

V SEMINÁRIO INTERNACIONAL DE FONOLOGIA

Dataset para documentação do contato Warao-Português em Belo Horizonte: parâmetros de compilação de dados

UFMG

U.PORTO



Utrecht University

Átila Vital – UFMG

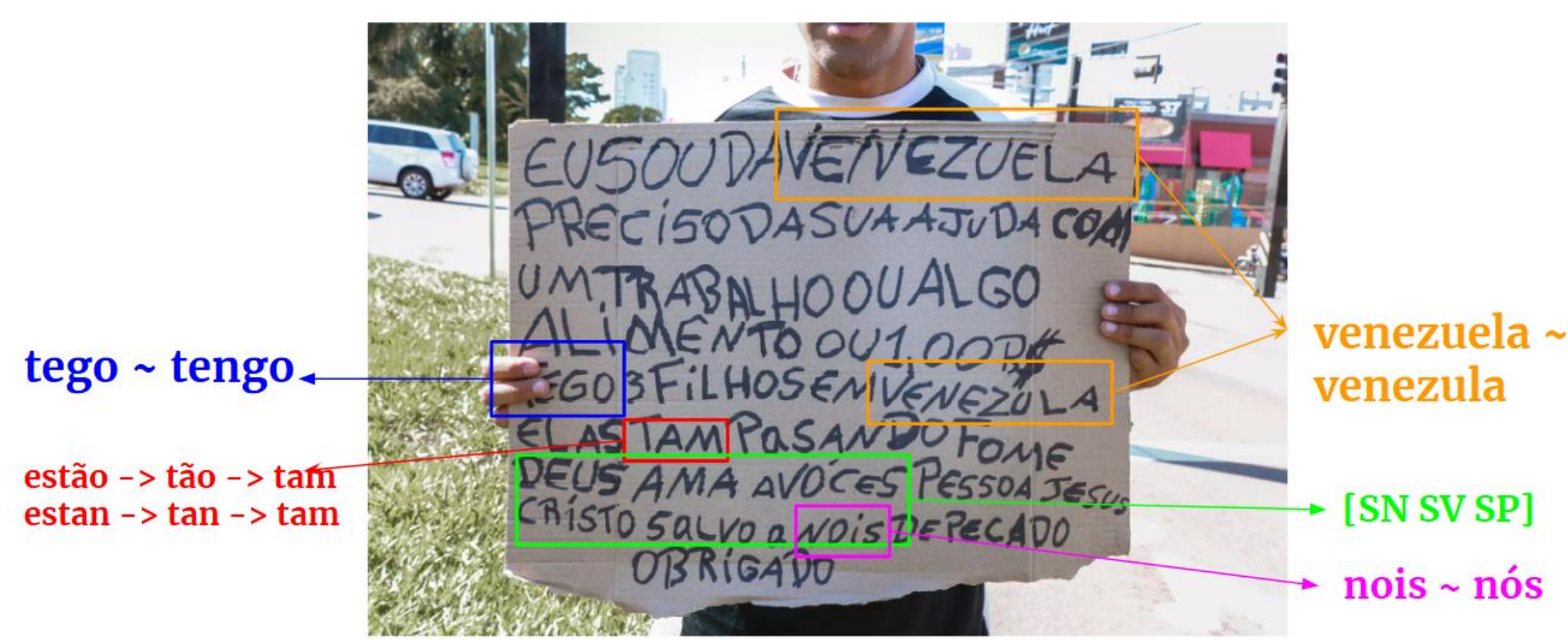
Dalmo Buzato – UFMG

Fábio Granja – Utrecht University

Carlos Silva – UPORTO

INTRODUÇÃO

- As línguas de contato costumam apresentar baixo prestígio sociolinguístico e, em geral, não são incluídas nos currículos escolares, tampouco possuem mecanismos institucionais de promoção social, como produção literária, presença midiática ou uso em contextos governamentais.
- A documentação pode se dar através da **compilação de corpora multimodais** de línguas de contato (Nagy, 2011; Mello, 2014; Adamou, 2016; Léglise and Alby, 2016);
- Este estudo relata o desenvolvimento de um dataset com **produções orais e escritas de refugiados venezuelanos** da etnia **Warao** no Brasil;
- População de +45.000 pessoas na Venezuela, falantes da língua Warao, tipologicamente isolada (Romero-Figueroa, 1997);



