

Normas adotadas pelo Grupo de Estudos em Linguagem Interação e Sociedade (GELINS) para a realização de transcrição ortográfica e catalogação das entrevistas coletadas

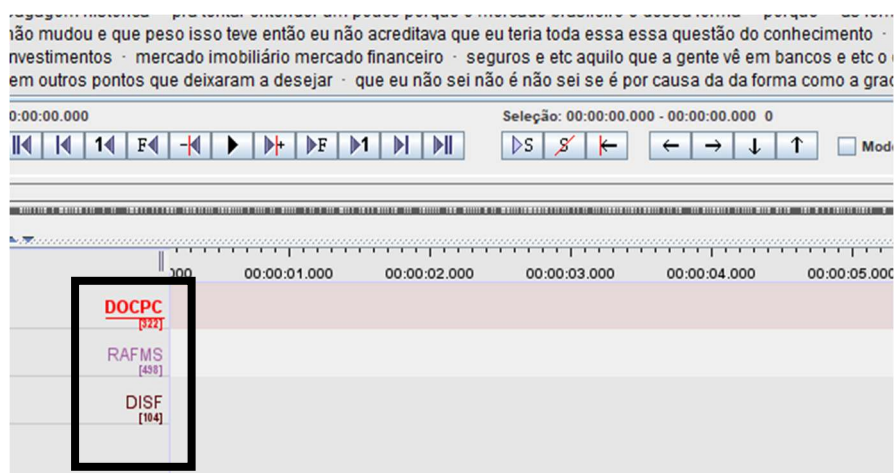
Atualizado por Paloma Cardoso, durante a coleta/anotação da amostra
Linguagem Corporificada 2023

1. Objetivos

- Catalogar usos linguísticos, preservando ao máximo possível informações referentes à situação de interação.
- Padronizar transcrições de dados de fala usados em pesquisas nos níveis fonético, fonológico, morfossintático, semântico e pragmático da língua.
- Facilitar a localização de fenômenos linguísticos marcados ortograficamente (por palavras ou etiquetas de disfluências), e o uso de ferramentas automáticas para busca em grande volume de dados.

2. Estrutura da transcrição

As transcrições dos dados de fala são feitas no ELAN (ver anexo 1), seguindo as normas de ortografia do português, em um modelo de divisão por trilhas (linhas), conforme permitido pelo software. A cada falante (incluindo a/o documentadora/o), atribui-se uma trilha.



Por padrão, atribui-se uma trilha para a/o documentadora/o, outra para a(s) pessoa(s) entrevistada(s) e uma trilha de disfluências. A trilha da/o documentadora/o é nomeada DOCXX, em que “XX” corresponde às iniciais de quem fez a entrevista. A trilha da pessoa entrevistada é nomeada com as três primeiras letras no nome, mais “m”, “f” ou “nb” para indicar o sexo/gênero desse falante (masculino, feminino, não-binário, respectivamente). A trilha de disfluências, nomeada DISF, engloba marcações sobre barulhos externos à entrevista (pigarro, tosse, risos, etc), hesitações (ah, eh... sons vocálicos que não têm função morfossintática, ou seja, não atuam como artigos ou preposições), estímulos de continuidade (uhum, hum).

Barulhos externos, hesitações e estímulos de continuidade, sejam eles realizados pela/o documentadora/o ou pela pessoa entrevistada, devem ser marcados na trilha DISF. Conforme listado na tabela abaixo, para marcar hesitações deve-se usar, sempre **em caixa alta**, (HES), nesse padrão. Para barulhos como tosse e risadas, deve-se usar ((TOSSE)) ((RISOS)). Para estímulos cujo objetivo é manter a interação, ou seja, de continuidade, deve-se usar (EST).

Na seção 4, constam as normas de nomeação das trilhas da/o documentadora/o e da pessoa entrevistada.

3. Diretrizes de transcrição

A transcrição dos dados de fala deve ser feita de acordo com as normas de ortografia do português, considerando algumas ressalvas:

1. Não utilizar sinais de pontuação: ponto, vírgula, ponto e vírgula, exclamação, reticências.
2. Utilizar apenas o hífen com palavras compostas: guarda-roupa, beija-flor
3. Não marcar variação fonética/fonológica (variações de pronúncia de tia, por exemplo)
4. Manter somente os acentos agudo, circunflexo e cedilha, nos casos previstos pelas normas ortográficas.
5. A variação morfossintática é preservada (as menina, a gente fomos, etc).

O objetivo da transcrição é manter o máximo de informações comunicativas possível. Na tabela abaixo, apresentamos diretrizes referentes à dúvida (por parte de quem transcreve) em relação ao que foi dito, truncamento de palavras, etc.

OCORRÊNCIA	SINAIS	EXEMPLIFICAÇÃO
Interrogação	?	sabe o é?
Comentário do transcritor sobre o que está acontecendo no ambiente	(())	((RISOS)) ((BARULHO)) – com exceção do riso/risada, todos os barulhos do ambiente (tosse, pigarro, batida na porta, etc) devem ser demarcados com ((BARULHO))
Estímulo do interlocutor	(EST)	(EST) Consideramos estímulo os sons cujo objetivo é atuar na manutenção da interação, no sentido de “estou ouvindo”, “estou prestando atenção ao que você está dizendo”. Essa marcação deve estar na trilha DISF.
Hesitação do locutor	(HES)	foi (HES) uma brincadeira bem interessante. Essa marcação deve estar na trilha DISF.
Truncamento de palavra	/	come/ começou
Nomes próprios, profissões, nomes de cursos, filmes	Inicial maiúscula	... fui à Petrópolis uma ...
Discurso direto	“ ”	eu saio pra apresentar trabalho fora eles têm orgulho “ah ela saiu pra outro estado tá apresentando trabalho da universidade” então de certa forma isso é um apoio...
Números	por extenso	eu tenho vinte e oito anos
Incompreensão do que ouviu	()	
Hipótese do que ouviu	(hipótese)	ter que estudar lá no no campus de São Cristóvão ia re/ ia reque/ requerer da minha (como a associação) que eu teria que pagar todos os meses...
Onomatopeias e siglas	caixa alta	a questão do incentivo de participação de eventos porque assim de eventos por exemplo a OCMEA ela é incentivado por todos os professores

Atenção: as marcações listadas na tabela em caixa alta devem ser usadas, igualmente, em caixa alta nas transcrições.

