figura 10 – Diagrama de classes da ferramenta Laws	30
Figura 11 – Diagrama de Sequencia do UC Abrir Projeto	31
Figura 12 – Diagrama de Sequencia do UC Criar Projeto	32
Figura 13 – Diagrama de Sequencia do UC Inserir Anotação	32
Figura 14 – Diagrama de Sequencia do UC Deletar Anotação	33
Figura 15 – Diagrama de Sequencias do UC Editar Anotação	33
Figura 16 – Arquitetura do LAWS	35
Figura 17 – Interface do LAWS	38
Figura 18 – Teste do LAWS na língua da comunidade Wajuru	39
Figura 19 – Tela abrir projeto do LAWS	45
Figura 20 – Tela Novo projeto do LAWS	46
Figura 21 – Tela de edição	46
Figura 22 – Editar e Deletar	47

LISTA DE TABELAS

Tabela 1 – Identificação dos Atores	28
Tabela 2 – Identificação dos Casos de Usos	29
Tabela 3 – Identificação das regras de negócios da ferramenta	29

LISTA DE ABREVIATURAS E SIGLAS

abnTeX	ABsurdas Normas para TeX
CSS	Cascading Style Sheets
ELAN	Eunuco Language Annotation
DOM	Document Object Model
ETST	ELAN Text-Sync Tool
FLEx	FieldWorks Language Explorer
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
LAWS	Language Annotation Web system
XML	Extensible Markup Language

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Contexto	12
1.2	O Problema	12
1.3	Justificativa	15
1.4	Objetivos	16
1.5	Estrutura do Trabalho	16
2	REFERENCIAS TEÓRICOS E TRABALHOS RELACIONADOS . . .	17
2.1	Referencias Teóricas	17
2.1.1	Documentação Linguística	17
2.1.2	Conceitos sobre a Arquitetura de software do LAWS	18
2.2	Trabalhos relacionados	19
2.2.1	ELAN	20
2.2.2	LingView	21
2.2.3	ELAN Text-Sync Tool(ETST)	24
2.2.4	Tycho Brahe Platform	26
3	METODOLOGIA	27
3.1	Levantamento e Definição de requisitos	28
3.1.1	Abrir Projeto	31
3.1.2	Criar Projeto	31
3.1.3	Inserir Anotação	32
3.1.4	Deletar Anotação	32
3.1.5	Editar Anotação	33
3.2	Preparar o ambiente	33
3.2.1	Backend	33
3.2.2	Banco de dados	34
3.2.3	Adaptação na web	34
3.3	Acrescentar funcionalidades	34
3.3.1	Implementação	35
3.4	Testes e ajustes da Ferramenta	35
3.5	Hospedagem da ferramenta	36
4	RESULTADOS E USO DO LAWS	38
4.1	Experiências clássicas e uso de casos de LAWS com uma língua ameaçada, a língua da comunidade indígena Wajuru	39
5	CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS	41
5.1	Conclusão e Sugestões para trabalhos futuros	41
	REFERÊNCIAS	42

	APÊNDICES	44
	APÊNDICE A – MANUAL DE USO DO LAWS	45
A.1	Abrir Projeto	45
A.2	Novo Projeto	45
A.3	Inserir, Editar, Deletar e Exportar	46

1 INTRODUÇÃO

1.1 Contexto

Existe no mundo uma diversidade linguística incalculável. Segundo o consenso de vários linguistas (AUSTIN; SALLABANK, 2011), há aproximadamente 7.000 línguas faladas em todo o mundo. Entretanto, Austin e Sallabank (2011) afirmam que "pelo menos metade delas pode não continuar a existir depois de mais algumas gerações, pois não estão sendo aprendidas pelas crianças como primeiras línguas". O autor denomina essas línguas como LÍNGUAS EM PERIGO.

Uma maneira de preservar línguas em perigo é a documentação linguística (LIMA-PADOVANI; MIRANDA; BARROS, 2019).

Documentação de línguas é uma prática constante de pesquisadores em linguística (LIMA-PADOVANI; MIRANDA; BARROS, 2019), os quais demonstram forte preocupação pelas línguas que estão extintas, como algumas línguas indígenas no Brasil, citadas pelas autoras.

O processo de documentação de pesquisadores de línguas ameaçadas começa quando linguistas realizam viagens etnográficas para coleta de gravação de áudio ou vídeo (PRIDE; TOMLIN; ANDERBOIS, 2020).

Anotações são ações de transcrever e traduzir uma gravação audiovisual. Existem dois tipos de anotações: anotações mínimas e anotações mais elaboradas. Anotações mínimas consistem em uma transcrição e tradução livre acompanhadas de dados primários (gravações audiovisuais) (GIPPERT et al., 2006).

As anotações são feitas por meio de ferramentas computacionais dedicadas. Uma das ferramentas mais poderosas para realizar a transcrição, tradução e posterior análise linguística é o ELAN (WITTENBURG et al., 2006). Tal ferramenta permite, simultaneamente, ver/escutar uma mídia (áudio e/ou vídeo) e realizar anotações (BEREZ, 2007). A autora informa que o ELAN foi originalmente desenvolvido para analisar línguas de sinais e gestos. O ELAN permite aos usuários vincular anotações a fluxos de mídia, vincular anotações a outras anotações, para configurar relacionamentos definidos pelo usuário entre camadas de anotação.

1.2 O Problema

Vimos que existem anotações mínimas e anotações elaboradas:

No entanto, a extensão exata e o formato das anotações que devem ser incluídas em cada sessão é uma questão de debate. É comum distinguir entre esquemas de anotação mínimos e mais elaborados. Um esquema de anotação mínima amplamente

assumido consiste em apenas uma transcrição e um tradução gratuita que deve acompanhar todos, ou pelo menos um número substancial de, segmentos de dados primários. Esquemas de anotação mais elaborados incluem vários níveis de glosas interlineares, gramaticais também comentário etnográfico e extensa referência cruzada entre as várias sessões e recursos compilado em uma determinada documentação (GIPPERT et al., 2006).

ELAN é uma ferramenta de anotação linguística que foi projetada para a criação de anotações de texto para arquivos de áudio e vídeo de uso da linguagem (BRUGMAN; RUSSEL; NIJMEGEN, 2004). É a mais usada atualmente.

Entretanto, devido suas muitas funcionalidades, resulta em uma interface rebuscada, e requer treinamento para novos usuários. Por exemplo, em um projeto de documentação de línguas indígenas do norte do Amapá, Campetela et al. (2017) informam que fizeram curso e oficinas, de ferramentas, incluindo o ELAN, para professores indígenas com o objetivo de apresentá-los às ferramentas de documentação e análise linguística e ao conhecimento teórico necessário para que eles possam implementar ações de documentação em suas próprias aldeias.

Além da interface rebuscada para realizar as anotações, vale lembrar também que o ELAN não é uma ferramenta ideal para apresentar o material ao público externo. São de diversas áreas as pessoas com interesse nesse tipo de material, além da comunidade de linguistas, antropólogos, historiadores, musicólogos, artistas de diversas áreas e o público em geral (pessoas sem conhecimento em documentação linguísticas).

Ferramentas profissionais são difíceis de aprender e de gerenciar. Entretanto, o avanço da tecnologia web tem fomentado o surgimento de diversas ferramentas computacionais. Arquitetura padrão de aplicativos da web são comumente estruturada por HTML, CSS e códigos JavaScript. Segundo Flanagan (2004)

JavaScript & a linguagem de programação da Web. A ampla maioria dos sites modernos usa JavaScript e todos os navegadores modernos em computadores de mesa, consoles de jogos, tablets e smartphones - incluem interpretadores JavaScript, tornando-a a linguagem de programação mais onipresente da história. JavaScript faz parte da triade de tecnologias que todos os desenvolvedores Web devem conhecer: HTML, para especificar o conteúdo de páginas Web; CSS, para especificar a apresentação dessas páginas; e JavaScript, para especificar o comportamento delas.

Muitas bibliotecas e projetos JavaScript surgiram nos últimos anos. Um desses projetos é npm, curto para Node Package Manager. É um repositório online para a publicação de projetos Node.js de código aberto e um utilitário de linha de comando para interagir com o referido repositório que ajuda na instalação do pacote, versão gerenciamento e gerenciamento de dependências.

Novos projetos web tem surgido visando sistemas de anotações na internet. Em (DOB-
BRIN; ROSS, 2017), os autores mostram uma ferramenta para exibição de material em lin-
guísticas. Chamada de "IATH ELAN Text-Sync Tool"(ETST). O ETST apresenta as anotações
sincronizada com o tempo. ETST funciona com três arquivos principais:

1. um arquivo HTML que pega as informações baseadas em XML em um arquivo ELAN e
as transforma em HTML para que um navegador da web possa lê-lo;
2. um arquivo JavaScript que gerencia os recursos interativos da tela, como rolagem e clique
reproduzir / pausar e
3. um arquivo Cascading Style Sheet que fornece as instruções de formatação.

Pride, Tomlin e AnderBois (2020) desenvolveram uma ferramenta web, denominada
LingView. LingView é uma interface da web para visualização de arquivos ELAN e FLEx
(*FieldWorks Language Explorer*, do SIL). Dobrin e Ross (2017, p. 2) afirmam que

A interface ELAN é otimizada para transcrição e análise, não exibição, portanto,
não há como livrar a tela da interface de seus muitos botões, guias de menu, visuali-
zadores e controles potencialmente perturbadores. Para reproduzir arquivos ELAN,
o programa precisa ser carregado em um computador e há etapas apropriadas para
associar o arquivo ELAN ao arquivo de mídia correto. Embora isso possa ser de
menor importância para um linguista que passou muitas horas produzindo anotações
em ELAN, pode ser confuso para alguém não familiarizado com o programa que
apenas deseja ouvir e ver uma gravação transcrita.

Assim, o LingView foi desenvolvido para tornar o compartilhamento de anotações de
materiais audiovisuais linguísticos com públicos que não entendem de ELAN.

Segundo Pride, Tomlin e AnderBois (2020) foi usado, na criação da interface do Ling-
View, o framework React. Para Aggarwal (2018) ReactJS é uma biblioteca JavaScript implantada
para desenvolver componentes reutilizáveis de interface de usuário (IU). De acordo com a Docu-
mentação oficial do React ¹, React é uma biblioteca para construção de interfaces de usuário
modulares. O React basicamente permite o desenvolvimento de aplicativos grandes e complexos
baseados na web que podem alterar seus dados sem atualizações de página subsequentes.

O pipeline de dados do LingView possui várias etapas, que será discutido na subseção
2.2.2. Permitindo aos usuários mesclar arquivos ELAN e FLEx em arquivos JSON proprietários.
Este formato JSON é então convertido em HTML pelo código ReactJS de front-end. O processo
de conversão de arquivos ELAN e FLEx em arquivos JSON é feito no backend, desenvolvido
em Node.js.

¹ <https://reactjs.org/>

Node.js é outro projeto JavaScript. Segundo Tilkov e Vinoski (2010) "Node.js - também chamado de Node - é um ambiente JavaScript do lado do servidor (consulte <http://nodejs.org>). É baseado no tempo de execução de implementação do Google - o motor apropriadamente denominado "V8". V8 e Node são implementados principalmente em C e C++, com foco em desempenho e pouco consumo de memória. Mas, enquanto o V8 suporta principalmente JavaScript no navegador (mais notavelmente, Google Chrome), o Node tem como objetivo oferecer suporte a longa duração a processos do servidor."

É importante ressaltar que a documentação de línguas é um trabalho exaustivo. Reiman (2010, p. 263) relata que "meio dia de gravações (3 horas - 60 eventos comunicativos @ 3 minutos cada), um anotador tem que trabalhar quase uma semana inteira (pelo menos 33 horas) para completar o processo de anotação". Assim, a participação de membros das comunidades de línguas ameaçadas, como indígenas autodidatas ou em formação acadêmica que tem habilidade de manusear um computador, é de extrema importância e contribuição para a comunidade linguística. A ameaça a uma língua significa a perda de diversos momentos de socialização de uma comunidade, tais como, rituais, canções tradicionais, narrativas orais e etc. Dessa forma, a documentação (gravação, armazenamento adequado, transcrição e tradução, minimamente) é de extrema relevância para salvaguardar essas práticas discursivas para a comunidade linguística envolvida e para o mundo.

