

Apresentação do Projeto DOC

(Didática, Oral, *Corpus*)



1.

PARCEIROS INSTITUCIONAIS



Ortolang

- Open Ressources and TTools for LANGuage
- Outils et Ressources pour un Traitement Optimisé de la LANGue

Website

<https://www.ortolang.fr/>



CORLI

Consortium CORpus, Langue et
Interaction

<https://corli.huma-num.fr/>



ULille

Universidade de Lille

<https://www.univ-lille.fr/>



STL – Savoirs, Textes, Langage

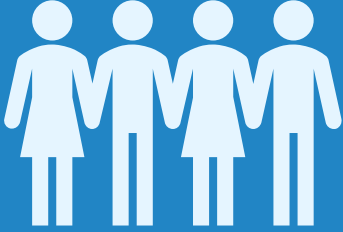
Equipe de pesquisa

<https://stl.univ-lille.fr/>



CNRS – Centre national de la
recherche scientifique

<https://www.cnrs.fr/>



2.

EQUIPE

▷ Duas coordenadoras científicas

▷ **Emmanuelle Canut**

- ▷ linguística da aquisição da linguagem, aprendizagem de língua segunda e língua estrangeira; didática do francês primeira língua e língua estrangeira

▷ **Juliette Delahaie**

- ▷ semântica, linguística, interação, francês língua estrangeira, francês falado, *corpora*, marcadores discursivos

▷ Dois professores-pesquisadores

▷ **Henry Hernández-Bayter** (espanhol)

- ▷ estudos da tradução, análise do discurso, fraseologia, *corpus*, lexicometria, América Latina contemporânea
- ▷ responsável pelo domínio “espanhol latino-americano (Colômbia, México)”

▷ **Liliane Santos** (português)

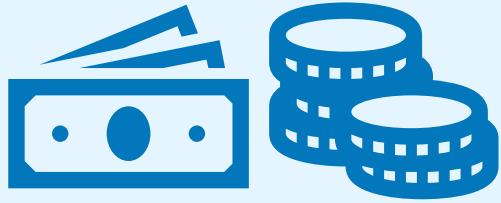
- ▷ marcadores discursivos, pragmática dos atos de fala, tradução, ensino/aprendizagem de PLM e PLN
- ▷ responsável pelo domínio “português (Brasil, Portugal)”

- ▷ Um pesquisador
 - ▷ **Driss Sadoun**
 - ▷ Diretor de PostLab (Marselha)
 - ▷ informática, processamento automático das línguas naturais
 - ▷ responsável pela integração das transcrições online; criação da página DOC na plataforma ORTOLANG

- ▷ Uma professora contratada
 - ▷ **Simona Anastasio**
 - ▷ aquisição contrastiva francês/italiano
 - ▷ responsável pelo domínio “italiano”

- ▷ Uma doutoranda
 - ▷ **Rou Chen**
 - ▷ responsável pela constituição do corpus em mandarim

- ▷ Uma mestranda
 - ▷ **Grace Mercy Kitcher**
 - ▷ responsável pela constituição de um corpus contrastivo francês (França/Benim) (debates argumentativos)



3.

FINANCIAMENTO



CORLI

Consortium CORpus, Langue et
Interaction

(TGIR Huma-Num)

<https://corli.huma-num.fr/>

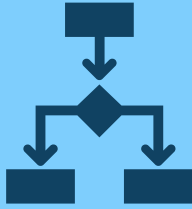


4.

OBJETIVOS

Um duplo objetivo

- ▷ implementar uma plataforma multimídia centrada no ensino da modalidade oral de línguas estrangeiras (chinês, espanhol, francês, inglês, italiano, português), numa perspectiva comparatista, intercultural e linguística;
- ▷ fornecer aos pesquisadores em didática de línguas-culturas e em linguística (principalmente em linguística da língua falada e em análise das interações, mas também em Processamento Automatizado das Línguas Naturais – PNL). dados multilíngues comparáveis, compostos de interações orais espontâneas.



5. “CORPUS COMPARÁVEL”



um conjunto de corpora composto de textos similares numa língua ou numa variante (Sinclair, 1996).

SINCLAIR, J. (1996). *EAGLES, Preliminary recommendations on Corpus Typology. EAG--TCWG--CTYP/P*. Version of May, 1996.

