

Estado da publicação: O preprint não foi publicado em outro meio.

Corpora anotados digitais de línguas indígenas brasileiras com traduções automáticas

Filomena Sandalo

<https://doi.org/10.1590/SciELOPreprints.13713>

Submetido em: 2025-10-13

Postado em: 2025-10-28 (versão 1)

(AAAA-MM-DD)

CORPORA ANOTADOS DIGITAIS DE LÍNGUAS INDÍGENAS BRASILEIRAS COM TRADUÇÕES AUTOMÁTICAS

DIGITAL ANNOTATED CORPORA OF BRAZILIAN INDIGENOUS LANGUAGES WITH AUTOMATIC TRANSLATIONS

Filomena Sandalo, UNICAMP & CNPq

<https://orcid.org/0000-0003-4595-7765>

RESUMO:

As línguas nativas do Brasil, já gravemente ameaçadas antes da pandemia, enfrentam uma ameaça crescente com a morte de anciãos nativos em decorrência da COVID-19. Nosso projeto visa promover a inovação digital inclusiva por meio do desenvolvimento de uma plataforma computacional para materiais em línguas indígenas que sejam cultural e gramaticalmente significativos. Esses materiais fornecerão uma base digital para o aprimoramento da educação bilíngue, das conexões intergeracionais e da transmissão de conhecimentos ancestrais nas comunidades indígenas

PALAVRAS-CHAVE: corpora, dicionários, anotações gramaticais, traduções automáticas, línguas indígenas, documentação

ABSTRACT:

The native languages of Brazil, already severely endangered before the pandemic, face increased menace from the death of native elders of COVID-19. Our project aims to advance inclusive digital innovation by developing a computational platform for indigenous language materials that are culturally and grammatically significant. These language materials will provide a digital foundation for enhanced bilingual education, intergenerational connections, and the transmission of ancestral knowledge in indigenous communities.

KEY WORDS: corpora, dictionaries, grammatical annotation, automatic translations, indigenous languages, documentation

Conflito de interesse: Declaro que os dados deste trabalho foram coletados por professores indígenas falantes das línguas ou são dados publicados na literatura, não exigindo parecer de comissão de ética. Declaro ainda que não tenho conflito de interesse com as fontes citadas aqui.

Declaração de Disponibilidade de Dados: Todos os dados estão disponíveis em

<https://www.tycho.iel.unicamp.br/catalog/C12>.

Introdução

Esse artigo apresenta os esforços em documentação linguística e cultural em realização com apoio da FAPESP (Processo 22/09158-5) através do projeto “Corpora anotados digitais de línguas indígenas brasileiras com traduções automáticas (DACILAT)”. E as atividades dentro deste projeto indicaram uma necessidade de organizar uma publicação que pudesse apresentar diferentes esforços no Brasil no trabalho de documentação e fortalecimento de suas línguas nativas, fortemente ameaçadas atualmente. Assim nasceu este dossiê.

Neste capítulo apresentamos o projeto DACILAT, que se baseia em uma concepção de construção de corpus e de tradução automática diferente das correntes principais de abordagens computacionais que partem de uma enorme quantidade de dados (*big data*). Para línguas em perigo de extinção, em particular, grandes corpora não podem ser constituídos.

Nosso trabalho é baseado em elaboração de corpora anotados automaticamente por um Parser sintático, elaborado pelo projeto, que garantem dados acrescidos automaticamente de informações gramaticais que abrem a possibilidade de novas pesquisas sobre as línguas de línguas documentadas e também materiais para ensino de língua e gramática nas escolas indígenas. Não usamos os métodos probabilísticos da IA e trabalhamos na elaboração de um método alternativo que usa um Parser de regras baseado na gramática gerativa chomskyana e na linguagem Corpus Search (<https://corpussearch.sourceforge.net>).

O projeto documenta Kadiwéu e Nheengatu, mas outras línguas estão em início de documentação, como o Tukano e o Sateré-Mawe. O corpus mais avançado neste momento é o do Kadiwéu, e, portanto, este artigo se concentra no corpus desta língua originária para demonstrar como nossa documentação é elaborada. Os dados do Kadiwéu estão sendo

coletados na aldeia Alves de Barros, Mato Grosso do Sul por falantes nativos de Kadiwéu e traduzidos por professores indígenas que fazem parte deste projeto. Kadiwéu é uma língua da família Guaikurú falada por, segundo Pires 2022, menos de 500 pessoas (as comunidades contam também por volta de 1000 pessoas que falam apenas português). Portanto é uma língua em perigo de grande enfraquecimento ou extinção.

Nosso corpora é abrigado na Plataforma Tycho Brahe. Esta plataforma é um conjunto de ferramentas totalmente online, e é pioneira em sua aplicação para as línguas originárias da América do Sul. A sua interface baseada na web permite a disseminação imediata para as comunidades indígenas e para outros pesquisadores de linguística ou antropologia que queiram trabalhar com dados de linguagem, além de permitir a criação de corpora anotados gramaticalmente significativamente maiores do que era possível anteriormente.

Entre essas ferramentas está o Parser sintático mencionado: além de fazer traduções de palavras e anotações automáticas, o Parser elabora automaticamente as análises sintáticas das sentenças nos corpora, base para a nossa proposta de tradução automática.