Entretanto, cursos de treinamento de ferramentas profissionais de sistemas anotações, como o ELAN, para usuários iniciantes, requer investimentos. Isso é um problema que implora por solução.

1.3 Justificativa

A presente monografia foi pensada, globalmente falando, como uma contribuição ao corpus da língua indígena Wayoro, armazenado no Museu Paraense Emílio Goeldi, cf. <https://arqling.museu-goeldi.br/>. A criação de uma ferramenta de transcrição e anotação de fácil manuseio permitirá a participação de indígenas no processo de documentação e contribuirá para o aumento de amostras de gravações anotadas sobre essa. Trata-se de uma língua em vias de desaparecimento (com apenas duas falantes fluentes idosas), ou seja, a documentação da mesma é urgente.

Denominamos esta ferramenta de Language Annotation Web System (LAWS) essa ferramenta. O LAWS não tem servidor web nem de banco de dados, é tudo processado no lado do cliente usando módulos em JavaScript. O usuário poderá adicionar anotações de duas camadas aos materiais audiovisuais, a saber: transcrição e tradução. É importante ressaltar que reutilizamos o código do LingView (PRIDE; TOMLIN; ANDERBOIS, 2020), que, por sua vez, também usou código-fonte de outros códigos abertos (DOBRIN; ROSS, 2017).

1.4 Objetivos

O presente trabalho tem como objetivo geral desenvolver uma ferramenta web de fácil apresentação para edição e criação de anotações linguísticas usando o formato padrão de arquivos do sistema de software EUDICO Linguistic Annotator (ELAN).

Os objetivos específicos do trabalho são:

- Analisar as funcionalidades da ferramenta, que será aproveitado o código do LingView, que o mesmo será estudado para adaptação de alguns processamentos para o lado do cliente;
- desenvolver o Sistema Web usando o framework React.js, denominado LAWS para ser usado para documentação de materiais em linguística;
- testar a ferramenta quanto a funcionalidade e usabilidade;
- escrever um manual de manuseio da ferramenta LAWS.

1.5 Estrutura do Trabalho

A parte desse capítulo de Introdução, Capítulo 1, este trabalho está estruturado da seguinte forma: o Capítulo 2, Referenciais Teóricos e Trabalhos Relacionados, que apresenta conceitos usados neste documento e projetos relacionados à ferramenta LAWS que está sendo apresentada neste trabalho. No Capítulo 3, temos a Metodologia utilizada neste trabalho, desde a etapa de desenvolvimento da ferramenta com suas modelagens. No Capítulo 4 é mostrado os resultados e no capítulo 5, é apresentada a conclusão do trabalho, também são apontados possíveis trabalhos futuros que possam agregar mais qualidade e utilidade ao processo.

2 REFERENCIAS TEÓRICOS E TRABALHOS RELACIONADOS

Neste capítulo, abordaremos conceitos importantes a respeito deste projeto, e também a apresentação de trabalhos relacionados junto com suas vantagens e desvantagens.

2.1 Referencias Teóricas

2.1.1 Documentação Linguística

Documentação linguística, resumidamente, é um registro duradouro e multifuncional de uma língua. Tal registro é feito em áudio e/ou vídeo, sobre os quais são produzidas anotações linguísticas que permitem a compreensão global dos mesmos. Existem dois tipos de anotações: anotações mínimas e anotações complexas (GIPPERT et al., 2006). Anotações mínimas são somente a transcrição e a tradução de gravações de vídeos ou áudios. Já anotações complexas, conhecida como análise descritiva de anotações, envolvem análises da gramática da língua, comparações de línguas etc. Geralmente, o trabalho de documentação é feito por pesquisadores em linguística com treinamento para coleta de materiais audiovisuais e outras práticas que envolve documentação.

Da mesma forma, Gries e Berez (2017) definem que a anotação linguística acontece quando um arquivo audiovisual em um determinado idioma é transcrito e traduzido. Audiovisual são gravações de áudios ou vídeos. (GIPPERT et al., 2006) denominam essas gravações como dados linguísticos primários, enquanto que a transcrição, a anotação e os metadados descritivos (anotações descritivas, como gramaticais) são conhecidos como aparatos dos corpus.

(GRIES; BEREZ, 2017) apresentam o seguinte conjunto de características de um corpus:

- consiste em um ou mais arquivos de texto Unicode legíveis por máquina;
- destina-se a ser representativo de um tipo específico de orador, registro, variedade, ou língua como um todo, o que significa que o esquema de amostragem do corpus representa a variabilidade da população que pretende representar;
- deve ser equilibrado, o que significa que os tamanhos das subamostras (de falantes, registros, variedades) são proporcionais às proporções de tais falantes, registros, variedades etc. na população que o corpus pretende representar;
- contém dados de configurações comunicativas naturais, o que significa os dados no corpus não foram produzidos apenas com o propósito de ser inserido em um corpus (não foram produzidos em entrevistas controladas)

Pesquisadores em linguística são os responsáveis por documentar e descrever anotações de línguas ameaçadas (GRIES; BEREZ, 2017). Geralmente, tais cientistas fazem viagens etnográficas para coletar materiais audiovisuais de uma determinada língua, quer seja áudio ou vídeo. Com o material audiovisual coletado, são elaboradas as anotações.

2.1.2 Conceitos sobre a Arquitetura de software do LAWS

As tecnologias usadas no desenvolvimento do LAWS foram as bibliotecas bem populares na linguagem JavaScript. O site Micreiros.com define o conceito de biblioteca:

Bibliotecas são um conjunto de funções, métodos e objetos que tem como objetivo facilitar e tornar mais rápido o desenvolvimento de aplicações, não sendo necessário recriar as funções todas as vezes que estas forem necessárias, é importante lembrar que uma biblioteca normalmente é menos complexa, e mais flexível que um framework ¹.

ReactJS é a tecnologia principal da arquitetura do LAWS. É uma biblioteca JavaScript para criar interfaces de usuário. É Muito conhecida por prometer desenvolvimento eficiente e rápido. A documentação oficial do React é bastante intuitiva, o que facilita a aprendizagem para os iniciantes. A documentação pode ser encontrada no site oficial do ReactJS ².

Detalhes do funcionamento do React são muito longos para explicar aqui, então apresentaremos o resumo da biblioteca encontrado no site Blog da Trybe ³.

Para entender como o React funciona é preciso conhecer a estrutura de uma página HTML e de que forma ela é manipulada. A página HTML é formada por uma série de tags que representam cada elemento de um conteúdo. Já o DOM — Document Object Model — é uma representação estruturada de todos os elementos da página e é utilizado para facilitar as alterações nesses elementos.

O React cria um DOM Virtual para permitir a manipulação dos elementos da página. Basicamente, ele realiza uma cópia do DOM em memória, faz as alterações necessárias conforme a determinação das instruções programadas e atualiza o navegador de forma inteligente. Esse processo oferece mais agilidade na manipulação das páginas ⁴.

A outra biblioteca utilizada na ferramenta LAWS é o Redux. Redux é uma biblioteca JavaScript para gerenciar o estado da aplicação, é similar e inspirada pela arquitetura Flux do

¹ <https://micreiros.com/o-que-e-e-como-funciona-o-react-e-o-redux/>

² <https://pt-br.reactjs.org/docs/getting-started.html>

³ <https://blog.betrybe.com/desenvolvimento-web/react/>

⁴ <https://blog.betrybe.com/desenvolvimento-web/react/>

Facebook, ela pode ser utilizada com o React ou outras bibliotecas JavaScript, foi criada por Dan Abramov e Andrew Clark em 2015.

O site [Micreiros.com](http://micreiros.com) explica que o Redux cria um local centralizado para armazenar os dados da aplicação. No LAWS, o dado centralizado é o JSON, que é detalhado na seção 3.2.3.

Durante o desenvolvimento do código, a biblioteca ReactJS é compilado pelo framework NodeJS, que é um ambiente JavaScript do lado do servidor. É baseado na implementação de tempo de execução do Google - o mecanismo apropriadamente chamado “V8”. V8 e Node são implementados principalmente em C e C ++, com foco no desempenho e baixo consumo de memória (TILKOV; VINOSKI, 2010).

O JavaScript possui o conceito de módulos, que são arquivos que podem conter classes ou bibliotecas de funções para usos específicos ⁵. Para quem usa ou já usou JavaScript, deve ter usado a palavra chave *require*, que é utilizada para importar módulos ⁶. Entretanto, o navegador não possui suporte para a função *require*.

Para resolver esse problema, utilizamos o empacotador de módulos Webpack. O empacotador vai gerar um novo arquivo a partir do arquivo principal do projeto, de forma recursiva, resolvendo as dependências das dependências ⁷ Esse arquivo vai conter código original e pode ser executado pelo navegador sem apresentar nenhum problema.

Webpack possui 5 conceitos principais ⁸: Entry, Output, Loaders, Plugins e Mode. Entry é o ponto de entrada o qual definimos para o webpack iniciar a construção do gráfico interno. No caso da ferramenta LAWS, definimos o arquivo `./src/App.js`. Output é a propriedade que define nome e local do pacote gerado pelo webpack. Outra propriedade do webpack é o Loaders. O webpack entende somente arquivos Javascript, e para suporte de outros tipos de arquivos é necessário o uso de Loaders, que são módulos instalados separadamente para que o empacotador converta esses arquivos em módulos válidos e os adicione ao gráfico de dependência. Quando a ferramenta esta em desenvolvimento, é necessário definir na propriedade Mode o modo *development*. Caso a ferramenta esteja finalizada, o modo é *production*.

2.2 Trabalhos relacionados

Nesta seção, serão apresentados projetos ou ferramentas relacionadas e suas características. As ferramentas que serão descritas são ELAN, LingView, ELAN Text-Sync Tool (ETST) e Tycho Brahe Platform.

⁵ <https://javascript.info/modules-intro>

⁶ <https://devheroes.io/webpack-2-para-iniciantes-o-que-e-porque-usar-e-como-iniciar/>

⁷ <https://blog.codecasts.com.br/ecossistema-javascript-parte-05-bundlers-builders-6809b17ddcf8#.ltj42buse>

⁸ <https://medium.com/rocketseat/entendendo-e-dominando-o-webpack-4b2e8b3e02da>

2.2.1 ELAN

Segundo Wittenburg et al. (2006), dentre as ferramentas de transcrição de línguas, ELAN (*Eudico Linguistic Annotator*) é a ferramenta profissional para documentação e análises de materiais linguísticos mais usada pela comunidade de pesquisadores. Foi desenvolvido pela Max Planck Institute for Psycholinguistics.

As anotações, segundo Wittenburg et al. (2006), são agrupadas em camadas, referidas no ELAN como “Níveis”. As camadas podem ter relacionamentos diferentes uma com a outra, por exemplo, podem estar independentes, alinhadas ou incorporadas. Na Figura 1, mostra-se a interface do ELAN.

