

**CURSO**

# Corpora Online Anotadas de línguas Originárias/Uma nova proposta acessível de documentação linguística e cultural

---

**MINISTRANTES**    **Profa. Dra. Maria Filomena S Sandalo**  
Universidade Estadual de Campinas  
[Lattes](#)

**Profa. Dra. Charlotte M. C. Galves**  
Universidade Estadual de Campinas  
[Lattes](#)

---

**LOCAL**                      Plataforma Abralin EaD  
**PERÍODO**                    Agosto, 9h às 12h

---

**CARGA**                      12h/a  
**HORÁRIA**

---

**IDIOMA**                    Português

---

Alunos de graduação;  
Alunos de pós-graduação;  
Professores do ensino fundamental e médio;  
Professores do ensino superior.

Sem avaliação

## **AVALIAÇÃO**

## **EMENTA**

Este trabalho se baseia em uma concepção de construção de corpus e de tradução automática diferente do mainstream de abordagens computacionais que partem de uma enorme quantidade de dados (big data). Assumimos uma posição oposta que contesta a obrigatoriedade de 'big data' para pesquisa e tradução automática de línguas com poucos materiais disponíveis. Para línguas em perigo de extinção, em particular, grandes corpora não podem ser constituídos, e isso por si só as exclui desse tipo de projetos. Defendemos que a anotação acrescida aos textos é que os torna relevantes para pesquisas e aplicações para fins educativos e sociais, adicionada a uma parser por regras. Afirmamos ainda que grandes corpora são inúteis se não contiverem informações extras que permitam aos pesquisadores recuperar dados para responder às suas perguntas sobre a linguagem e sua dinâmica. Essas informações extras são adicionadas por meio de anotação e a anotação é baseada em modelos linguísticos. Em suma, este curso trará uma proposta de documentação linguística e cultural baseada em corpora anotados sintaticamente e morfologicamente e elaboração de uma parser sintático para garantir o trabalho em documentação de línguas em perigo de extinção, bem como para o trabalho de tradução automática destas línguas. Não usamos os métodos probabilísticos da IA e trabalhamos na elaboração de um método alternativo para línguas nativas das Américas através de um parser baseado na gramática gerativa chomskyana. O material produzido é totalmente online e acessível para comunidades indígenas.

## **OBJETIVOS**

1. ensinar elaborar corpora anotados de línguas originárias na Plataforma Tycho Brahe.
2. Ensinar sobre o uso de um parser sintático
3. Produção de traduções automáticas e análises sintáticas automáticas.

A Plataforma Tycho Brahe é o nosso 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. A plataforma contempla edição de textos, parser sintático, sistemas de buscas lexicais e sintáticas, e dicionários, uma vez que um corpus é desenvolvido.

---

## **PROGRAMA**

Quer saber mais sobre a documentação de línguas indígenas? Então, este curso é para você! A plataforma Tycho Brahe, desenvolvida para linguistas e pesquisadores, é uma ferramenta fundamental na preservação e estudo de línguas indígenas. Este curso apresentará como a plataforma Tycho Brahe pode apoiar o trabalho de documentação linguística, destacando suas funcionalidades e aplicações. Além disso, mostraremos como o material desenvolvido na plataforma pode ajudar a pesquisa em linguística e antropologia, bem como o ensino de língua e arte verbal nas escolas indígenas. Toda a documentação e anotação em português e em línguas nativas.

O que você vai aprender?

Uso e funcionalidades da plataforma Tycho Brahe para o registro de dados de línguas indígenas.

---

## **BIBLIOGRAFIA**

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. <https://doi.org/10.20396/cel.v59i3.8651003>

---