As ferramentas de busca, assim como os corpora, são abertas e é possível fazer buscas por palavras ou por etiquetas sintáticas, já que o corpus é anotado sintaticamente.

Os corpora geram dicionários online. O dicionário é alimentado por palavras dos corpora, dicionários das línguas publicados ou não e de dados coletados em trabalho de campo. Contém informações de POS e paradigmas morfológicos.

Resta ainda mencionar que os tamanhos dos corpora e dicionários são variáveis, pois são corpora abertos. Quaisquer novos dados podem ser alimentados no corpora e serão automaticamente anotados, e, por sua vez, alimentará o dicionário com cada nova palavra que ocorrer. Tudo de maneira online. Enfim, é um trabalho de documentação que não se esgota e cujas informações gramaticais alimentam novos trabalhos sobre as línguas. Além disso, o corpus Kadiwéu, por ser formado por narrativas originárias tem potencial de colaborar com

estudo etnológico/antropológico e histórico do Kadiwéu e permitir comparações linguísticas e mitológicas na região do Chaco. Os corpus e o dicionário tem alimentado pesquisas/publicações sobre a língua e o ensino da língua, podendo dar subsídios para discutir a análise sintática das línguas, sendo uma importante ferramenta de ensino de gramática nas escolas na comunidade Kadiwéu. Portanto, nosso trabalho tem implicações científicas e sociais.

Há várias outras plataformas para depositar corpora de línguas indígenas, mas salientamos aqui que nenhuma conta com um Parser capaz de analisar sintaticamente as línguas podendo servir no futuro de uma base de revitalização linguística, uma vez que as anotações gramaticais oferecem uma visão ampla da gramática das línguas em suas dimensões morfológicas e gramaticais. E todas as informações estão abertas pela internet podendo ser acessadas de qualquer parte do mundo.

Segue a página do projeto DACILAT para visitaç o: <https://www.tycho.iel.unicamp.br/dacilat>.

1. A Plataforma Tycho Brahe

A *Plataforma Tycho Brahe*, como já mencionado, é o um conjunto de ferramentas **totalmente online**. A Plataforma Tycho Brahe (<https://www.tycho.iel.unicamp.br/home>) foi desenvolvida pelo aluno de doutorado Luiz Henrique Lima Veronesi, sob supervisão de Charlotte Galves, na Universidade Estadual de Campinas. Abaixo apresentamos as ferramentas que constituem a Plataforma Tycho Brahe, as abertas para o público e as fechadas:

1.1. As ferramentas abertas ao público

1.1.1. O Visualizador

A ferramenta Visualizador permite que qualquer pessoa usando a internet possa visitar e baixar dados de quaisquer de nossos corpora. A Plataforma Tycho é uma Biblioteca Digital:

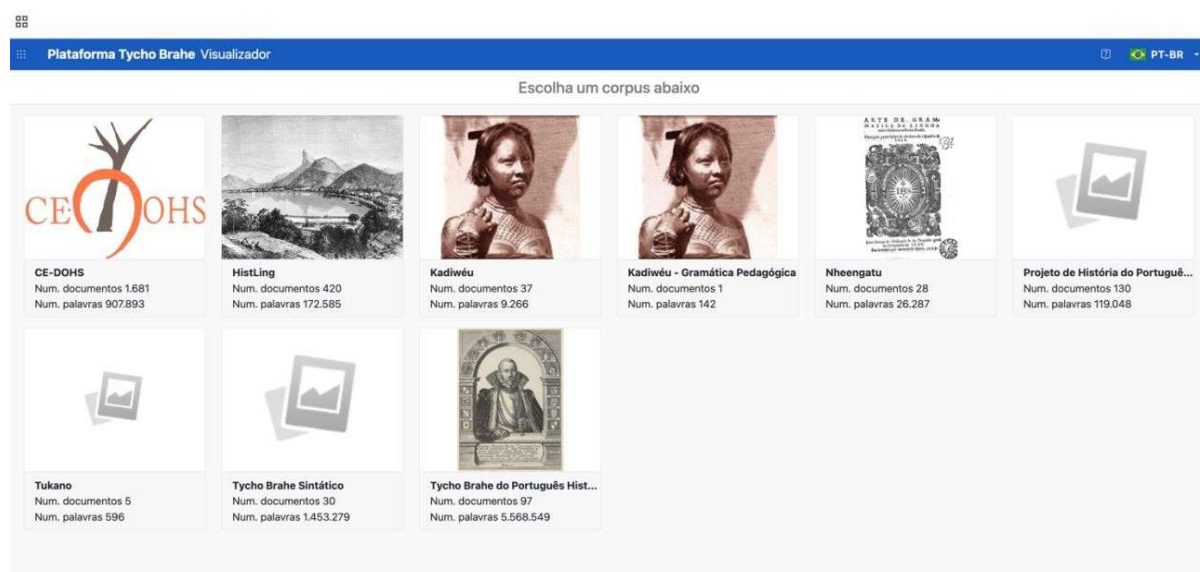


Figura 1: Entrada do Visualizador (ferramenta para visitantes aos corpora)