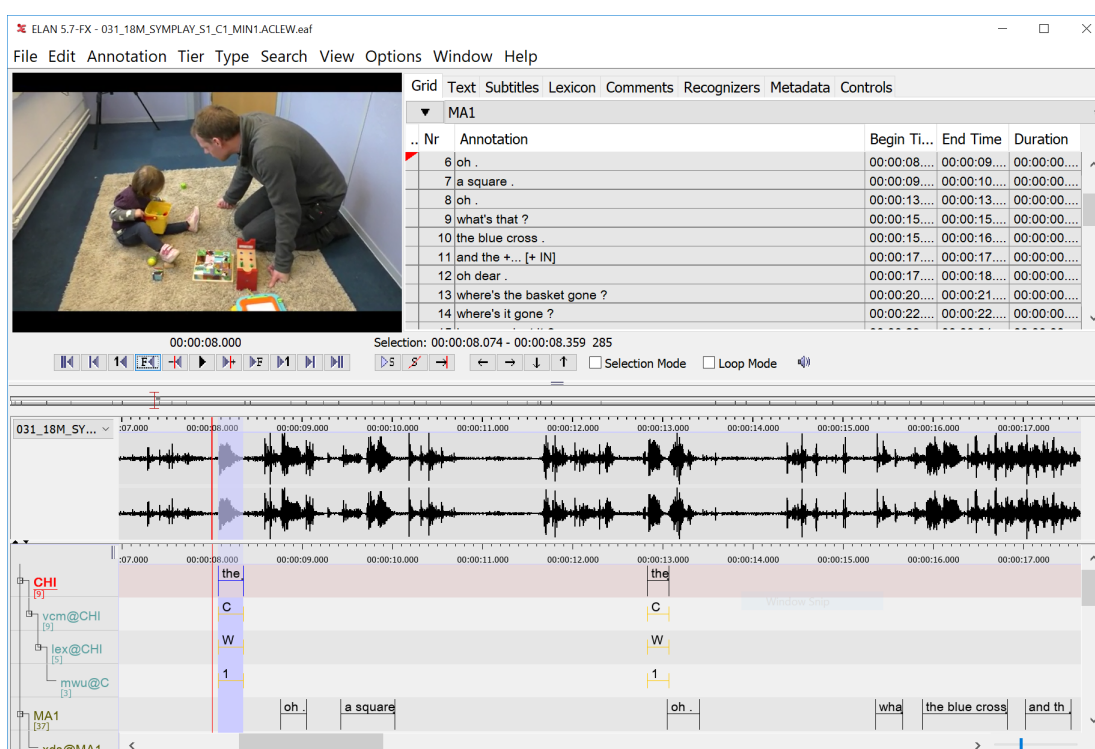


Figura 1 – Interface do ELAN
 Fonte: site <<https://archive.mpi.nl/tla/elan>>

Na documentação do ELAN, encontramos algumas características da ferramenta (HELLWIG, 2021), tais como as seguintes:

- Em ELAN, é possível criar anotações com uma precisão máxima de 1 ms. Em combinação com a capacidade de reprodução precisa de fragmentos, os linguistas podem alcançar a alta precisão em suas anotações que eles necessitam em suas pesquisas.
- Suporte para arquivos de vídeos nos formatos 'mpg' e 'mp4'. E suporte para arquivos de áudios nos formatos 'wav'.

- Pode ter 1 ou até 4 vídeos e podem ser adicionados ou removidos do documento a qualquer momento, em qualquer estágio da transcrição, fornecendo a máxima flexibilidade para o usuário no processo de anotação.

Entretanto, ELAN é uma ferramenta que requer treinamento para iniciantes, ou seja, ferramentas profissionais são difíceis de aprender, como discutido na seção 1.3.

2.2.2 LingView

Conforme (PRIDE; TOMLIN; ANDERBOIS, 2020), o LingView é uma ferramenta web para exibição de materiais a públicos que não necessariamente conhecem ou não fazem parte da comunidade de pesquisadores em linguística, mas tem interesse em conhecer a documentação de determinada língua. Na Figura 2, temos o LingView apresentando materiais.

The screenshot shows the LingView interface. At the top, there is a blue header with the text "LingView: ELAN and FLEx Web Display" on the left and "Index of Texts" on the right. Below the header, there is a video player on the left and a list of transcribed segments on the right. The video player shows a play button and a progress bar at 1:20 / 3:20. The list of segments includes the following text:

1:15 S1: tsa'ma ja'ñiongaetangi panzapa canse'fa potae'ngoi'ccu
tsa='ma ja'ño=ni=ngae=ta=ngi panza=pa canse-'fa potae'ngo-'ccu
ANA=FRUST now=LOC=MANN=TOP=1 hunt=SS live-PL.SUBJ gun=INST

1:20 S1: Coragandeccu tsu yajema cu'ipa
Coraga=ndeccu tsu yaje=ma cu'i=pa
curandero=HUM.PL 3 ayahuasca=ACC drink=SS

1:23 S1: a'indeccuma se'jepa canse'fa
a'=ndeccu=ma se'je=pa canse-'fa
person=HUM.PL cure=SS live-PL.SUBJ

1:25 S1: poi a'i can'qqueni tsu coragandeccu canse'fa
poi a'i can'qque=ni tsu coraga=ndeccu canse-'fa
all person town=LOC 3 curandero=HUM.PL live-PL.SUBJ

Below the video player, there is an "Intro" section with the following information:

Speakers:
• S1: Martin Criollo

Show/hide tiers:

- A'ingae
- English
- Morphemes
- Glossed Morpheme
- Show video

Figura 2 – Interface do LingView

Fonte: baseado em Pride, Tomlin e AnderBois (2020)

O pipeline do Lingview é mostrado na Figura 3. Pride, Tomlin e AnderBois (2020) informam que este pipeline foi elaborado dessa forma para atender às seguintes necessidades:

- prevenção de perda de informações;
- capacidade para o site para carregar de forma rápida e corretamente em várias plataformas, e
- a capacidade do site para funcionar offline.

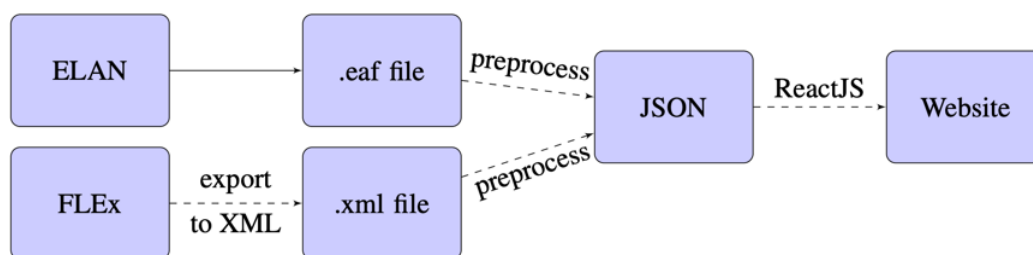


Figura 3 – Pipeline da ferramenta LingView

Fonte: baseado em Pride, Tomlin e AnderBois (2020)

LingView tem suporte para arquivos ELAN e FLEx. O primeiro passo do pipeline do LingView é o 'preprocess' (pré-processamento) dos arquivos. O objetivo do 'preprocess' é converter os arquivos ELAN ou FLEx no formato JSON. Depois de ter os dados no formato JSON, as informações são exibidas em uma página em html com auxílio de códigos em JavaScript (PRIDE; TOMLIN; ANDERBOIS, 2020).

Pride, Tomlin e AnderBois (2020) informam que o código do pré-processamento é um pacote NodeJS, scripts que permitem ao usuário converter arquivos FLEx e ELAN em um único formato JSON. A estrutura do JSON é mostrado na Figura 4.

Os detalhes de todos os parâmetros do arquivos JSON são muito longos para explicar aqui, então apresentaremos apenas um resumo dos parâmetros **metadados** e **sentences**.

O **metadados** contém informações do autor, título, data, mídia de áudio, mídia de vídeo e assim por diante. Por fim, temos as listas de IDs de alto-falante e de nível. IDs de alto-falante fornecem uma abreviatura como "S1" e "S2", conforme mostrado na Figura 4. Mais detalhes do **metadados** são encontrados no documento de (PRIDE; TOMLIN; ANDERBOIS, 2020).

No **sentences** temos uma lista, gerada a partir de anotações ELAN independentes ou de sentenças FLEx. Cada frase tem alguns dados associados, como o locutor, nível e hora de início / término. Mais detalhes de todos os parâmetros do **sentences** são encontrados no documento de (PRIDE; TOMLIN; ANDERBOIS, 2020).

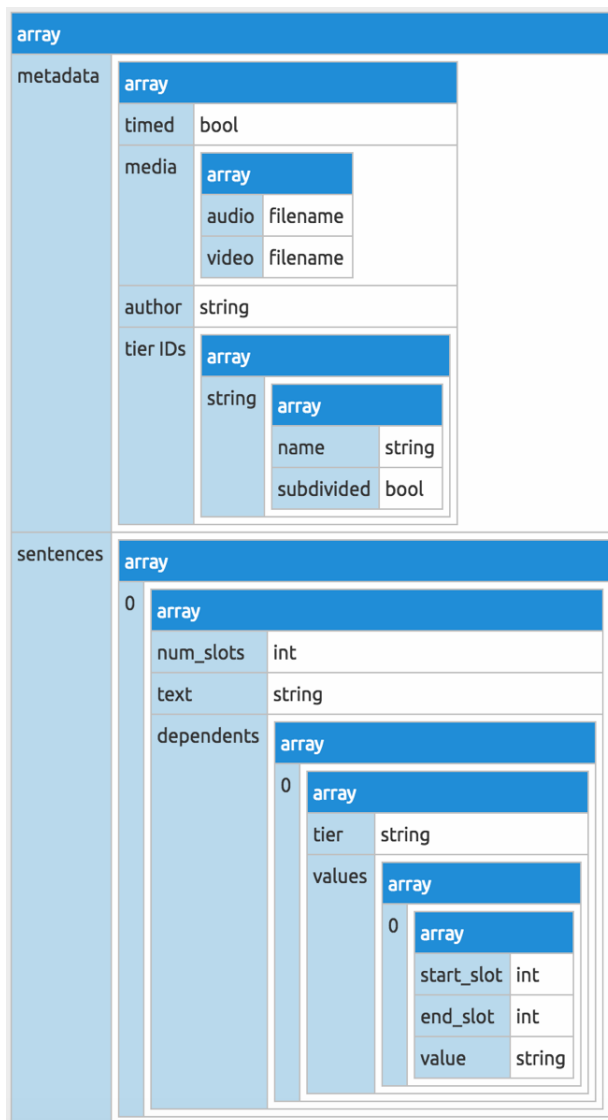


Figura 4 – Estrutura do JSON
 Fonte: baseado em Pride, Tomlin e AnderBois (2020)

O LingView tem outro arquivo JSON chamado de **database.json**. Este arquivo simplesmente contém uma lista de histórias, representada como acima (arquivo JSON, da Figura 4), junto com um índice que é usado para gerar o índice de histórias. Este arquivo também é usado para gerar o site. O JSON individual as representações podem ser visualizadas em data/json_files.

O LingView segue o modelo de cliente-servidor (PRIDE; TOMLIN; ANDERBOIS, 2020). Como já mencionamos na seção 1.3, a interface do site do LingView é construída em ReactJS, que é uma biblioteca Javascript baseada em componentes para construir interfaces de usuário. Um dos componentes criados para representações de dados do arquivo JSON é mostrado na Figura 5.

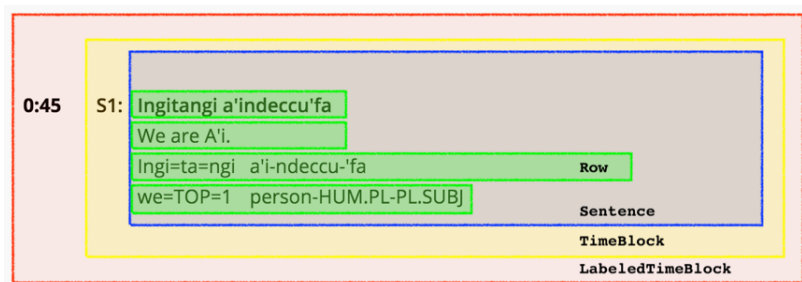


Figura 5 – Componente de representação de uma sentença

Fonte: baseado em Pride, Tomlin e AnderBois (2020)

A instalação da ferramenta consiste em seguir alguns passos, como é mostrado a seguir:

1. Baixe o projeto no link <<https://github.com/BrownCLPS/LingView.git>>
2. Extraia o projeto
3. Baixe o Node.JS e os pacotes NPM, utilizando o comando *npm install*.