METODOLOGIA

- Compilação de dados escritos e orais: placas com pedidos de ajuda e entrevistas do Youtube com migrantes Warao em Belo Horizonte;
- Anotação de fenômenos fonético-fonológicos, code-switching; os dados orais são transcritos semi-ortograficamente e segmentados em unidades tonais;

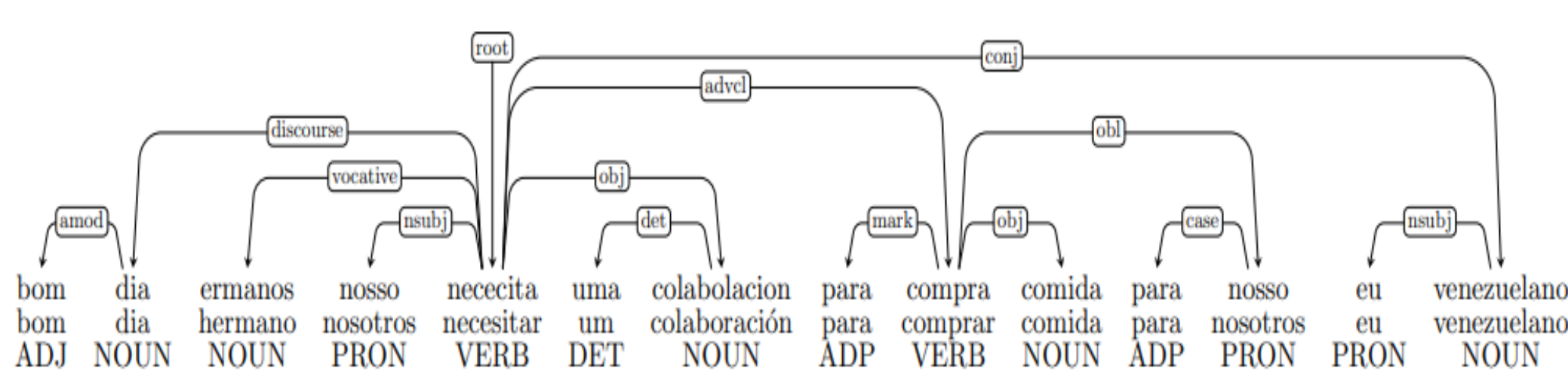


Família Warao da Venezuela mantém tradições em meio à procura por sustento em BH

"Socorro", grita Povo Indígena Warao, da Venezuela, refugiado em BH/MG (300), no Brasil + de 7 mil

"Socorro", grita Povo Indígena Warao, da Venezuela, refugiado em BH/MG (300), no Brasil + de 7 mil

Transcrição
bom dia ermanos
nosso nececita uma
colabolacion
para compra
comida para nosso
eu venezuelano