<http://www.ilc.cnr.it/EAGLES96/corpustyp/node21.html#SECTION00011000000000000000000000000000>

não existe um consenso estabelecido sobre a natureza dessa similaridade. Assim, o domínio e os gêneros textuais podem ser características pertinentes para a composição do *corpus*. é este o critério que escolhemos para o nosso projeto.

- ▷ situações interacionais clássicas dos manuais de ensino de línguas
- ▷ dois campos de utilização da língua ⇒ *Quadro Comum de Referência para as Línguas* (QECRL; ver Conselho da Europa, 2001):
 - *domínio pessoal*
 - encontrar (vizinhos, um desconhecido numa festa); convidar (para jantar, por exemplo); dar a opinião sobre um filme, elaborar um projeto (por exemplo, planejar as férias)
 - *domínio público*
 - pedir conselhos sobre um produto; interagir em comércios; ir ao restaurante; interagir numa agência de viagens

▷ três etapas simultâneas:

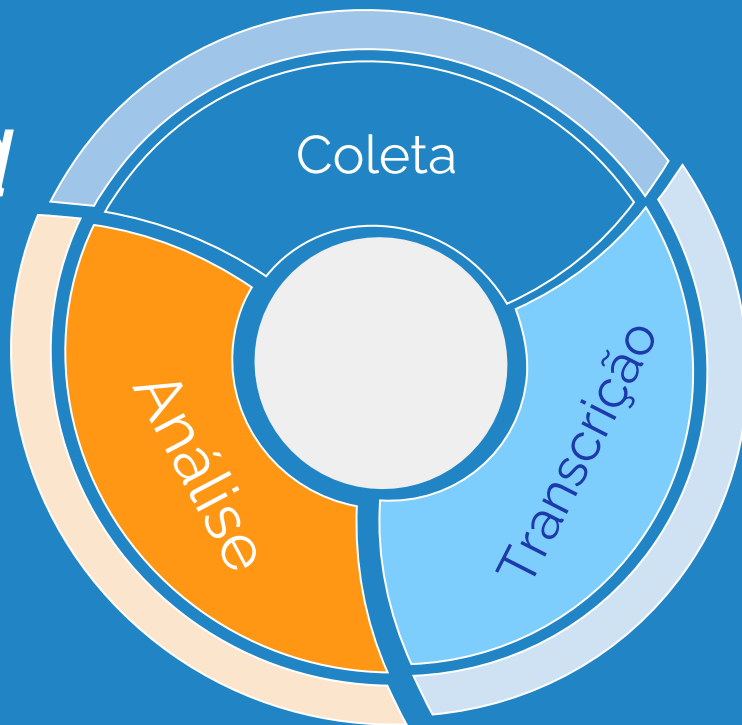
- **Etapa 1:** constituição de um *corpus* multilíngue de dados autênticos de interação em situações de comunicação ordinária (encontros na rua, conversas com os vizinhos, interações comerciais e telefônicas, entrevistas de emprego, etc.), que impliquem a coleta de dados nos países envolvidos.
- **Etapa 2:** tratamento dos dados (anotação linguística) no quadro do consórcio *CORLI* (base de dados *ORTOLANG*), linha de pesquisa multilinguismo/plurilinguismo (coordenadores: Antonio Balvet e Evangelia Adamou).
- **Etapa 3:** análise linguística e exploração didática dos dados para a criação da plataforma aberta e colaborativa DOC, sob a forma de atividades pedagógicas, de um dicionário da modalidade oral, de interações simuladas e de dados autênticos disponibilizados online. O caráter multilíngue dos dados permitir-nos-á inscrevermo-nos plenamente numa abordagem intercultural das interações e do ensino de línguas, tal como valorizada pelo Conselho da Europa desde o advento do *QECRL*.

Estado atual do projeto

- ▷ janeiro de 2022
- ▷ etapas 1 e 2
- ▷ dados já coletados: plataforma ORTOLANG (<https://www.ortolang.fr/market/corpora/doc-stl/>).

6.

Corpus/ Corpora



Corpora

- ▶ Mesmas características
 - mesmos gêneros interacionais
 - dados comparáveis

Três gêneros interacionais

- ▷ Conselhos (sobre um produto, num estabelecimento comercial)
- ▷ Encontros (na rua, na faculdade)
- ▷ Opinião (sobre um filme, uma série, um vídeo)



7.