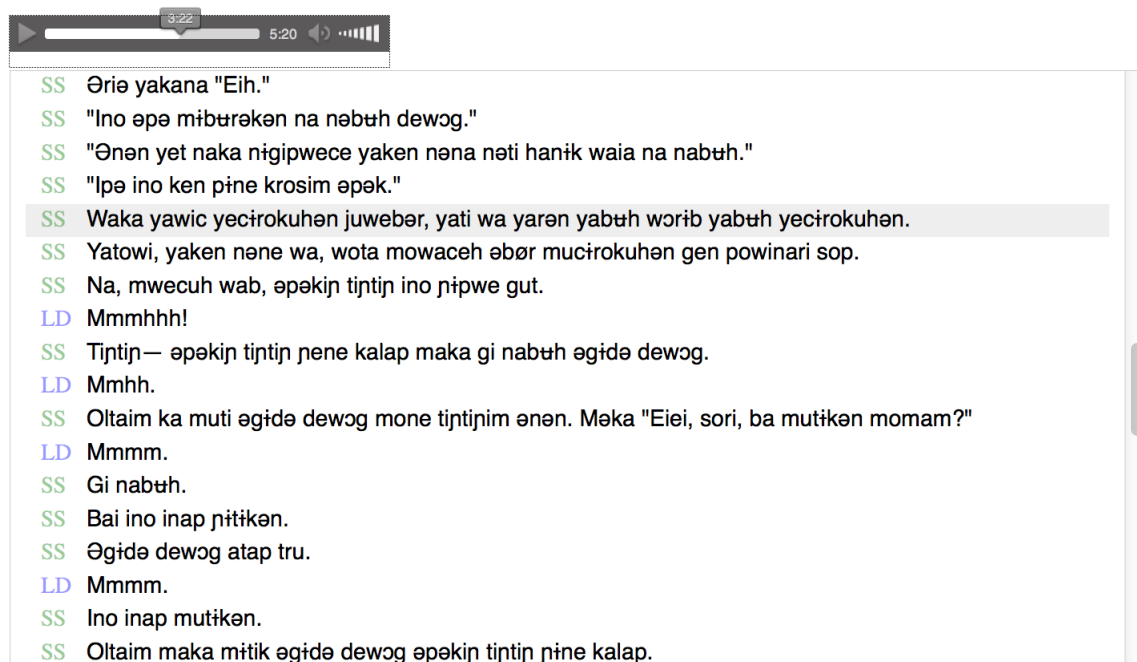
A ferramenta LingView é um excelente projeto e muito essencial para exibição de materiais. É certo que, para alguns iniciantes pode haver dificuldades em instalar a ferramenta. Para mais detalhes do funcionamento do LingView, consulte o documento (PRIDE; TOMLIN; ANDERBOIS, 2020).

2.2.3 ELAN Text-Sync Tool(ETST)

ETST é um simples reproduzidor de transcrição baseado em navegador que pode ser usado como está, "pronto para uso" ou como base para um maior desenvolvimento. Codificado inteiramente em HTML compatível com os padrões, sem depender de plug-ins de terceiros, o ETST é extremamente eficiente, confiável e compatível com outros sistemas. ETST também foi projetado para ser responsivo a diferentes ambientes digitais, trabalhando em dispositivos e resoluções de tela (DOBRIN; ROSS, 2017).

When Timothy Fell in The Latrine

LD · Lise Dobrin SS · Scola Sonin



3:22 5:20

SS Əriə yakana "Eih."
 SS "Ino əpə mibərəkən na nəbəh dewəg."
 SS "Ənən yet naka nigipwece yaken nəna nəti hanik waia na nabəh."
 SS "Ipə ino ken pine krosim əpək."
 SS Waka yawic yecirokuhən juwebər, yati wa yarən yabəh wərib yabəh yecirokuhən.
 SS Yatowi, yaken nəne wa, wota mowaceh əbər mucirokuhən gen powinari sop.
 SS Na, mwecuh wab, əpəkiŋ tiŋtiŋ ino niŋpwe gut.
 LD Mmmhhh!
 SS Tiŋtiŋ— əpəkiŋ tiŋtiŋ nene kalap maka gi nabəh əgide dewəg.
 LD Mmhh.
 SS Oltaim ka muti əgide dewəg mone tiŋtiŋim ənən. Məka "Eiei, sori, ba mutikən momam?"
 LD Mmmm.
 SS Gi nabəh.
 SS Bai ino inap niŋtikən.
 SS Əgide dewəg atap tru.
 LD Mmmm.
 SS Ino inap mutikən.
 SS Oltaim maka mitik əgide dewəg əpəkiŋ tiŋtiŋ niŋe kalap.

Gloss:

SS Nau mi kisim em igo insait nau wasim em long hotwara nau mi lukim nogat nau mi kisim em igo daun long wara nau mi wasim em gen.

Figura 6 – Interface do ETST
 Fonte: baseado em Dobrin e Ross (2017)

Existem três arquivos principais que constituem o sistema ETST:

- um arquivo de pré-processamento de hipertexto (txt_sync.php) que pega as informações baseadas em XML e um arquivo ELAN e as transforma em HTML para que possam ser lidas por um navegador da web,
- um arquivo JavaScript (txt_sync.js) que conduz os recursos interativos da tela, como rolagem e reprodução / pausa clicável e
- um arquivo Cascading Style Sheet (txt_sync.css) que fornece as instruções de formatação para a exibição.

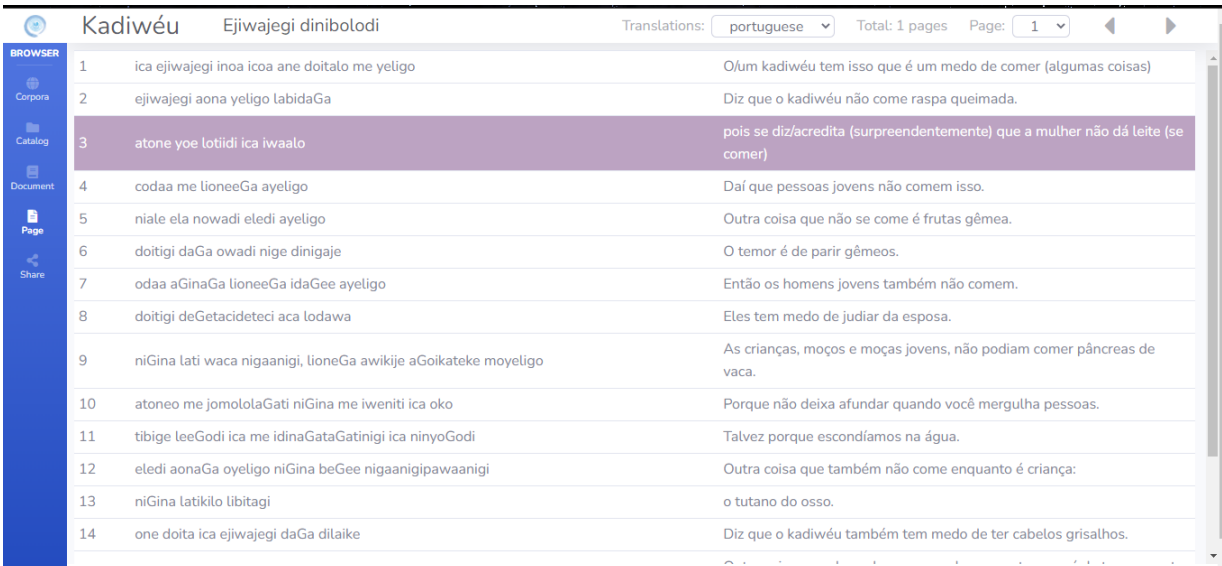
ETST é ferramenta de visualização simples, entretanto, para uso, é necessário que o arquivo de entrada ELAN esteja configurado para que seja legível para o sistema (DOBRIN; ROSS, 2017). Ou seja, o processamento de dados do ETST não é robusto e o usuário, "muitas vezes deve modificar o arquivo para que sua estrutura de camadas esteja em conformidade com as especificações do sistema" (DOBRIN; ROSS, 2017). Para usuários iniciantes será difícil configurar arquivo na ferramenta ELAN. Assim, temos um problema que precisa de uma solução.

2.2.4 Tycho Brahe Platform

A plataforma Tycho Brahe possui ferramentas para criar, editar, gerenciar e pesquisar seus próprios corpora anotados sintaticamente com identificador de PDV integrado e ferramentas de análise sintática <<https://www.tycho.iel.unicamp.br/home>>. A seguir, são detalhadas as ferramentas do Tycho Brahe.

- Tycho Brahe Browser é a sua ferramenta de plataforma para navegar pelos corpora disponíveis. Na Figura 7 temos apresentação de um documento da língua Kadiwéu. Kadiwéu é uma língua Guaikuru falada em Mato Grosso do Sul, Brasil <<https://www.tycho.iel.unicamp.br/browser/catalog/C12>>.
- Tycho Brahe eDictor é a ferramenta de edição de anotações da plataforma Tycho. No entanto, o acesso a esta ferramenta é restrito <<https://www.tycho.iel.unicamp.br/edictor/unauthorized>>.
- Tycho Brahe Search é a sua ferramenta de pesquisa linguística nos corpora disponíveis. Seleciona-se um corpus clicando sobre ele ao lado ou use o filtro acima <<https://www.tycho.iel.unicamp.br/search/>>.
- Tycho Brahe Parser tem mecanismos de analisador e interpretador. O acesso a esta ferramenta é restrito <<https://www.tycho.iel.unicamp.br/parser>>.

Tycho Brahe é uma plataforma muito completa, porém o acesso às ferramentas de sistema de anotação é restrito. Ou seja, precisa de permissão para acesso. Outra limitação é que ainda não é possível trabalhar com arquivos de vídeo na plataforma.⁹



Kadiwéu	Ejiwajegi dinibolodi	Translations: portuguese
1	ica ejiwajegi inoa icoa ane doitalo me yeligo	O/um kadiwéu tem isso que é um medo de comer (algumas coisas)
2	ejiwajegi aona yeligo labidaGa	Diz que o kadiwéu não come raspa queimada.
3	atone yoe lotiidi ica iwaalo	pois se diz/acredita (surpreendentemente) que a mulher não dá leite (se comer)
4	codaa me lioneeGa ayeligo	Daí que pessoas jovens não comem isso.
5	niale ela nowadi eledi ayeligo	Outra coisa que não se come é frutas gêmeas.
6	doitigi daGa owadi nige dinigaje	O temor é de parir gêmeos.
7	odaa aGinaGa lioneeGa idaGee ayeligo	Então os homens jovens também não comem.
8	doitigi deGetacideteci aca lodawa	Eles tem medo de judiar da esposa.
9	niGina lati waca nigaanigi, lioneeGa awikije aGoikateke moyeligo	As crianças, moços e moças jovens, não podiam comer pâncreas de vaca.
10	atoneo me jomololaGati niGina me iweniti ica oko	Porque não deixa afundar quando você mergulha pessoas.
11	tibige leeGodi ica me idinaGataGatinigi ica ninyoGodi	Talvez porque escondíamos na água.
12	eledi aonaGa oyeligo niGina beGee nigaanigipawaanigi	Outra coisa que também não come enquanto é criança:
13	niGina latikilo libitagi	o tutano do osso.
14	one doita ica ejiwajegi daGa dilaike	Diz que o kadiwéu também tem medo de ter cabelos grisalhos.

Figura 7 – Apresentação da língua Kadiwéu na ferramenta Tycho Brahe

Fonte: baseado em <<https://www.tycho.iel.unicamp.br/browser/translation/6aacec70-e929-11ea-bdb7-024>>

⁹ SANDALO, 2021, congresso Seminário do Grupo de Estudos Linguísticos de São Paulo, evento online.

3 METODOLOGIA

Este capítulo descreve o desenvolvimento, teste e utilização da ferramenta de apoio à produção de anotações linguísticas por usuários, sem precisar de treinamento prévio. A metodologia utilizada nesse projeto foi o estudo de caso, que tem como objetivo estudar e detalhar profundamente todas as etapas deste trabalho.

Com base no modelo Interativo (BRAATZ; OLIVEIRA; ROCHA, 2018), de desenvolvimento de software, sugerimos uma especialização neste modelo para o desenvolvimento da ferramenta. O pipeline é mostrado na Figura 8.