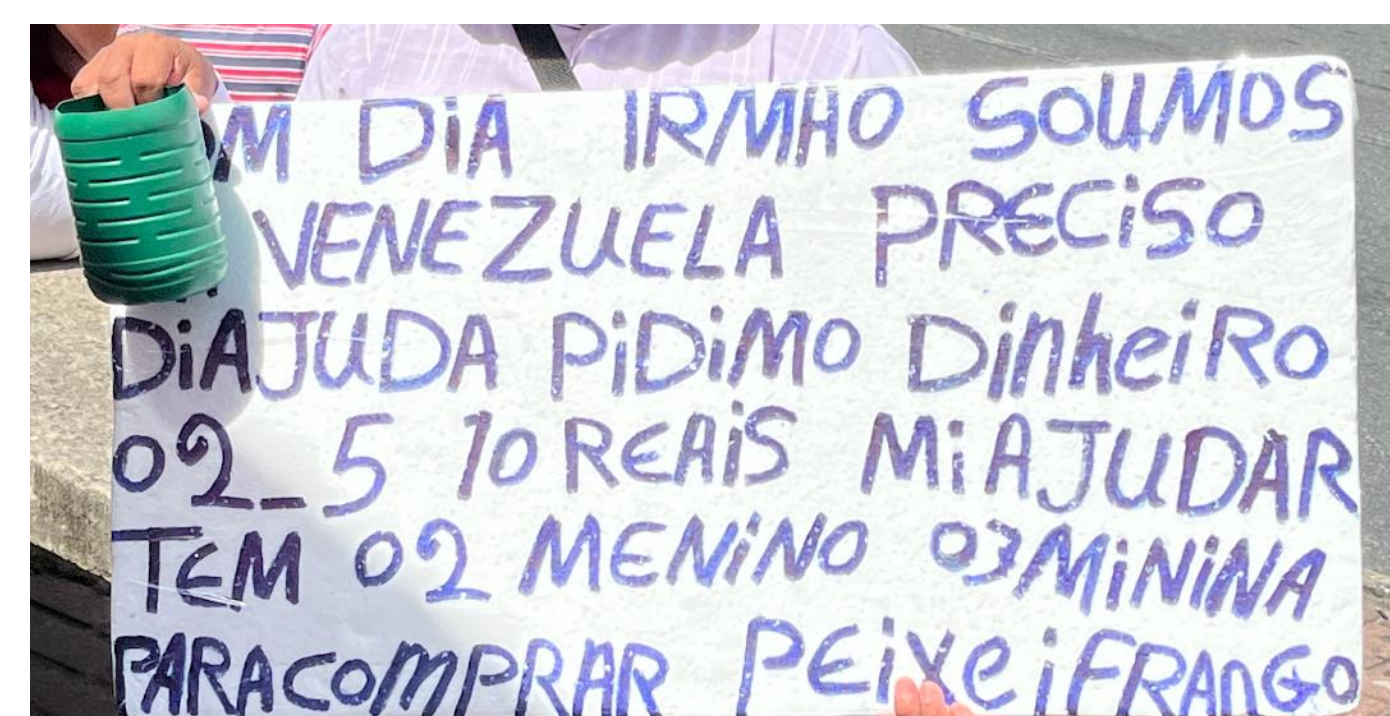


- Critérios de transcrição adaptados dos critérios do C-ORAL-BRASIL I (Raso & Mello, 2012); e coletas de palavras da lista Swadesh em PB para verificar a estabilidade lexical em falantes de Warao.

RESULTADOS E DISCUSSÃO

Para as placas

- 60 placas transcritas e morfossintaticamente anotadas em UD;
- Registro de fenômenos fonéticos resultantes das línguas em contato e da grande influência do Espanhol;



Para os dados orais

- Até o momento, foram transcritos trechos de entrevistas no Youtube;
- São previstas 20 transcrições de fala espontânea dos Warao para o dataset;