Como mencionado, concentraremos no kadiwéu. O corpus Kadiwéu é um dos corpora abrangidos na Plataforma Tycho Brahe. Trata-se de um corpus totalmente formado por narrativas originárias coletadas em campo e escritas usando a ortografia da língua. A Plataforma conta com um teclado virtual baseado no teclado Linklado (<https://www.linklado.com>) para permitir escrever nas línguas do Brasil. As narrativas do kadiwéu são de vários gêneros: narrativas mitológicas, narrativas histórias e cantos (choro Kadiwéu). O choro é um importante gênero ritualístico da arte ameríndia brasileira (Urban 1988). O choro Kadiwéu são cantos contendo poemas na língua, sendo um gênero que irá trazer um desafio especial para a tradução. Nossa tradução baseada em um Parser sintático

pretende ocorrer com qualquer gênero e qualquer assunto, como será demonstrado mais adiante. A página de entrada do corpus Kadiwéu segue abaixo e cada figura histórica da arte Kadiwéu é a entrada de uma narrativa distinta (Figura 2):

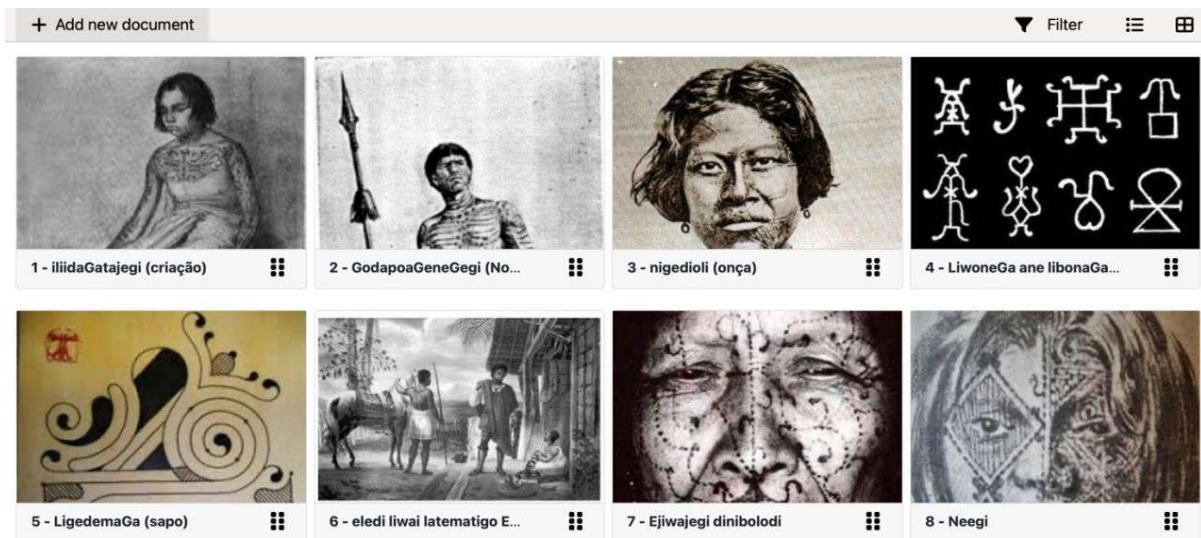


Figura 2: Entrada do Corpus Kadiwéu

Ao clicarmos na figura de entrada de cada narrativa, a narrativa abre-se em kadiwéu com uma tradução dada por um professor indígena. O áudio pode ser baixado pelos usuários:

1	natigide jaaga jatematiga natematigigi ica maleeka iliidagadi Aneotedogoji oko .	Agora vamos contar a história da criação do homem por Deus.	▶ 🔊
2	ica jotigide Aneotedogoji maleka iliidagiteda ina oko .	Antigamente Deus criou esta gente	▶ 🔊
3	one enidite etolitoli	Dizem que chamou caburé	▶ 🔊
4	oniketa etolitoli : owieni iawate	diz que ele disse para ele: cuide de minha neta	▶ 🔊
5	onete damiida ica inoenegegi ?	Ele disse: qual sera o meu canto?	▶ 🔊
6	onete idagida idoda gadaxacenegegi .	ele disse : esse mesmo, esse seu cantar simples compassado	▶ 🔊
7	oda jonagidagee ica enoale jonagidagee etolitoli eotibige ica layagigegi me anoe .	Então esta noite o caburé ficou fazendo seu barulho que era gritado	▶ 🔊
8	nigatibece ica ane wajibata .	às vezes ele escutava algo	▶ 🔊

Figura 3: Narrativas

E ao clicar em uma dada sentença as anotações gramaticais podem ser visualizadas. É importante mencionar que a anotação dos corpora segue a necessidade de seus criadores e é totalmente parametrizada desde que contenham etiquetas POS para alimentar o Parser

sintático, que será apresentado mais adiante. A anotação do Corpus Kadiwéu conta com etiquetas POS, morfológicas e traduções de palavras e morfemas, além de árvores sintáticas e áudios. Os áudios são acrescentados manualmente através da Ferramenta Edictor, que será apresentada mais adiante. As anotações gramaticais são automáticas depois de o Parser ser elaborado:

The screenshot displays the Edictor software interface for the Kadiwéu sentence "one@ @likeeta etoitoli : owieni iwawte". The interface is divided into several sections:

- Annotations Table:** A table with columns for morphemes and their corresponding annotations. The morphemes are "one", "likeeta", "ta", "etoitoli", ":", "owieni", "iwawte". The annotations include POS tags (e.g., EV, VBAPL, N, PUNC, VBI, NS), glosses (e.g., "show to", "caburé", "take care of", "my granddaughter"), and morphological information (e.g., "1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25", "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39", "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53", "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67", "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81", "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95", "96", "97", "98", "99", "100").
- Audio Player:** A section with a waveform and playback controls (play, stop, volume).
- Translations:** A section with input fields for Portuguese and English translations. The Portuguese translation is "diz que ele disse para ele: cuide de minha neta" and the English translation is "my granddaughter".
- Syntactic Tree:** A tree diagram showing the hierarchical structure of the sentence. The root node is IP-MAT, which branches into EV, VBAPL, NP-APL, PUNC, and IP-APL. The EV node branches into one@ (show to / mostrou para). The VBAPL node branches into @likeeta. The NP-APL node branches into etoitoli (caburé / caburé). The PUNC node branches into :. The IP-APL node branches into VBI (owieni / take care of / cuide de) and NP (iwawte / my granddaughter / minha neta).

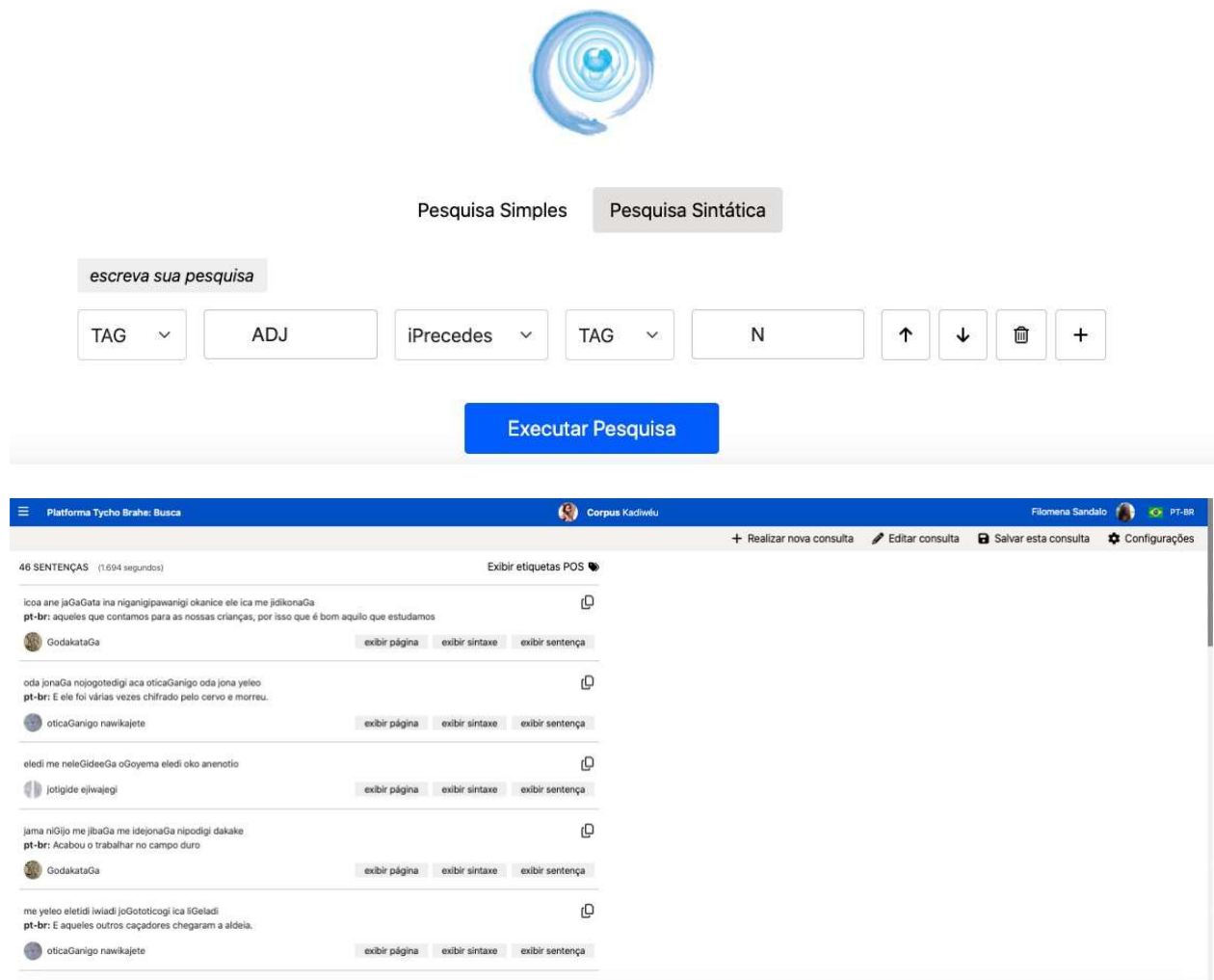
Figura 4: anotações gramaticais e áudio do Corpus Kadiwéu

Na seção sobre o Parser, a anotação será mais discutida. A análise morfológica é feita automaticamente pelo Parser, mas dado que a língua Kadiwéu é uma língua de grande complexidade morfológica, nem todas as palavras já contam com sua divisão morfológica feita automaticamente. Neste momento, ainda entramos análise manualmente através da ferramenta Edictor. Também é através do Edictor que alimentamos novas narrativas ao corpus. Trata-se de um corpus aberto que recebe novos materiais constantemente.