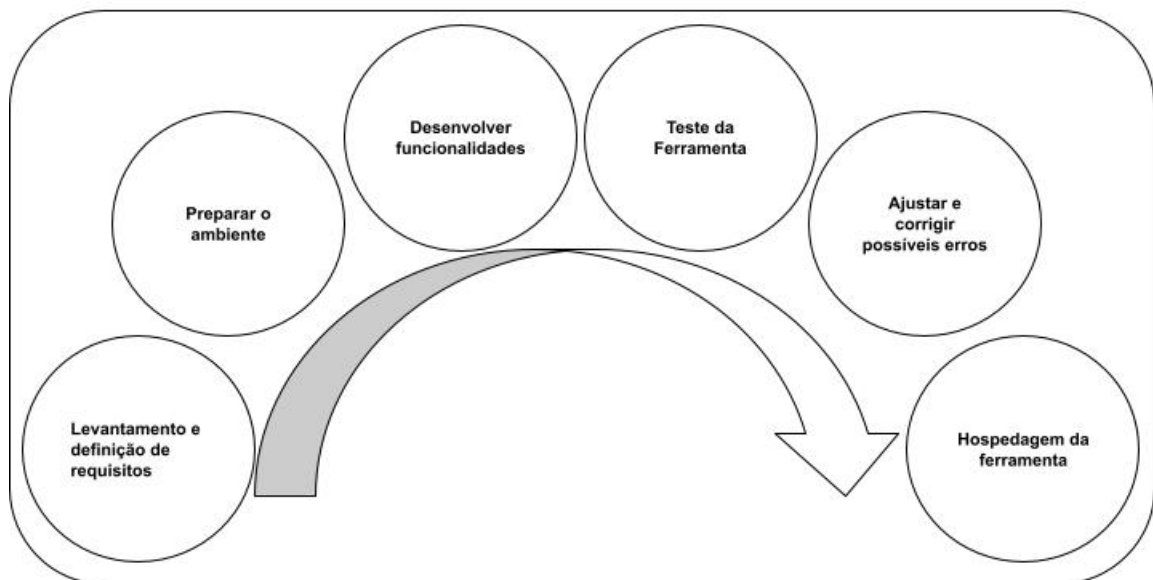


Figura 8 – Pipeline do modelo Interativo do LAWS

Fonte: Autoria própria

É importante ressaltar que poderíamos ter utilizado outros modelos, ao invés do modelo Interativo. Optou-se pelo modelo Interativo pelo fato de que o LAWS seja uma ferramenta simples e de poucas funcionalidades. E também, os requisitos não eram bem conhecidos. O que foi necessário que o usuário interagisse com o sistema para que fosse descoberta suas verdadeiras necessidades (TAROCO; WERNER, 2007).

3.1 Levantamento e Definição de requisitos

Esta etapa consiste na análise de requisitos de acordo com a necessidade do usuário. Tal análise define as novas funcionalidades que o software deverá atender.

Como adaptamos o **pré-processamento**, que é o próprio backend, no frontend, será detalhado a seguir somente os requisitos da plataforma web.

Os requisitos do LAWS foram passados pela Professora Dra Antonia Fernanda de Souza Nogueira, linguista que trabalha na documentação da língua indígena Wayoro. A plataforma web aqui apresentada tem como objetivos principais, até o presente momento, os três itens abaixo:

- Abrir ou criar projeto para começar a produzir anotações em linguística.
- Edição, onde o usuário poderá editar, deletar e adicionar anotações.
- Exportação do material, na qual será gerado arquivo XML que seja compatível com o software ELAN.

Não é possível criar uma ferramenta web de grande poder de processamento sem utilizar o backend (servidor remoto). Entretanto, limitando a entrada de mídias de duração maior que 10 minutos, temos uma ferramenta em potencial. Alertando o usuário quanto à limitação, possivelmente podemos ter uma planejamento em gravações de mídias e uma organização de materiais em módulos sobre temas comuns, por exemplo, vários módulos, tratando de cultura, comidas típicas, tradição e etc.

Atualmente, ferramentas como o ELAN, software profissionais tem muito recursos, e os novos usuários precisam de treinamento. Com esse desafio em mente, adicionamos somente funcionalidades básicas e uma interface amigável.

Com base nesses levantamentos, foram definidos e detalhados o ator, requisitos e regras de negócio da ferramenta web. Os requisitos foram associados a casos de uso, conforme mostrado na tabela 2.

Ator	Nome	Descrição
01	Usuário	Pesquisadores em linguística, estudantes ou pessoas autodidatas interessadas em contribuir para um determinada língua. Serão responsáveis por criar documentações de línguas que estão em perigo.

Tabela 1 – Identificação dos Atores

Fonte: Autoria própria

UC	Nome
UC01	Criar novo projeto
UC02	Abrir Projeto
UC03	Deletar anotação
UC04	Inserir anotação
UC05	Editar anotação
UC06	Exportar

Tabela 2 – Identificação dos Casos de Usos

Fonte: Autoria própria

#RN	Descrição	UC
01	Mídia de áudio deve pertencer ao seguintes formatos: MP3 e WAV.	UC01
02	Mídia de video deve pertencer ao seguintes formatos: MP4, MPEG-4 e WebM.	UC01
03	As duas entradas de arquivos devem ser preenchidas.	UC02
04	O arquivo de entrada XML deve ser do formato EAF.	UC02
05	Todos os campos (Participante, Tempo Inicial, Tempo Final, Transcrição e Tradução) devem ser preenchido.	UC04
06	Tempo Inicial e Tempo Final da nova anotação (não conflitar com o Tempo Inicial e Tempo Final, respectivamente, das anotações salvas) não deve cobrir outras anotações.	UC04
07	O campo (transcrição ou tradução) para edição desejada, deve ser preenchido	UC05

Tabela 3 – Identificação das regras de negócios da ferramenta

Fonte: Autoria própria

A partir do detalhamento de requisitos, foi possível elaborar os diagramas de casos. A Figura 9 representa os casos de uso da ferramenta LAWS.

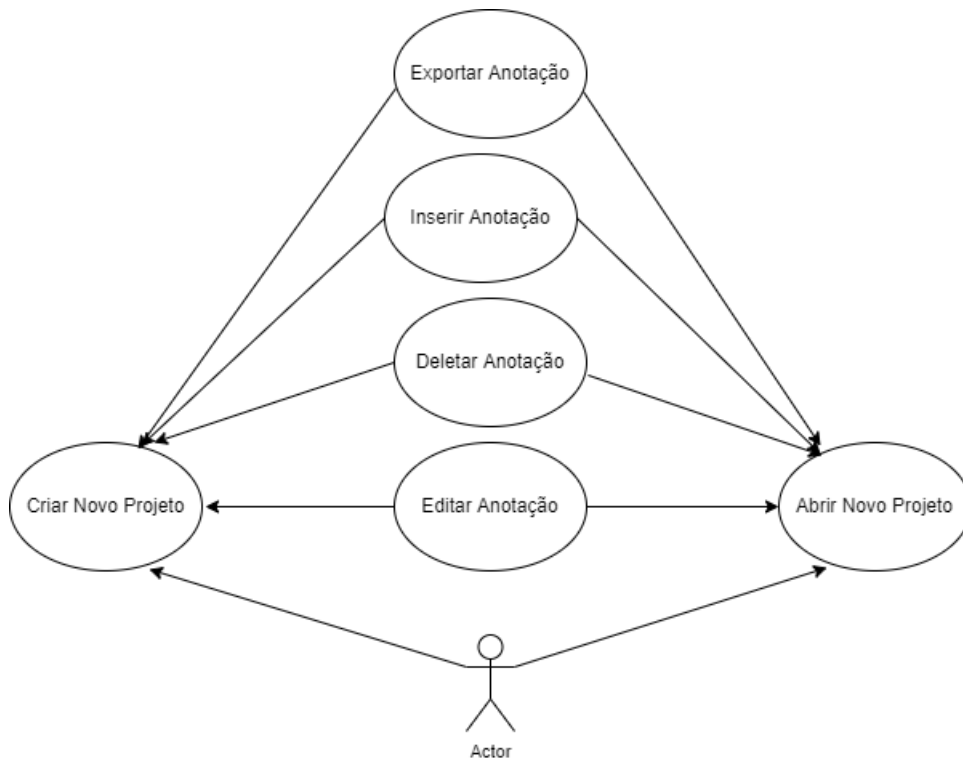


Figura 9 – Diagrama de Casos de uso da ferramenta Laws
 Fonte: Autoria própria

É apresentado na figura a seguir o diagrama de classes para aplicação ReactJS. Cada funcionalidade possui uma classe. Optamos por criar classe para cada funcionalidade, por motivos de que a ferramenta possui apenas 6 casos de usos simples, e também a aplicação está desenvolvida em React, que também foi utilizado componentes para cada funcionalidade, como é mostrado na seção 3.3. No centro do diagrama, temos a classe que cria e gerencia um JSON global usando o Redux, como mostrado na figura 10.

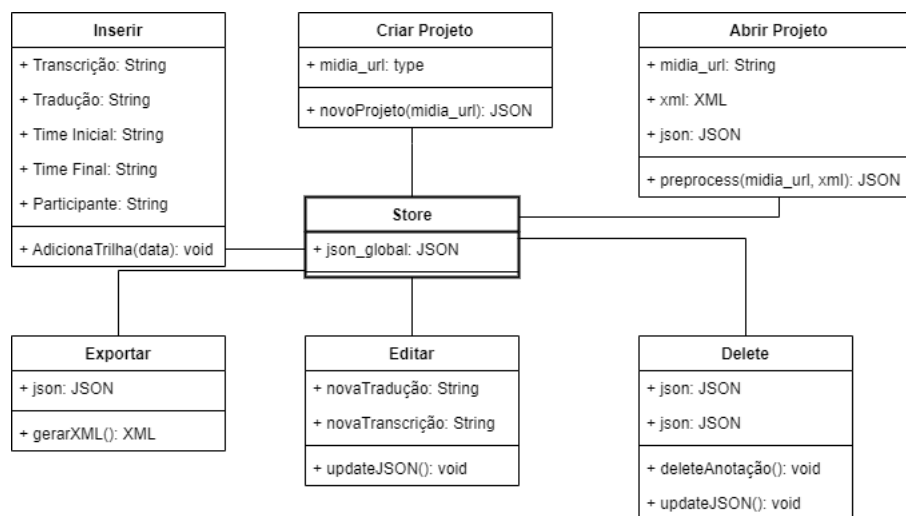


Figura 10 – Diagrama de classes da ferramenta Laws
 Fonte: Autoria própria

Nas figuras a seguir desta seção são apresentados os diagramas de sequências, baseado nas regras de requisitos. Foram detalhados passo a passo a interação entre as classes.

3.1.1 Abrir Projeto

O usuário escolhe os arquivos de entrada, (xml e media) para abrir o projeto. Depois será verificado se a mídia é suportada pela ferramenta; caso seja aceito, será apresentado as anotações ao lado da mídia. Caso contrário, informa o usuário que a ferramenta não suporta tal mídia. Como mostrado na Figura 11.

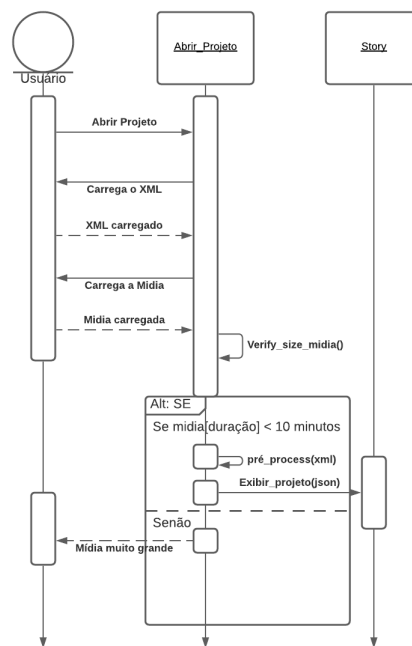


Figura 11 – Diagrama de Sequencia do UC Abrir Projeto
 Fonte: Autoria própria

3.1.2 Criar Projeto

Nessa funcionalidade, o usuário informa somente a mídia. A mídia é verificada, e se for compatível, o novo projeto é criado. Caso contrário, informa o usuário para anexar outro arquivo de mídia.