Critérios de transcrição

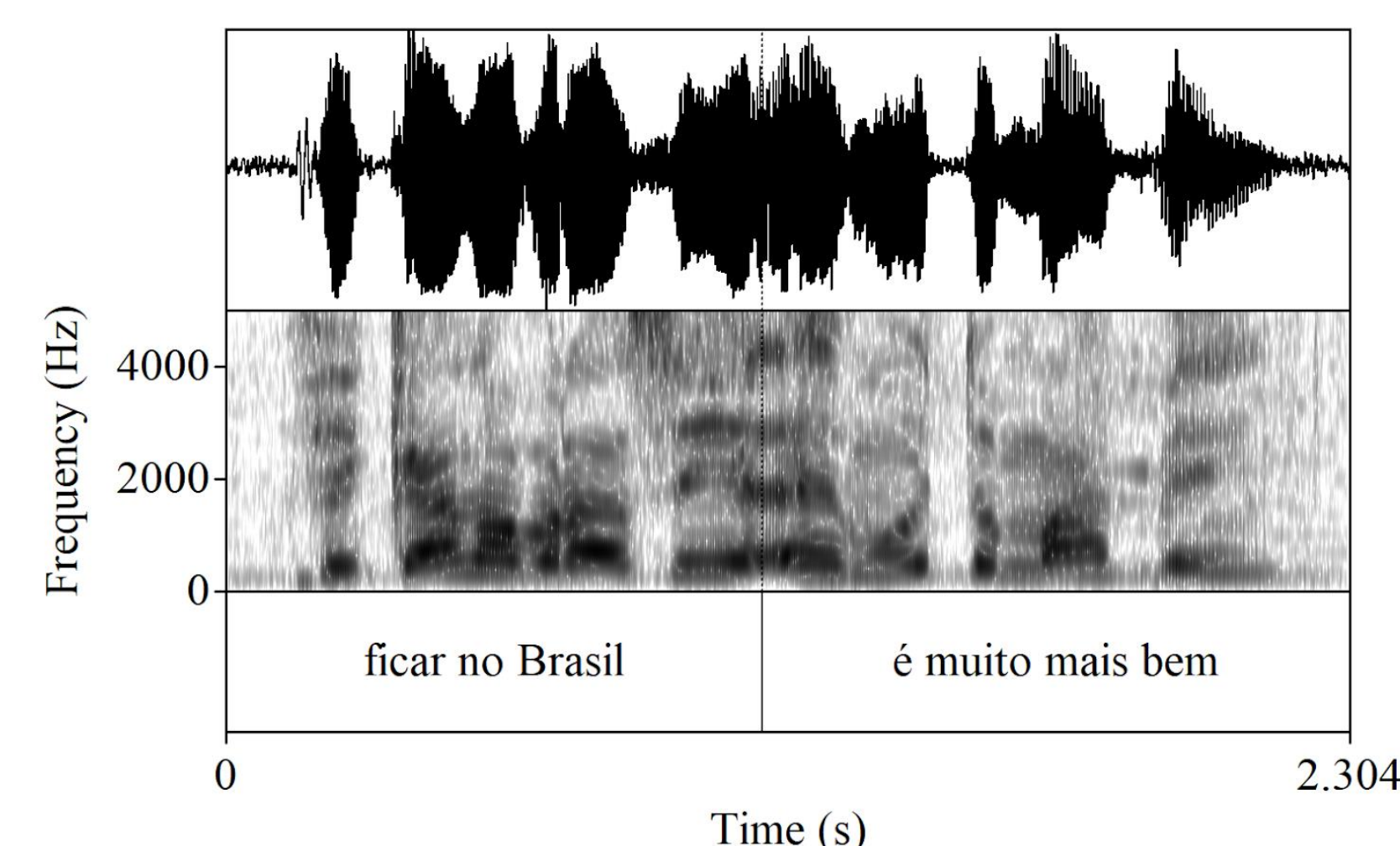
- Transcrição e anotação seguindo os critérios do C-ORAL-BRASIL I;
- Veiculação da estrutura informacional através da prosódia (Izre'el et al., 2020);
- Critérios semi-ortográficos capazes de capturar cliticizações, formas aféricas, apagamento de morfologia verbal, novos paradigmas pronominais, disfluências e muitos outros fenômenos;

Example 2 (documentary_VAR)

Negação	não	nũ	n' é não	non	no
"Mais"	mais	ma			
"Sim"	sim	si			
"Mas"	mas	mai			

VAR: lá / passava muita / dificuldade / por falta de / &m [1] da medicamento // porque / muita [1] muita criança // &he / muito / homem / mulher / vovó / &fa [1] faleciam / porque / faltava de [1] de medicamento lá // si / mas na [1] na mi [1] alimentação / não nos chega //

- Até o momento, as entrevistas do Youtube transcritas possuem qualidade acústica razoável;