A Ferramenta Visualizador simplesmente permite a qualquer visitante explorar os corpora abertos abrigados na Plataforma que contem com permissão da comunidade.

2.1.2. A Ferramenta de Buscas

As buscas nos corpora são também abertas para quaisquer visitantes fomentando novas pesquisas sobre as línguas. É possível fazer buscas por palavras ou por etiquetas sintáticas, já que o corpus é anotado sintaticamente. Segue uma imagem de um exemplo de busca sintática, que são feitas usando a linguagem Corpus Search. Para facilitar para usuários não familiarizados com essa linguagem computacional, há caixas de escolha para cada alternativa:



The screenshot displays the search interface for the Corpus Kadiwéu. At the top, there is a logo and two search options: 'Pesquisa Simples' and 'Pesquisa Sintática'. Below this is a search bar with the placeholder text 'escreva sua pesquisa'. To the right of the search bar are several controls: a 'TAG' dropdown menu, an 'ADJ' button, an 'iPrecedes' dropdown menu, another 'TAG' dropdown menu, an 'N' button, and four navigation buttons (up, down, trash, and plus). A large blue button labeled 'Executar Pesquisa' is positioned below the search bar.

The search results are shown in a table with the following columns: the original Kadiwéu text, a Portuguese translation, and a list of morphemes with their part-of-speech tags. The results are as follows:

Original Text	Portuguese Translation	Morphemes and POS
icoa ane jaGaGata ina niganigipawanigi okanice ele ica me jidikonaGa	pt-br: aqueles que contamos para as nossas crianças, por isso que é bom aquilo que estudamos	GodakataGa (exibir página, exibir sintaxe, exibir sentença)
oda jonaGa nojogotedigi aca oticaGanigo oda jona yeleo	pt-br: E ele foi várias vezes chifrado pelo cervo e morreu.	oticaGanigo nawikajete (exibir página, exibir sintaxe, exibir sentença)
eledi me neleGideeGa oGoyemaeledi oko anenotio		jotigide ejiwajegi (exibir página, exibir sintaxe, exibir sentença)
jama ni'oijo me jibaGa me idejonaGa nipodigi dakake	pt-br: Acabou o trabalhar no campo duro	GodakataGa (exibir página, exibir sintaxe, exibir sentença)
me yeleo eletidi iwidi joGototocigi ica iGeladi	pt-br: E aqueles outros caçadores chegaram a aldeia.	oticaGanigo nawikajete (exibir página, exibir sintaxe, exibir sentença)

Figura 5: Buscas Sintáticas

O trabalho com o Corpus Kadiwéu vem evidenciando novas estruturas da língua e as buscas permitem encontrar todas suas ocorrências. E assim novos artigos sobre a gramática desta língua têm sido publicados (cf. Sandalo 2023, Sandalo & Galves 2025). Por exemplo, Sandalo (2023) avança no estudo de morfemas evidenciais da língua, e Sandalo & Galves (2025) sobre finitude e estratégias de subordinação da língua. Em um novo estudo das mesmas autoras, as anotações das narrativas vêm ajudando e discriminar tipos de aplicativos no kadiwéu (cf. Sandalo & Galves 2025b ms.). Verbos aplicativos podem selecionar

argumentos nominais (argumentos indiretos) ou sentenças na língua. A anotação sintática consegue discriminar entre estes argumentos pelas etiquetas automáticas NP-APL e IP-APL respectivamente:

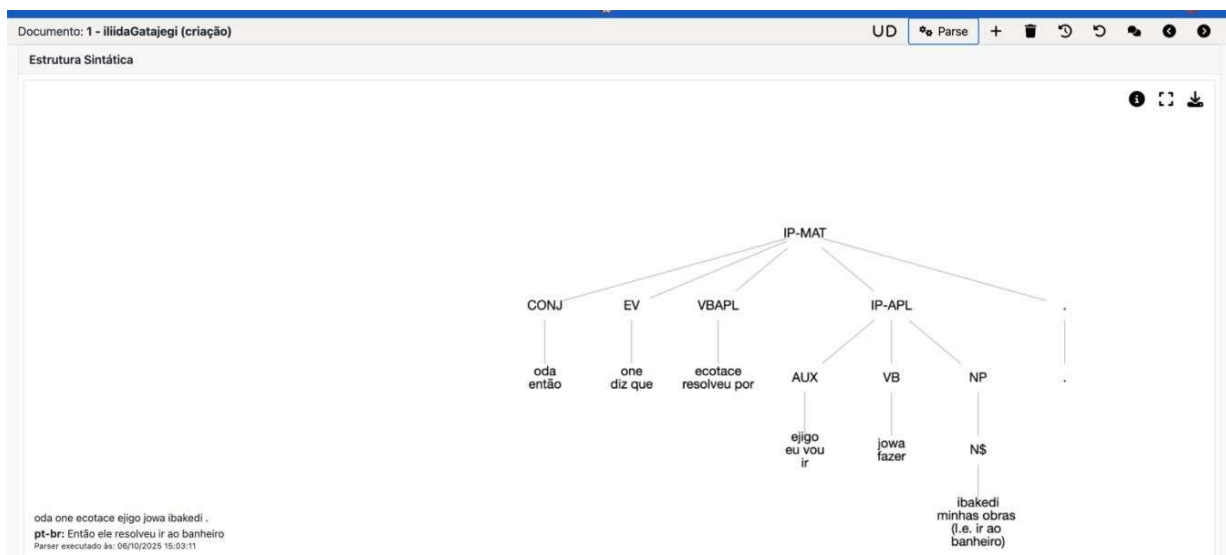


Figura 6: Aplicativos do kadiwéu

1.1.1. Dicionários (Léxicos)

Geramos ainda dicionários das línguas dos corpora automaticamente. Os dicionários são cruciais para as escolas indígenas. E são também visitados livremente. Os dicionários podem também ser encontrado na entrada da página da Plataforma.

O dicionário Kadiwéu conta com palavras como entradas e estas palavras são anotadas com etiquetas POS e informações gramaticais. Verbos e nomes contam ainda com seus paradigmas:

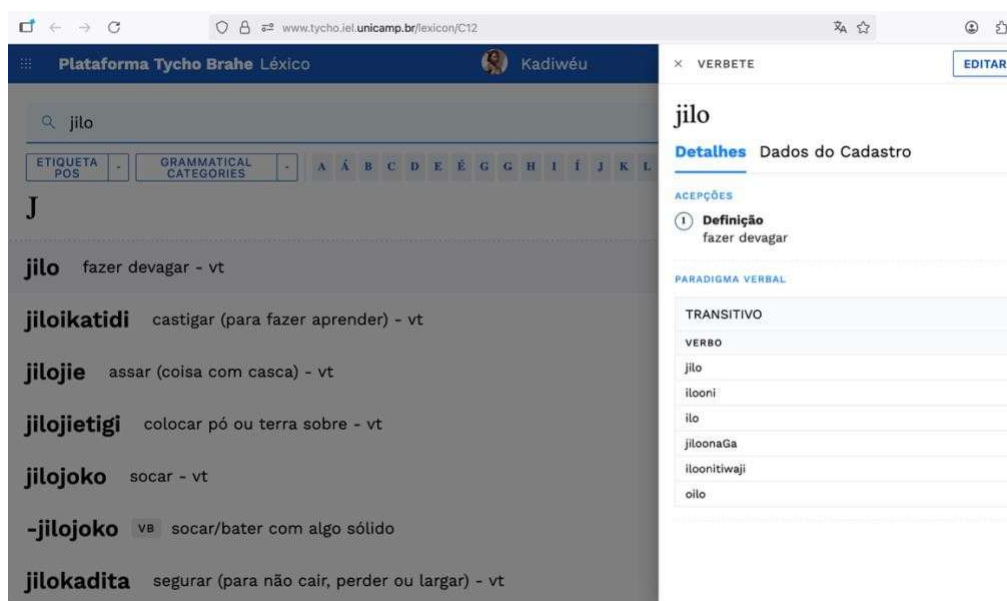


Figura 7: Dicionário Kadiwéu

Os materiais que alimentam os dicionários são as palavras das próprias narrativas, mas também dados coletados nas comunidades indígenas e em publicações sobre a línguas. Além disso, os dados do dicionário de Griffiths (2002) também foram alimentados ao dicionário. As fontes dos dados são discriminadas dentro do dicionário. Atualmente o dicionário Kadiwéu conta com 7058 palavras.

2.1.2. A Ferramenta Syntree

Esta ferramenta permite ao visitante elaborar automaticamente análises sintáticas de quaisquer sentenças no corpora ou em sentenças criadas pelo próprio falante usuário, por exemplo um professor indígena, colaborando, portanto, na educação escolar indígena. O professor indígena pode criar uma frase em frente aos alunos e o analisador sintático (Parser) fará sua análise, que poderá ser discutida e comparada com estruturas do português. Além da visualização da estrutura gerada, é possível fazer o download de imagens das sentenças.

Segue abaixo uma imagem de uma sentença que não está no corpus para exemplificar. A Plataforma contém a ortografia do Linklado (<https://www.linklado.com>) permitindo escrever em qualquer língua brasileira. Observe que algumas palavras não puderam ser traduzidas pelo Parser, mas mesmo assim a estrutura sintática foi gerada corretamente.