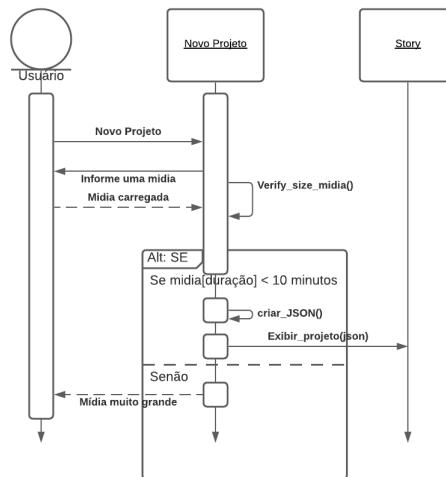


Figura 12 – Diagrama de Sequencia do UC Criar Projeto
 Fonte: Autoria própria

3.1.3 Inserir Anotação

A Figura 13 mostra de maneira organizada o diagrama de sequência do Caso de uso Inserir Anotação.

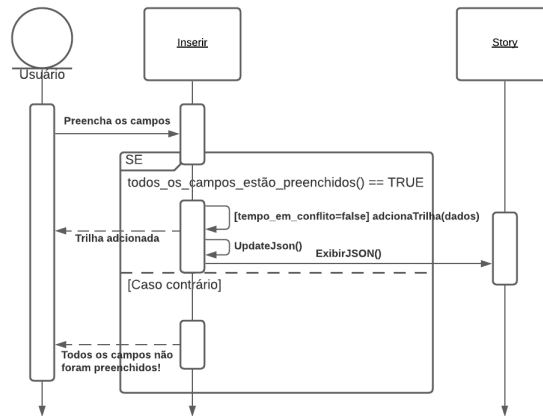


Figura 13 – Diagrama de Sequencia do UC Inserir Anotação
 Fonte: Autoria própria

3.1.4 Deletar Anotação

A Figura 14 mostra de maneira organizada o diagrama de sequência do Caso de uso Deletar Anotação.

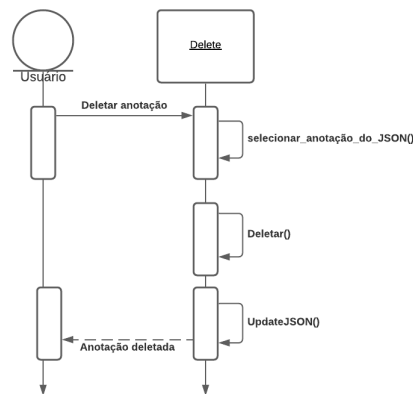


Figura 14 – Diagrama de Sequencia do UC Deletar Anotação
 Fonte: Autoria própria

3.1.5 Editar Anotação

Temos na Figura 13 de maneira organizada o diagrama de sequência do Caso de uso Editar Anotação.

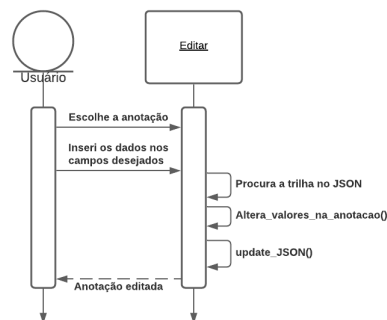


Figura 15 – Diagrama de Sequencias do UC Editar Anotação
 Fonte: Autoria própria

3.2 Preparar o ambiente

Nesta etapa foi realizado o estudo do funcionamento do backend (pré-processamento), conhecimento das bibliotecas usadas e adaptação no frontend.

3.2.1 Backend

O backend é um servidor responsável por fornecer serviços a computadores e a sites

¹. Os serviços oferecidos são, geralmente, de alto poder de processamento. E, dependendo da

¹ <https://www.treinaweb.com.br/blog/o-que-e-front-end-e-back-end/>

necessidade do usuário, essa estrutura é executada em outros computadores de configuração de processamento superior a computadores pessoais.

Uma das tecnologias usadas no backend do LingView é o framework NodeJs, pertencente a linguagem JavaScript. O NodeJS fornece uma gama de módulos que fornece operações específicas, como o `fs`². O `fs` é responsável por ler e escrever dados em diretórios.

Vale relatar que bibliotecas também foram usadas no LingView, como o `XML2js`, que é responsável por transformar XML no objeto JSON³.

3.2.2 Banco de dados

O banco de dados utilizado no backend do lingview é um arquivo JSON, denominado de `'database.json'`. O qual não precisa de SGBD (Sistema de Gerenciamento de Banco de dados). A escolha de uma arquivo JSON como banco de dados pela equipe de desenvolvimento do LingView, partiu da necessidade de apresentar solução simples e de fácil instalação. Devido que o backend é instalado por usuários de pouco conhecimento na área de tecnologias.

3.2.3 Adaptação na web

O frontend possui processamento leve. Assim, adaptar o backend no frontend é limitar o poder de processamento. Como por exemplo, projetos com mídias de duração muito grande, o frontend não terá suporte.

O backend possui duas etapas: processamento e armazenamento no banco de dados. Na adaptação, foi removido certas estruturas, como o processamento de arquivos flex, banco de dados `'database.json'`.

Muitas alterações foram feitas no pré-processamento, como remover o pacote `'fs'`. Nessa adaptação, não precisaríamos mais escrever o JSON em arquivos. Precisávamos de uma variável que armazenasse o JSON. E com isso, tivemos conhecimento de que o Redux é uma biblioteca ideal para atender nossas necessidades. Redux gerencia o estado global, funcionalidade muito importante do framework React, o qual é responsável pelo frontend. Quando um estado global é alterado, por exemplo, recebia uma nova transcrição, todos os componentes ligados ao estado, era renderizado novamente. Mais detalhes do React e Redux é mostrado na subseção 2.1.2.

3.3 Acrescentar funcionalidades

Nesta seção, foi implementado no frontend as novas funcionalidades levantadas. Usando o framework React, que já estava sendo utilizado no projeto.

Os componentes principais, que foram criados, são:

² <https://nodejs.dev/learn/the-nodejs-fs-module>

³ <https://www.npmjs.com/package/xml2js>

- NewProject - Componente responsável por criar novo projeto.
- OpenProject - Componente responsável por abrir e carregar um projeto.
- Edit - Componente responsável por editar anotações.
- Delete - Componente responsável por deletar anotações.
- Export - Componente responsável por gerar e baixar o XML do tipo *.eaf*, *formato suportado pela ferramenta ELAN*.
- Insert - O objetivo deste componente é adicionar novas anotações no JSON.

3.3.1 Implementação

A estrutura de dados para armazenar os dados de anotações e os metadados (informações que são título do projeto, título da mídia, nomes dos participantes, e etc) é o objeto JSON. O JSON é um estado global que precisa ser acessado por todos os componentes de edição. Utilizamos a biblioteca Redux para gerenciar esse estado.

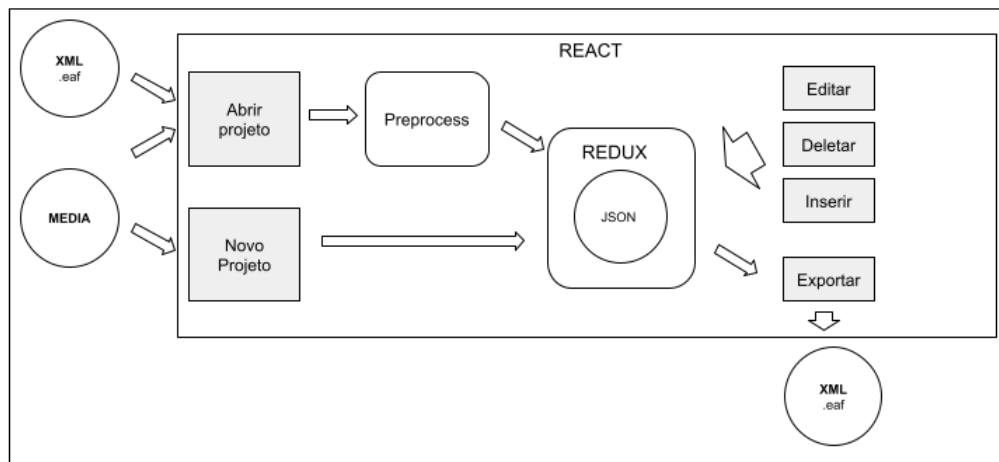


Figura 16 – Arquitetura do LAWS

Fonte: Autoria própria

3.4 Testes e ajustes da Ferramenta

O teste da ferramenta foi realizado por pesquisadores: Antônia Fernanda de Souza Nogueira⁴ e Ivan Rocha da Silva⁵.

O teste foi realizado com arquivos audiovisuais primários, ou seja, gravações feitas em viagens etnográficas. A realização dos testes foi feito de maneira simples, ou seja, os

⁴ <https://www.escavador.com/sobre/437826/antonia-fernanda-de-souza-nogueira>

⁵ <https://www.escavador.com/sobre/5889659/ivan-rocha-da-silva>

pesquisadores tentaram fazer a documentação sem ter as instruções da parte do desenvolvedor da ferramenta. Depois de testado, os pesquisadores retornaram uma lista de erros e sugestões para correção de erros e melhorias na ferramenta.

Dentre o feedback, estão documentados nesta seção os principais erros e sugestões, os quais foram:

- **Anotações não sincronizadas com o tempo:** quando o usuário adicionava anotações e posteriormente reproduzia a mídia, as anotações não estavam sincronizadas com tempo. Isso acontecia porque o script `setupTextSync` que destacava a anotação certa era chamado somente uma vez no `componentDidMount()`, método do ciclo de vida do React. Então, para solução deste problema, adicionamos também o script no método do ciclo de vida `componentDidUpdate()`.
- **Mídia de duração longa não foi carregada:** Arquivos de mídias muito grande não eram carregados no site. Então, limitamos a entrada de mídia com duração maior que 10 minutos.
- **Ausência do manual de uso na ferramenta:** nos testes, o usuário não conseguiu abrir o projeto sem as instruções. Então, adicionamos na pagina *Abrir Projeto* os procedimentos necessários.
- **Erros na função editar:** na função editar, são apresentados dois campos: transcrição e translação. Se o usuário adicionasse texto só no campo transcrição, o campo transcrição era editado e campo translação, que não foi preenchido com textos, apagado. Assim, adicionamos no código para verificar e alterar somente o campo desejado.
- **Ausência de aviso, caso o usuário mude de rota antes de exportar os dados:** durante o uso do site, é normal o usuário clicar pra mudar de rota, e foi sugerido que quando acontecesse esse evento, alertasse o usuário a exportar a documentação antes de mudar de rota. Então, adicionamos esse aviso na pagina de edição.
- **Desabilitar botão editar:** quando o botão editar é clicado, abre um pequeno frame com dois campos de entrada de texto, e a pesquisadora Antonia⁶ sugeriu que adicionasse um botão de desabilitar o frame, caso o usuário termine ou desiste de editar a anotação.
- **Ajustes no texto de abertura:** os pesquisadores que realizaram o teste, elaboraram um texto de apresentação plausível na abertura do site.

3.5 Hospedagem da ferramenta

A hospedagem da ferramenta é feita na 'GitHub Pages'. A escolha do GitHub Pages foram pelos motivos de os serviços serem gratuitos e de fácil atualização do projeto quando

⁶ <https://www.escavador.com/sobre/437826/antonia-fernanda-de-souza-nogueira>

alterado. O código do projeto está carregado na plataforma **Github**, que pode ser acessada por este link: <<https://github.com/marcosamaris/laws>>.

4 RESULTADOS E USO DO LAWS

Neste capítulo, apresentaremos a interface do LAWS, explicaremos suas funcionalidades e apresentaremos resultados qualitativos.

Na Figura 17, temos a interface do LAWS, depois de importado e processado um projeto. Na barra superior, observado na imagem, temos 3 links, que são *Home* (página de apresentação do LAWS), *Abrir Projeto* e *Novo Projeto*. Abaixo dos links de navegação, temos o botão 'Exportar XML', o qual gera o arquivo 'Data.eaf' no formato xml.