4. Metadados das transcrições e dos arquivos correspondentes

No Gelins, trabalhamos com dados coletados em entrevistas (situação 1x1) e interações em grupo (mais de 2 pessoas). Esse tipo de informação deve estar na

transcrição (no número de trilhas, uma para cada mais a trilha DISF) e nos nomes dos arquivos de áudio, .pfsx e .txt (estes dois últimos são gerados pelo ELAN) correspondentes às transcrições.

4.1 Metadados nos arquivos de áudio e .pfsx

O tipo de dado de fala depende do tipo de amostra. Esse tipo de informação faz parte dos metadados de transcrição. São metadados:

1. O tipo de amostra: entrevista, interação, reunião, demarcado como, respectivamente: ent, int, reu.
2. Nome da comunidade: UFS-SaoCristovao, IFS-Itabaiana (siglas em maiúsculo, nomes próprios com iniciais maiúsculas, sem acento agudo, circunflexo ou cedilha).
3. Ano de gravação.
4. Três letras iniciais do nome do informante.
5. Sexo/gênero: f (feminino), m (masculino), nb (não-binário)
6. Escolaridade: fundamental (f) ensino médio (m), superior (s)
7. Idade

A indicação de todas essas informações resulta na seguinte etiqueta, que deve ser utilizada para nomear o arquivo de áudio, o arquivo .pfsx, e o .txt:

40.ent.UFS-SaoCristovao.2023.raf.ms.22

Os dados de fala são documentados durante o processo de constituição de amostras. Cada gravação é enumerada para fins de controle. O número atribuído a cada uma é aleatório. A etiqueta acima corresponde à gravação 40, uma entrevista (ent) realizada no campus de São Cristóvão (UFS-SaoCristovao), em 2023. raf corresponde às iniciais do informante, uma pessoa do sexo/gênero masculino, de escolaridade superior e com 22 anos.

Cada amostra é constituída para atender a objetivos específicos. O banco de dados Falares Sergipanos, constituído pelo Gelins, possui amostras cujos objetivos são: analisar a influência da mobilidade (onde nasceu > onde mora) na fala, analisar a relação entre fala e gestos corporais. Essas amostras são nomeadas, respectivamente, Deslocamentos e Linguagem Corporificada. No caso da amostra Deslocamentos, há uma divisão em 4 subgrupos. Essa particularidade também está demarcada nos nomes dos arquivos das

referidas amostras, por algarismos romanos na amostra Deslocamentos 2019, e por algarismos arábicos na amostra Deslocamentos 2020:

21ent.UFS-SaoCristovao2018__desl. II_final_dou.ms.25

66ent.UFS-SaoCristovao2020_desl4_inicio_isa_quimica.fs.19

Atenção: a divisão entre informações nas amostras Deslocamentos 2019 e 2020 foi feita por “_”. **Para fins de sistematização, essa mesma divisão deve ser feita por ponto:**

40.ent.UFS-SaoCristovao.2023.raf.ms.22

Quando houver informação sobre mobilidade (de acordo com o desenho da amostra Deslocamentos), deve-se marcá-la com algarismos arábicos (1, 2, 3, 4).

4.2 Metadados no arquivo .txt

Todas as transcrições realizadas no ELAN têm um arquivo .txt (ver tutorial do ELAN sobre como ele é gerado). No arquivo de texto, que deve ter o mesmo nome do arquivo de áudio e .pfsx, deve-se acrescentar, manualmente, exatamente como exposto abaixo (em caixa alta), as seguintes informações:

TIPO DE AMOSTRA: ENT

DOCUMENTADOR(A): PALOMA

COMUNIDADE: UFS

ANO: 2020

DESLOCAMENTO: 1

TEMPO DE CURSO: INICIO

INF: ADR1MI

SEXO: M

ESCOLARIDADE: S

IDADE: 18

CIDADE NATAL: ARACAJU/SE

GRAVAÇÃO: CABINE

QUALIDADE DO ÁUDIO: VOZ AUDÍVEL, SEM RUÍDOS

As informações acima devem ser acrescentadas no arquivo .txt exatamente como exposto. Nas linhas em que a informação requerida for inexistente/não estiver disponível, deve-se digitar apenas um hífen (-). Esse será o caso de amostras que não forem desenhadas de acordo com a metodologia deslocamentos. No .txt, acontecerá isso:

DESLOCAMENTO: 1

Quanto à informação referente à GRAVAÇÃO, deve-se indicar o local no qual a entrevista foi gravada: CABINE, CEPITEC

Quanto à avaliação da qualidade do áudio, deve-se indicar apenas se a voz é audível ou inaudível, com ou sem ruído, desse modo:

QUALIDADE DO ÁUDIO: VOZ INAUDÍVEL, COM RUÍDOS

QUALIDADE DO ÁUDIO: VOZ AUDÍVEL, SEM RUÍDOS