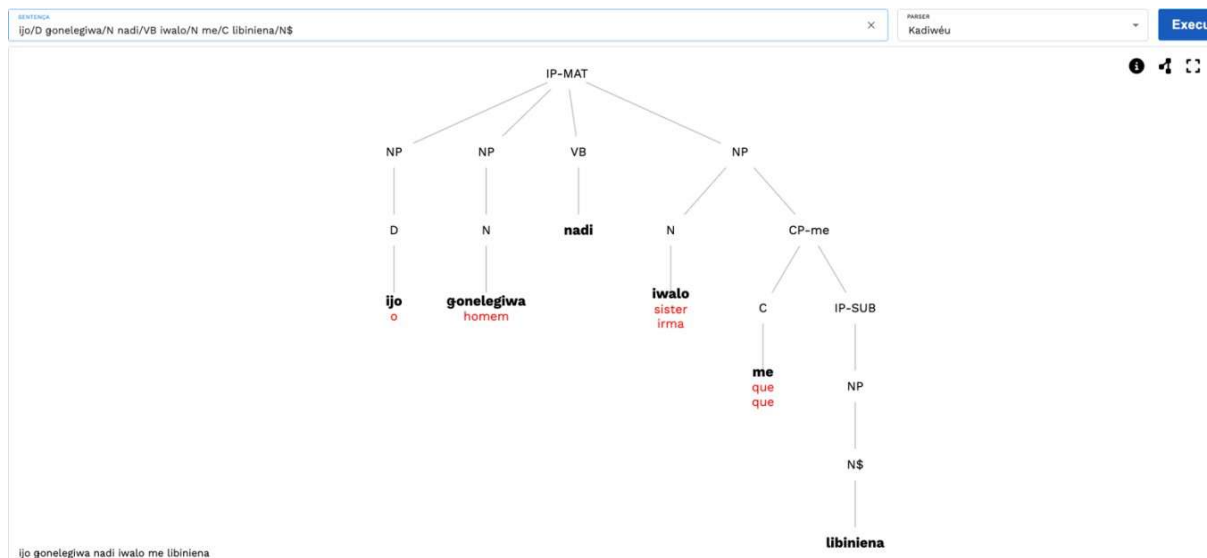


Figura 7: Parser de uma sentença não presente no corpus

1.2. A Plataforma Tycho Brahe: Ferramentas Fechadas (área reservada na página da plataforma)

1.2.1. O Parser

O Parser é uma ferramenta crucial neste projeto de documentação, e original para línguas indígenas: além de fazer traduções de palavras e anotações automáticas palavra por

palavra e morfema por morfema, ele elabora automaticamente as análises sintáticas das sentenças nos corpora, base para a nossa proposta de tradução automática e de educação linguística. Uma vez que a sentença é traduzida palavra por palavra e as palavras são organizadas sintaticamente, a tradução para o português está feita. Nosso parser é um parser de regras não probabilístico e, como já mencionado, toma como modelo o parser elaborado por Beatrice Santorini na Universidade da Pensilvânia para o francês antigo e adaptado ao espanhol e ao português por Catarina Magro do Centro de Linguística da Universidade de Lisboa (Magro & Galves 2019). É baseado na linguagem Corpus Search e foi adaptado neste projeto para o Kadiwéu (cf. Galves et al 2017, Sandalo & Galves 2023). O Parser Kadiwéu conta com 148 regras sintáticas atualmente:

NOME	REGRAS	ETIQUETAS POS	ETIQUETAS DE MORFEMA	ETIQUETAS SINTÁTICAS
Kadiwéu	148	157	58	24
Nheengatu	139	107	51	111
Portuguese	209	295	0	111
Sateré-mawé	139	107	51	111

Figura 8: Regras dos Parsers

A imagem abaixo mostra uma sentença traduzida pelo Parser Kadiwéu:

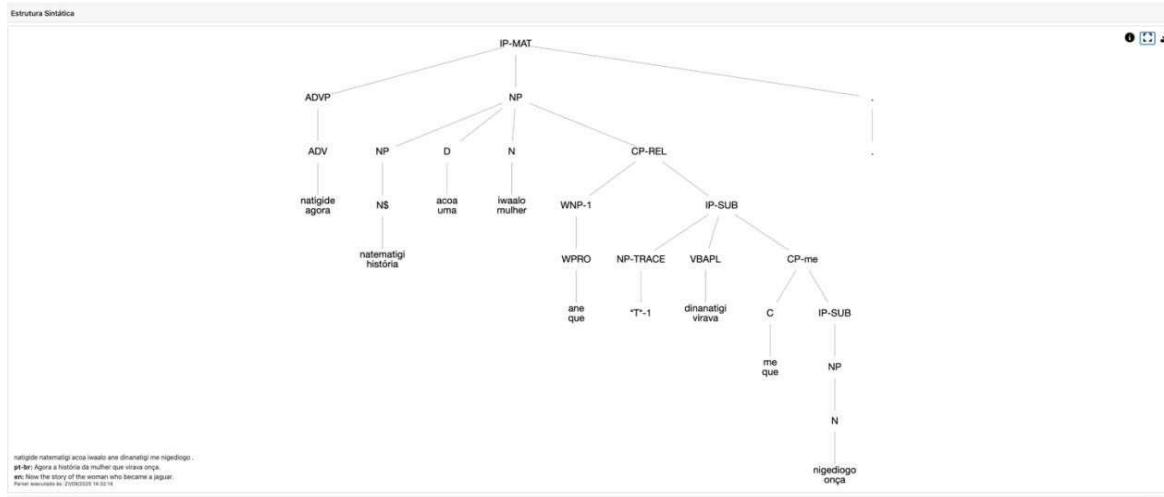


Figura 9: Tradução e análise pelo Parser

Nosso parser é baseado em uma concepção de gramática universal, portanto é facilmente adaptado para novas línguas, bastando acrescentar regras particulares e desativar regras não ativas na língua em questão. Segue abaixo uma regra do kadiwéu para exemplificação:

☰
49 cp-me1
⊙

Nome	Nó
<input type="text" value="cp-me1"/>	<input type="text" value="\$METAROOT"/>

Query

```

query: (IP-MAT iDoms {1}C)
      AND (C iDoms me*|me@)
      AND (C iPrecedes {2}ADJP|ADVP|NP)

add_internal_node{1, 2}: CP-me
                    
```

Vezes para executar - 1 +

Remove
Salvar

Figura 10: Uma Regra do Parser Kadiwéu

1.2.2. O Edictor

O Edictor é nossa ferramenta de edição manual, que permite a entrada de narrativas e correção manual de qualquer erro gerado pelo Parser.

O Parser do kadiwéu está pronto e a anotação é automática, mas alguns erros são gerados e corrigidos manualmente. Toda narrativa é também entrada através do Edictor, bem como seus áudios.

Documento: B.10 - eeyomoda

UD Parse +

nige dagadiagica degeyomoditaga jotatibige me anoo ,

palavra	nige	dagadiagica	degeyomoditaga	jotatibige	me	anoo	,
etiqueta POS	ADV	NEG+ADV+NEG+D	vazio	VBAPL	C	VB	PUNC
gloss-br	quando	não estiver mais aquela	vazio	falar de (alguém não presente)	que	chegar	vazio
gloss	vazio	vazio	vazio	vazio	vazio	arrive here	vazio
morfemas							
etiqueta							
gloss-br							
gloss							
	✓	✓	✓	✓	✓	✓	✓

Figura 10: Correção Manual pelo Edictor

2. Considerações Finais

Nossa proposta é inovadora tanto em organizar dados como em providenciar análises e traduções das línguas nos corpora. Todo o material de análise é gerado automaticamente e qualquer pessoa pode criar análises das sentenças no corpora ou de quaisquer sentenças da línguas que contam com o Parser desenvolvido. Cada língua nova entrada necessita de um tempo de trabalho manual para etiquetar algumas narrativas e para o desenvolvimento do Parser. Dado que partimos de uma concepção gerativa de linguagem, como mencionamos,

em que acreditamos que parte das gramáticas seguem princípios universais, o trabalho é acelerado em se resumir em adaptar os Parsers já prontos de acordo com a família e tipologia linguística. Este texto exemplificou nosso trabalho com base em uma língua, mas a Plataforma é um repositório capaz de ser uma biblioteca que analisa e traduz materiais de quaisquer tipologias linguísticas dado que temos linguistas e falantes da língua trabalhando nas adaptações necessárias. Não é de nosso conhecimento nenhum outro projeto de documentação com tamanho alcance. E totalmente desenvolvido no Brasil por uma equipe da Universidade Estadual de Campinas (UNICAMP).

Todos nossos dados (com transcrições e áudios) e anotações estão disponíveis na Plataforma Tycho Brahe. Nada no corpus, até o momento está fechado, e pode ser visitado online e usado para outras pesquisas ou para educação indígena livremente. Obviamente as comunidades indígenas que estão colaborando com o projeto podem pedir para fechar determinado assunto e isso será feito.

Referências:

FARIA, P., GALVES, C., MAGRO, C. (2023) Syntactic annotation for Portuguese corpora: standards, parsers, and search interfaces. *Language Resources and Evaluation, Especial Issue on Computational Approaches to Portuguese*.

GALVES, C., SANDALO, F., SENA, T.A., VERONESI, L. (2017) Annotating a polysynthetic language: From Portuguese to Kadiwéu. *Cadernos de Estudos da Linguagem*, (59.3), pp. 631-648.

GRIFFITHS, G. (2002). Dicionário da Língua Kadiwéu. SIL ms. <https://www.sil.org/system/files/reapdata/74/06/08/74060839706011162756896570533590209458/KDDict.pdf>.

MAGRO, C., GALVES, C. (2019) Portuguese Syntactic Annotation Manual.
<http://alfelul.clul.ul.pt/portuguesesyntacticannotation/>

SANDALO, Filomena. (2023). Evidencialidade Reportativa, Tempo e Negação em kadiwéu. *Liames* 23(1) <https://doi.org/10.20396/liames.v23i00.8671197>

SANDALO, Filomena & GALVES, Charlotte. 2025. A brief description and analysis of tense, finiteness, and verbal embedding strategies in Kadiwéu. *Cadernos de Estudos Linguísticos*. (a sair)

SANDALO, Filomena; GALVES, Charlotte Marie Chambelland. (2023). Anotando sintaticamente uma língua originária do Brasil: o problema de Anchieta. *Cadernos de Estudos Linguísticos*, v. 65, p. e023007-e023007.

SANTORINI, B. (2022) Annotation manual for the Penn Historical Corpora and the York-Helsinki Corpus of Early English Correspondence <https://www.ling.upenn.edu/hist-corpora/annotation/index.html>

URBAN, Greg. (1998). *Ritual Wailing in Amerindian Brazil*.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.