No lado esquerdo, temos um formulário HTML solicitando informações para inserir nova anotação. As informações solicitadas são: transcrição, tradução, tempo inicial, tempo final e participante.

Abaixo da anotação, temos um componente de mídia, que consiste em uma tag HTML de vídeo para carregamento de arquivo audiovisual. No próprio componente de mídia, temos controles, a saber: pause, barra deslizante de duração do vídeo, volume e etc.

No lado direito, observa-se as anotações com suas transcrições, traduções, participantes, botões de Editar e Deletar. Isto é, será fácil deletar, clicando no botão de deletar, abaixo de cada anotação, tal como é apresentado na Figura 17. A mesma coisa para Editar, que, quando clicado, é apresentado um pequeno formulário abaixo para edição de transcrição e tradução.

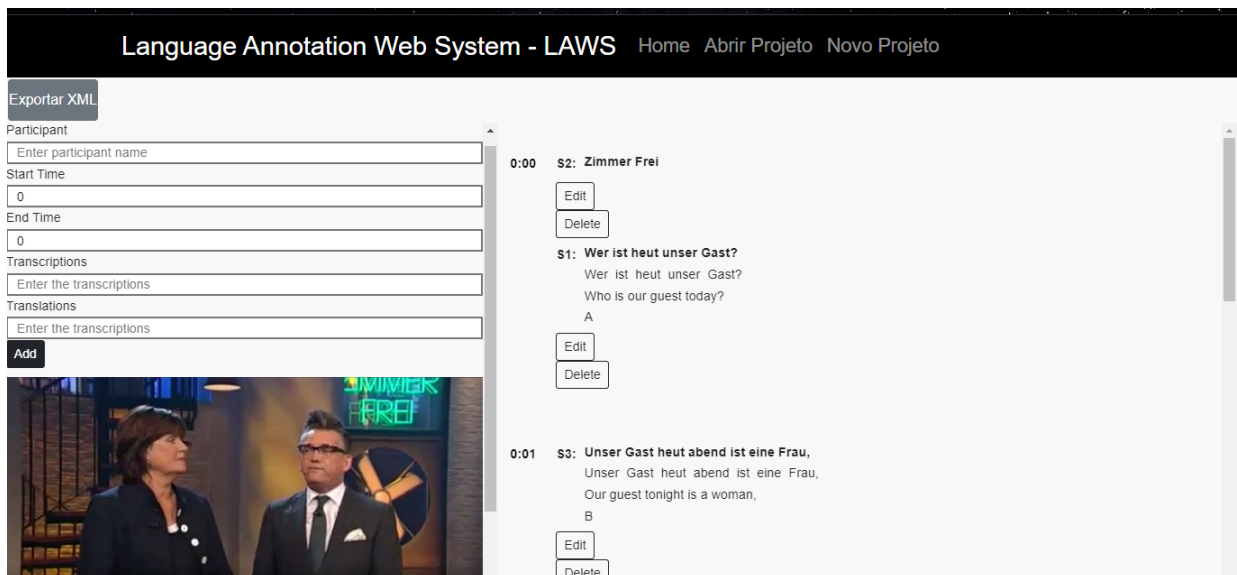


Figura 17 – Interface do LAWS

Fonte: Autoria própria

Conforme explicado no Capítulo 1, estávamos interessados na facilidade de interação do sistema. A ferramenta pode ser acessada no seguinte endereço <<https://marcosamaris.github.io/laws/#/>>.

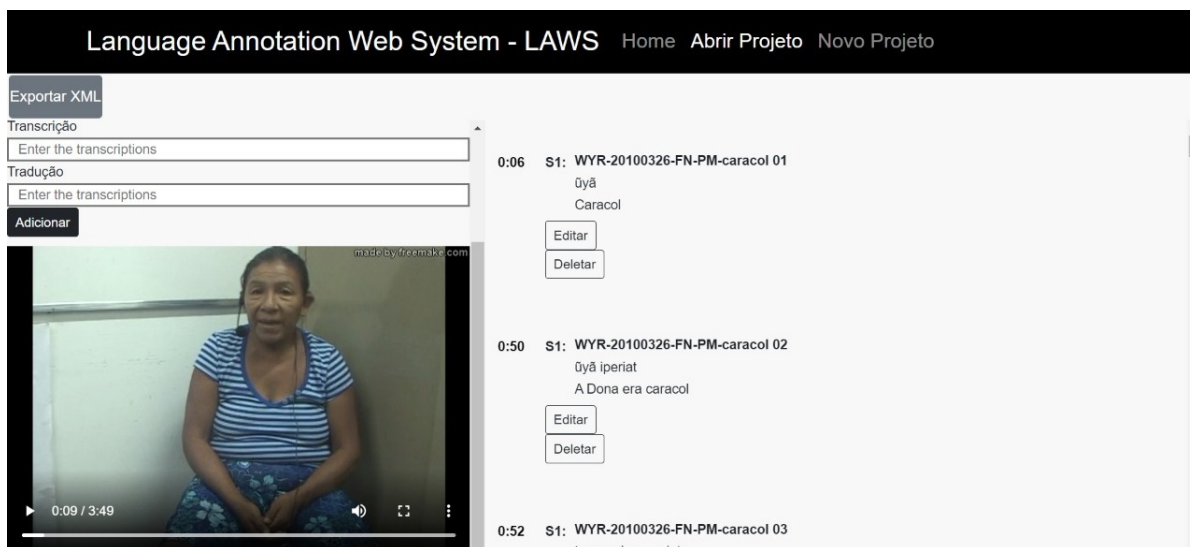
4.1 Experiências clássicas e uso de casos de LAWS com uma língua ameaçada, a língua da comunidade indígena Wajuru

Esta seção apresenta alguns resultados qualitativos de uma linguista que pesquisa uma língua ameaçada de extinção na Amazônia brasileira, a língua do Povo Wajuru.

O povo Wajuru, também denominado Wayoro, está localizado no estado de Rondônia(Brasil), especificamente na Terra Indígena (TI) Rio Guaporé, (município de Guajará-Mirim) e em Porto Rolim de Moura (município de Alta Floresta d'Oeste) (NOGUEIRA; MACURAP; WAJURU, 2019).

Nogueira, cf. Nogueira, Macurap e Wajuru (2019), faz, desde 2008, pesquisa linguística e documentação da língua Wayoro. Segundo a autora, existem apenas 02 falantes fluentes além de "onze semifalantes da língua Wayoro que vivem na Terra Indígena (TI) Rio Guaporé e em Porto Rolim de Moura". Uma das falantes fluentes é Paulina Makurap, de 63 anos de idade, que vive na TI Rio Guaporé.

Antonia Fernanda de Souza Nogueira fez o teste do LAWS, documentando a língua Wayoro, como mostrado na Figura 18.



The screenshot displays the LAWS interface. At the top, it says "Language Annotation Web System - LAWS" with navigation links for "Home", "Abrir Projeto", and "Novo Projeto". On the left, there are input fields for "Transcrição" and "Tradução", both with the placeholder text "Enter the transcriptions", and an "Adicionar" button. Below these is a video player showing a woman speaking. On the right, a list of annotations is shown, each with a timestamp, an ID, and the text in the indigenous language and Portuguese. Buttons for "Editar" and "Deletar" are provided for each annotation.

Timestamp	ID	Text	Actions
0:06	S1: WYR-20100326-FN-PM-caracol 01	Üyã Caracol	Editar Deletar
0:50	S1: WYR-20100326-FN-PM-caracol 02	Üyã iperiat A Dona era caracol	Editar Deletar
0:52	S1: WYR-20100326-FN-PM-caracol 03	tuarn nrlaran ariat	

Figura 18 – Teste do LAWS na língua da comunidade Wajuru

Fonte: Antonia Fernanda de Souza Nogueira

Os testes de Antonia Nogueira demonstraram que, utilizando o LAWS, a experiência de criação e modificação de projetos ELAN foi aprimorada. Muito menos tempo foi gasto fazendo a respectiva transcrição e tradução do material multimídia. O usuário indicou a facilidade de excluir e adicionar anotações mais rápido que o ELAN. Após esta fase de transcrição e tradução, a ferramenta ELAN pode ser usada para análises mais sofisticadas.

Dada a situação atual de pandemia, trabalhos de campo com comunidades indígenas tem sido adiados. No entanto, Nogueira planeja apresentar o LAWS para a comunidade Wajuru, assim que um projeto de pesquisa obtenha financiamento (uma submissão foi feita ao edital 2022 do Programa de Documentação de Línguas Ameaçadas, cf. <https://www.eldp.net/>). Nogueira e Silva (2021) informam que a escola da aldeia Ricardo Franco recebeu, em 2020, um ponto de internet de banda larga. Em um cenário ideal de documentação linguística, a comunidade realiza as gravações e elabora as anotações, com rotinas de backup do material em nuvem (BISCHOFF; JANY, 2018). O LAWS pode, assim, contribuir com esse cenário, fornecendo uma ferramenta de transcrição e tradução acessível aos indígenas.

5 CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS

Este capítulo apresenta as conclusões obtidas com o trabalho e as sugestões para trabalhos futuros que possam vir a contribuir tanto na melhoria deste processo como de outros trabalhos relacionados.

5.1 Conclusão e Sugestões para trabalhos futuros

Apresentamos um sistema web amigável para anotação linguística manual de arquivos audiovisuais. Esse sistema foi denominado LAWS - Language Annotation Web System. Usamos uma arquitetura web para disponibilizar a ferramenta na web. Todas as funcionalidades desenvolvidas são executadas no lado cliente, utilizando a linguagem JavaScript. Usamos o framework React, e outras bibliotecas como o Redux, BootStrap e o Webpack para modularização de códigos.

Códigos de bloco JavaScript são usados para importar especificações XML para arquivos JSON, que são usado para renderizar códigos de bloco HTML no navegador. Dessa forma, os usuários excluem e adicionam transcrição interativamente e anotações de tradução alinhadas ao tempo dos arquivos audiovisuais.

Este trabalho foi de grande aprendizagem para equipe, como conceitos de codificação. E isto, foi alcançado graças a dedicação dos participantes deste projeto. Além de possibilitar uma experiência com um produto final, tivemos a oportunidade de vivenciar os desafios dos estudantes ao se prepararem para um futuro no mercado de trabalho. E com isso, afirmamos que este projeto é de suma importância, pois trata-se de uma solução prática que será aplicado na sociedade.

Como trabalhos futuros, estamos interessados em desenvolver novas funcionalidades no sistema. Essas novas funcionalidades estão relacionadas à criação e exportação de um dicionário com os termos da camada de transcrição, que é uma camada intrínseca do sistema LAWS.