COLETA DOS DADOS

- ▷ Conselhos numa loja
- ▷ Encontros fortuitos (na rua, na faculdade)
- ▷ Opinião sobre um filme

Coleta dos dados

- ▷ Diferentes países
 - França, Espanha, Brasil, Itália, etc.
- ▷ Professores
 - Implicados na formação dos alunos
- ▷ Formação dos alunos
 - Transcriber
 - Alfabeto SAMPA
 - Explicação dos dados a coletar
 - Formato da transcrição
 - Autorizações

Quem coleta os dados?

- ▶ Trabalho com os alunos
 - Convenções traduzidas
 - Bianca Mori
 - Cíntia do Nascimento Severino
 - Transcrições
 - Transcriber (Emmanuelle Canut)

Em andamento

- ▷ *Corpora* contrastivos
 - Francês/espanhol (opinião, encontros, conselhos)
- ▷ *Corpus* nativo/não nativo em francês
 - França/Benim (Grace Mercy Kitcher)
- ▷ *Corpora* comparáveis francês/chinês
 - Opinião sobre um filme
 - Dados por enquanto não disponíveis

Aspectos técnicos

- ▷ Como nomear os *corpora*
- ▷ De que arquivos precisamos e como nomeá-los
- ▷ Como anonimizar os dados
- ▷ Convenções de transcrição



8.

ANÁLISE DOS DADOS

▷ Perspectiva comparativa



Verbos parentéticos

FR: je pense, je trouve, je crois, je (ne) sais (pas), tu vois

ES: creo que, me parece que



Opinar

FR: falar de si mesmo

ES: apresentar um ponto de vista exterior



Ponto de vista intercultural

Identidade social



Outros

Semântica

Análise do discurso



Didática

Gramática do oral

Interação

Expressão do desacordo



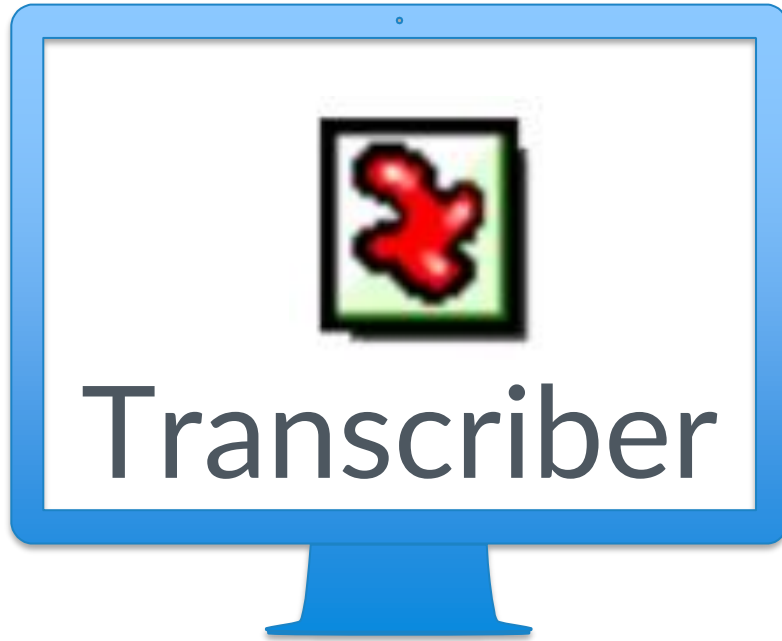
9.

FERRAMENTAS



Conversão de arquivos no
formato .wav
Utilização simples

<https://www.audacityteam.org/download/>



Transcrição ortográfica
Utilização fácil

<https://sourceforge.net/projects/trans/files/transcriber/>

Para particularidades da pronúncia

Speech Assessment Methods Phonetic Alphabet (“Alfabeto fonético de métodos de avaliação da fala”)

<https://pt.wikipedia.org/wiki/X-SAMPA>

<https://www.phon.ucl.ac.uk/home/sampa/>





Sketch Engine
 (<https://www.sketchengine.eu/>)?

Outro(s)?





Documentos
escritos

Traduzir
Assinar

10. CONVENÇÕES

Muito obrigada!



Free templates for all your presentation needs



For PowerPoint and
Google Slides



100% free for personal
or commercial use



Ready to use,
professional and
customizable



Blow your audience
away with attractive
visuals