@Title: documentary_VAR

@File: VAR

@Participants: VAR, John Vargas (male, unknown, unknown, Warao immigrant, participant, Venezuela)

@Date: unknown

@Place: Belo Horizonte (MG)

@Situation: documentary made by "Jornal o Tempo" about the Warao immigration @Topic: the life in Venezuela and the reasons why his family came to Brazil

@Source: YouTube

@Length: 39"

@Words: 64

@Transcriber: Átila Vital

@Comments: The audio has a music in a very low volume from the documentary

1) Forms originated by contact: at 10", VAR speaks "bobó", instead of "vovó" (grandmother). At 36", VAR speaks "possible", instead of "possível" (possible).

2) External noises: in some moments, there are sounds of children playing.

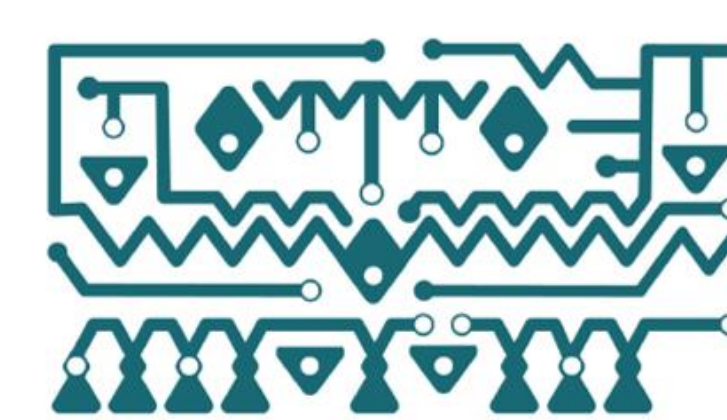
Próximos passos

- Coleta e processamento da lista Swadesh para 11 categorias semânticas (Cooking / Human body / Creatures / Kinship / Houses / Tools & Household utensils / Earth and sky, Seasons and weather & Geographical features / Professions / Numbers / Calendar / Botanical terms / Clothing);
- Análise léxico-estatística no Cog

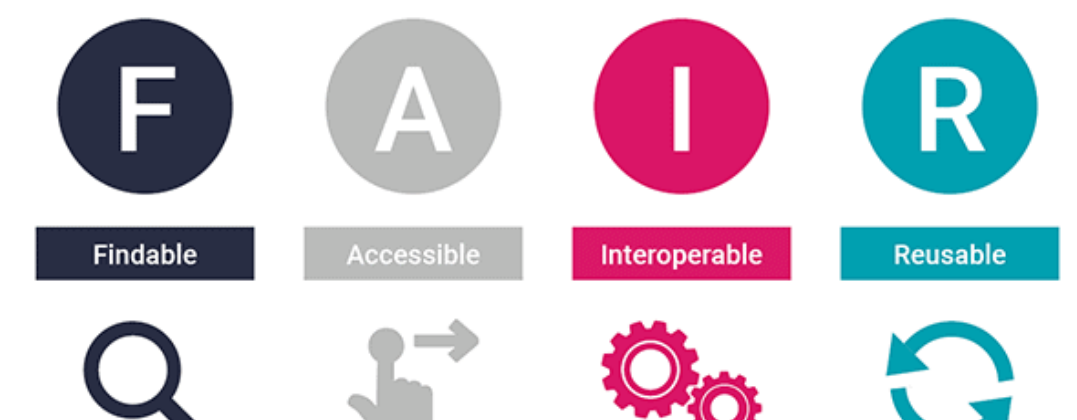


CONSIDERAÇÕES FINAIS

- O dataset reunirá conjuntos de dados escritos e orais de imigrantes venezuelanos e estará disponível para pesquisas em variados campos linguísticos;



CARE Principles
for Indigenous
Data Governance



UFRJ, Rio de Janeiro/Brasil
6-7 de novembro de 2024



APOIO:



Programa de Pós-Graduação em
Letras Vernáculas
Universidade Federal do Rio de Janeiro

PrInt