REFERÊNCIAS

- AGGARWAL, S. Modern web-development using reactjs. **International Journal of Recent Research Aspects**, v. 5, n. 1, p. 133–137, 2018.
- AUSTIN, P. K.; SALLABANK, J. **The Cambridge handbook of endangered languages**. [S.l.]: Cambridge University Press, 2011.
- BEREZ, A. L. Review of eudico linguistic annotator (elan). **Language Documentation & Conservation**, University of Hawai'i Press, v. 1, n. 2, 2007.
- BISCHOFF, S. T.; JANY, C. **Insights from Practices in Community-Based Research**. [S.l.]: De Gruyter Mouton, 2018.
- BRAATZ, D.; OLIVEIRA, K. d. S. de; ROCHA, T. R. da. O modelo de ciclo de vida iterativo/incremental para desenvolvimento de software. In: **7ª MOEPEX**. [S.l.: s.n.], 2018.
- BRUGMAN, H.; RUSSEL, A.; NIJMEGEN, X. Annotating multi-media/multi-modal resources with elan. In: **LREC**. [S.l.: s.n.], 2004. p. 2065–2068.
- CAMPETELA, C. et al. Documentação linguística, pesquisa e ensino: revitalização no contexto indígena do norte do amapá. **Revista Linguística**, v. 13, n. 1, p. 151–167, 2017.
- DOBRIN, L. M.; ROSS, D. The iath elan text-sync tool: A simple system for mobilizing elan transcripts on-or off-line. University of Hawaii Press, 2017.
- FLANAGAN, D. **JavaScript: o guia definitivo**. [S.l.]: Bookman Editora, 2004.
- GIPPERT, J. et al. **Essentials of language documentation**. [S.l.]: Mouton de Gruyter Berlin, 2006.
- GRIES, S. T.; BEREZ, A. L. Linguistic annotation in/for corpus linguistics. **Handbook of linguistic annotation**, Springer, p. 379–409, 2017.
- HELLWIG, B. **ELAN - Linguistic Annotator version 6.2**. <https://www.mpi.nl/corpus/manuals/manual-elan.pdf>, 2021.
- LIMA-PADOVANI, B.; MIRANDA, C. C.; BARROS, J. B. A importância da documentação e descrição linguística para revitalização de línguas ameaçadas. **Domínios de Lingu@ gem**, v. 13, n. 3, p. 900–926, 2019.
- NOGUEIRA, A. F. d. S.; SILVA, R. C. S. Linguistic annotation visualization of audio and/or video files: Elan, flex and lingview. **Cadernos de Linguística**, v. 2, n. 2, p. e369, Jul. 2021. Disponível em: <<https://cadernos.abralin.org/index.php/cadernos/article/view/369>>.
- NOGUEIRA, A. F. de S.; MACURAP, P.; WAJURU, J. Wayoro (wajuru)-tuero nderap eriat-a dona do preparo da chicha. **Revista Linguística**, v. 15, n. 1, p. 336–356, 2019.
- PRIDE, K.; TOMLIN, N.; ANDERBOIS, S. Lingview: A web interface for viewing flex and elan files. **Language Documentation & Conservation**, University of Hawaii Press, v. 14, p. 87–107, 2020.
- REIMAN, D. W. Basic oral language documentation. **Language Documentation & Conservation**, University of Hawai'i Press, v. 4, p. 254–268, 2010.

TAROCO, B.; WERNER, C. **Uma Abordagem sobre Modelos de Processos Ágeis**. [S.l.]: Universidade Paranaense (Unipar). Paraná, Brasil, 2007.

TILKOV, S.; VINOSKI, S. Node. js: Using javascript to build high-performance network programs. **IEEE Internet Computing**, IEEE, v. 14, n. 6, p. 80–83, 2010.

WITTENBURG, P. et al. Elan: A professional framework for multimodality research. In: **5th International Conference on Language Resources and Evaluation (LREC 2006)**. [S.l.: s.n.], 2006. p. 1556–1559.

Apêndices

APÊNDICE A – MANUAL DE USO DO LAWS

Apresentaremos agora, o manual de uso do LAWS. Que tem as seguintes funcionalidades: Novo Projeto, Abrir Projeto, Inserir anotação, Deletar, Editar e Exportar.

A.1 Abrir Projeto

Para abrir um projeto, o usuário terá que informar, como entrada o arquivo XML e o arquivo audiovisual, como mostrado na figura 19. Depois de carregado, clique em 'COMEÇAR'. Então o usuário já pode começar a edição.

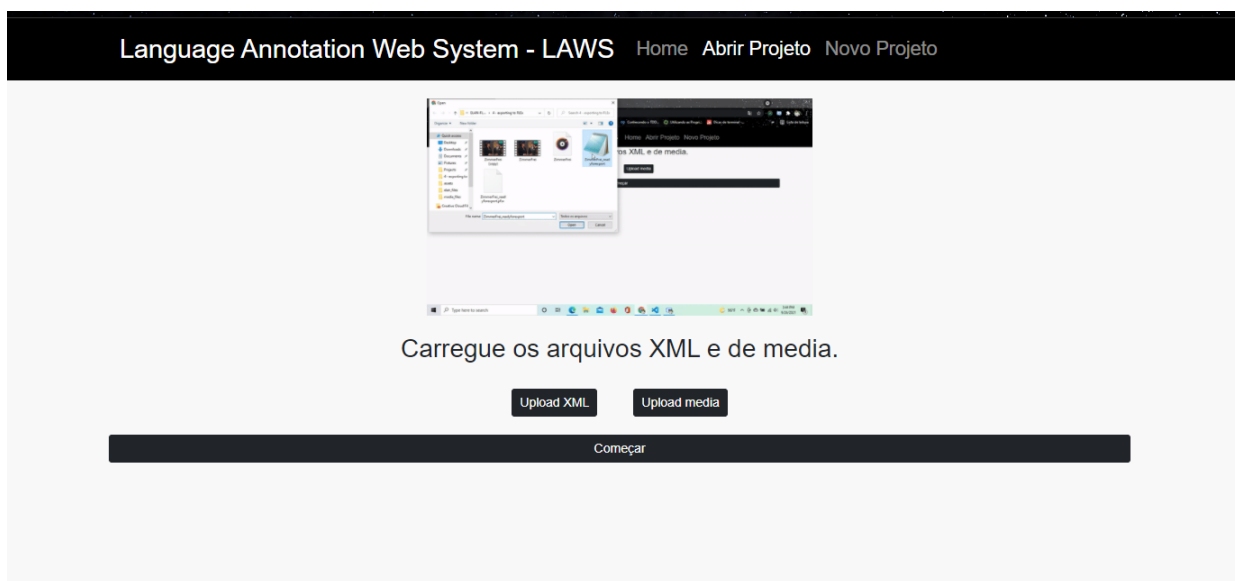


Figura 19 – Tela abrir projeto do LAWS

Fonte: Autoria própria

A.2 Novo Projeto

Para criar um novo projeto, é necessário o usuário informar um arquivo audiovisual, como mostrado na figura 20.

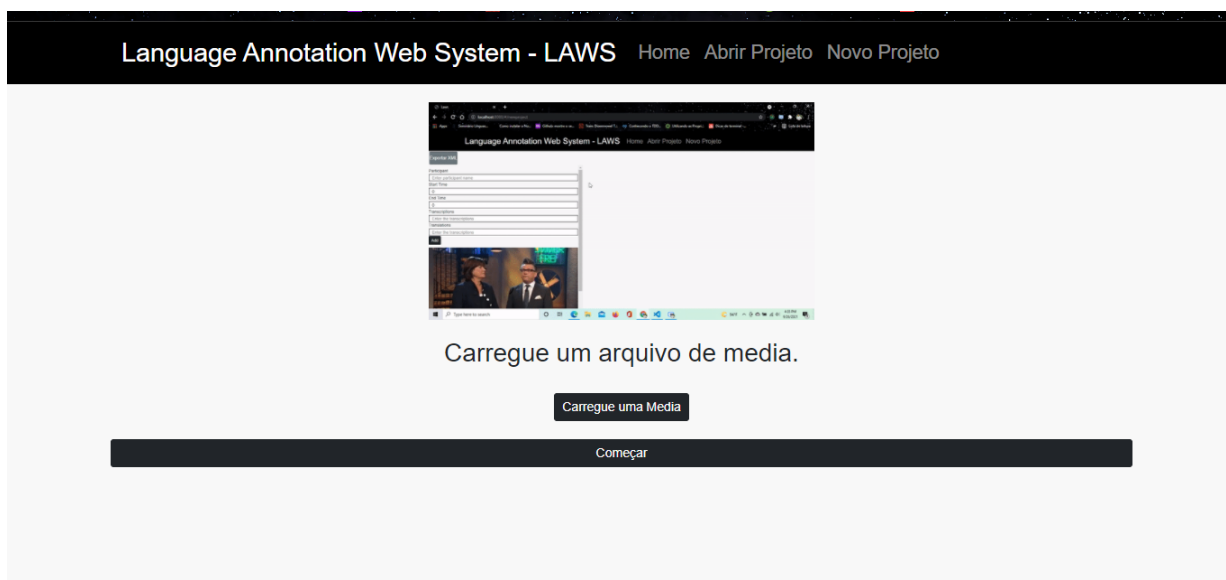


Figura 20 – Tela Novo projeto do LAWS

Fonte: Autoria própria

A.3 Inserir, Editar, Deletar e Exportar

No lado esquerdo, temos o botão 'EXPORTAR', que quando clicado gera e baixa um arquivo xml do tipo '.eaf'.

Abaixo do botão é exibido um formulário para adicionar anotações, como mostrado na figura 21. O usuário preenche todos os campos e depois, clica no botão 'Adicionar'.

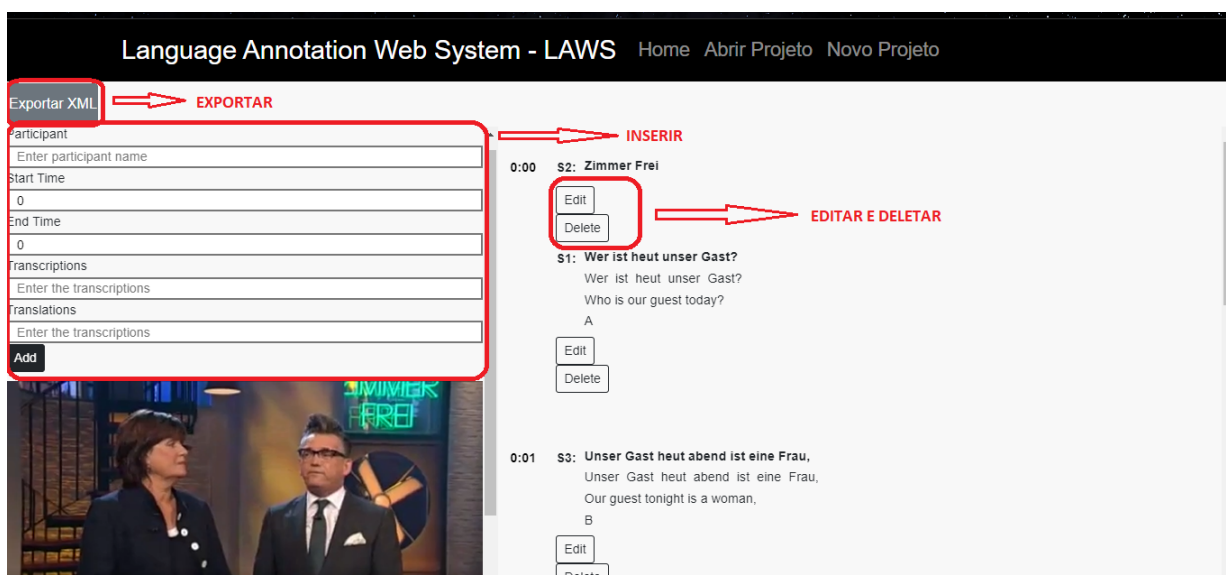
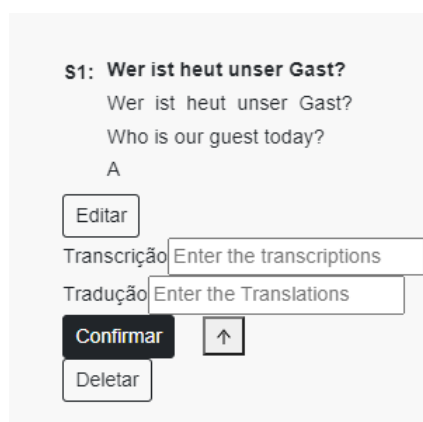


Figura 21 – Tela de edição

Fonte: Autoria própria

Para apagar anotação, o usuário clica no botão Deletar da anotação desejada. E da mesma forma para editar, o usuário clica no botão Editar, e será apresentada um formulário abaixo, que

contem os campos de transcrição e tradução. Depois de alterado, o usuário clica no botão Alterar para salvar as alterações, como mostrado na figura 22.



The screenshot displays a user interface for editing a transcription and translation. At the top, it shows a question in German: "S1: Wer ist heut unser Gast?" followed by its English translation: "Wer ist heut unser Gast?" and "Who is our guest today?". Below this, the letter "A" is visible. The interface includes several interactive elements: an "Editar" button, a text input field labeled "Transcrição" with the placeholder "Enter the transcriptions", another text input field labeled "Tradução" with the placeholder "Enter the Translations", a "Confirmar" button (highlighted in black), an upward-pointing arrow button, and a "Deletar" button.

Figura 22 – Editar e Deletar

Fonte: Autoria própria